HAL open science

The choice of shipment size in freight transport

François Combes

To cite this version:

François Combes. The choice of shipment size in freight transport. Architecture, space management. Université Paris-Est, 2009. English. NNT : 2009PEST1075 . tel-00489371v2

HAL Id: tel-00489371

https://pastel.hal.science/tel-00489371v2

Submitted on 6 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris-Est

Ecole Doctorale Ville Environnement

Laboratoire Ville Mobilité Transport

Thèse de doctorat

Spécialité Transport

The choice of shipment size in freight transport

Présentée et soutenue publiquement

Par François Combes

Le 14/12/2009

Sous la direction de Fabien Leurent

Composition du jury:

Elisabeth Gouvernal, Présidente

Lóránt Tavasszy, Rapporteur

Alain Bonnafous, Rapporteur

Bart Jourquin, Examinateur

Michel Savy, Examinateur

Fabien Leurent, Directeur

Abstract Summary

Classic spatialised freight transport models are based on a four stage representation of the decisions shippers and carriers take. The decisions this representation distinguishes are: emitted and received flow rates, supplier or receiver choice, transport mode, itinerary. However, freight transport is discrete by nature: commodities are moved by bundles called shipments. Shipments are absent from the four stage representation.

This study is aimed at investigating the role of the choice of shipment size in the freight transport system. After a review of recent freight transport modelling advances and of some logistic models, we proceed to a systems analysis of freight transport. Agents are identified, their behaviours and relationships are described. This allows for a clarification of the linkage between logistics and freight transport.

Then, attentions is paid to the empirical observation of the freight transport system. Propositions are made to improve French roadside surveys, so as to better observe the productivity and technical choices of motor carriers. The seminal microeconomic model of optimal shipment size called Economic Order Quantity is assessed econometrically using the ECHO database.

Lastly, two particular issues are addressed using microeconomic models. First, the equilibrium freight rates of a schematic road freight transport market are modelled, on the basis of an explicit representation of the logistic imperatives of shippers and of the technology of carriers (the consolidation of shipments in vehicles is accounted for). Second, the logistic imperatives of shippers are analysed in detail to explain why a shipper would use two transport modes simultaneously for a unique commodity flow.

Résumé Court

La modélisation spatialisée de la demande de transport de fret est classiquement fondée sur une représentation à quatre étapes des décisions que prennent les chargeurs et les transporteurs; cette représentation distingue les décisions de volume émis et reçus, de choix de fournisseur ou destinataire, de mode de transport et enfin d'itinéraire. Mais le transport de marchandise est une opération de nature discrète : les marchandises sont transportées par blocs, ou envois. Ces envois sont absents de la représentation à quatre étapes.

Ce travail a pour but d'étudier le rôle du choix de la taille d'envois dans le fonctionnement du transport de fret. Après une revue de la modélisation du transport de fret et de certains problèmes logistiques, le transport de fret est analysé et décrit de façon systémique. Les agents en jeu et leurs comportements sont identifiés. La distinction entre consommation et production du transport de fret est établie, ce qui permet de clarifier le lien entre logistique et transport de fret.

Ensuite, l'attention est portée sur l'observation empirique du système de transport de fret. Des propositions sont faires pour améliorer les enquêtes en bord de route menées en France auprès des poids lourds. Elles concernent principalement la productivité et les options techniques des transporteurs routiers. Une validation économétrique du modèle microéconomique de taille d'envoi optimale Economic Order Quantity est effectuée au moyen de la base de données ECHO.

Enfin, la modélisation microéconomique est employée pour traiter deux sujets en particulier. Premièrement, pour analyser en détail la formation des prix d'équilibre de transport de fret, en représentant les impératifs logistiques des chargeurs et la technologie des transporteurs (notamment la consolidation d'envois). Deuxièmement pour représenter en détail le lien entre la logistique des chargeurs et leur demande de transport de fret, afin, entre autre, de pouvoir modéliser l'usage simultané de deux modes de transport par un unique transporteur pour un unique flux de marchandises.

Contents

List of Figures

List of Tables

Notations

a	:	value of time per unit of time and weight.
a_c	:	commodity value of time, per unit of time and weight.
$\bar{a}_c$	:	maximum commodity value of time, over which the heavy transport mode is not attractive at all in the two modes case.
a_{dens}	:	commodity density value.
a_w	:	warehousing cost.
α	:	shipper's value of travel time savings.
α_l	:	shipper's value of light mode travel time savings.
α_r	:	shipper's value of heavy mode travel time savings.
α_{rho}	:	value of travel time reliability.
b	:	access cost.
$\mathcal{B}(n,p)$	:	multinomial distribution of parameters $n \in \mathbb{N}$ and $p \in [0;1]$.
β_X	:	estimated coefficient before variable X in an econometric specification.
c	:	haulage cost.
c_t	:	transport unit cost.
c_l	:	light mode transport unit cost.
c_h	:	heavy mode transport unit cost.
C	:	(expected) cost function.
C_c	:	customer cost.
C_d	:	inventory cost.
C_p	:	pipeline inventory cost.
C_τ	:	transport cost.
d	:	distance.
d_h	:	heavy mode shipment size.
D_t	:	daily demand at time t.
D_t^r	:	sum of the demand over the l past days at time t.
Δ	:	number of exceeding shipments of size 1.
δ	:	expected value of the exceeding shipments of size 1.

ε	:	small variation.
f_a	:	distribution density of the commodity unit values of time.
F_a	:	c.d.f. of the commodity unit values of time.
f_d	:	daily demand distribution density.
F_D	:	daily demand c.d.f.
φ	:	centered unit normal distribution density.
Φ	:	centered unit normal c.d.f.
g	:	unit generalised transport cost.
I_t	:	inventory at time t.
I_t^E	:	excess inventory at time t.
I_t^p	:	pipeline inventory at time t.
K_x	:	various constants.
λ	:	loading factor of a vehicle.
λ^e	:	equilibrium loading factor.
l	:	transport lead time.
l_l	:	light mode transport lead time.
l_h	:	heavy mode transport lead time.
μ_I	:	destination inventory expected value.
μ_{I^E}	:	excess inventory expected value.
n_i^s	:	industry demand for services of transport of shipments of size i.
N_i^s	:	amount of services of transport of shipments of size i asked from a carrier.
$\mathcal{N}(m, \sigma^2)$	:	normal distribution of mean m and variance σ^2.
p	:	transport price.
$p(s)$	:	transport price schedule, function of s.
π	:	(expected) profit function.
q_i^s	:	amount of services of transport of shipments of size i supplied by a carrier.
Q_i^s	:	industry supply of services of transport of shipments of size i.
Q	:	yearly amount of goods of a given type sent by a firm to a given receiver.
Q_{tot}	:	yearly amount of goods of any types sent by a firm to a given receiver.
ρ	:	ratio of n_1^s over n_2^s.
s	:	shipment size.
s_t	:	size of the shipment sent at time t.
s_t^l	:	size of the shipment sent at time t by the light mode.

s_t^h	:	size of the shipment sent at time t by the heavy mode.
S	:	vehicle capacity.
σ	:	daily demand standard deviation.
σ_I	:	destination inventory standard deviation.
σ_{I^E}	:	excess inventory standard deviation.
σ_l	:	standard deviation of travel time.
$\mathcal{U}(a,b)$	:	uniform distribution over interval $[a;b]$.
X_i	:	1 if mode i is used, 0 else.
ζ	:	marginal cost of an increase in σ_I.

Aknowledgements

I am deeply indebted to Fabien Leurent, who was the mentor of this thesis. His constant availability, willingness to help, teach, and support me during this work, were invaluable.

I would like to thank Lóránt Tavasszy and Alain Bonnafous for the attention the brought my work, and for their useful and kind commentaries. I would also like to than Bart Jourquin and Michel Savy, who kindly agreed to participate to my examination board, and Elisabeth Gouvernal, who accepted to preside over it.

This work also owes much to the numerous and particularly interesting discussions I had with a number of my colleagues, including: Nicolas Wagner, Vincent Breteau, Nicolas Coulombel, Olivier Bonin, Thierno Aw, Laetitia Dablanc and Frederic Meunier. Without their help, some chapters of this thesis would certainly have been shorter, and less relevant.

For their useful comments, I would like to thank David Meunier, Valentina Carbone, Floriane Torchin, Pascaline Cousin, Patrick Niérat, Michèle Guilbault, Antoine Frémont, as well as Eric Jeannière.

For their hearty welcome and generous collaboration, I would like to thank Pierre Baillet, Yves Saige, Pierre Samblat, and Mathieu Jacquot. I'd also like to thank especially Michel Drougard, and Franck Goblot from ACL Consultants, who learnt me much of what I know from the activities of motor carriers.

For their welcome, I would like to thank the *Commission Transport et Mobilité de l'Académie des Technologies*, with whom I learned a lot. In particular, I would like to thank the members I most often met during my thesis, including: Jean-Claude Raoul, Bernard Decomps, Pierre Castillon, Olivier Maurel, Marc Pelegrin, Michel Frybourg, Xavier Karcher and Paul Caseau.

Finally, my gratitude for the support, patience and trust of my family, notably during this work but not only, is beyond my limited capacity of expression. I trust they understand how much I owe them, I dedicate them this thesis.

Abstract

The purpose of this work is to improve freight transport modelling, by the identification of qualitative facts and regularities which characterise the freight transport system, and by modelling them microeconomically. Particularly, it is aimed at understanding the behaviour of both carriers and shippers, and their relationships. Much attention is paid to the role of logistics in the freight transport system.

The current state of the art of freight transport modelling is by and large based on the four stages representation of transportation systems. In the case of freight transport, this representation is endowed with major shortcomings, which have been partially addressed by recent modelling advances. First, the definition of the origins and destinations of commodity trips is unclear, in particular if transshipment operations take place. Second, decision makers, and their microeconomic behaviours, are not explicitly identified. Third, the discrete nature of freight transport operations is absent.

One option to try to address some of these issues is to consider explicitly shipments, which are, from a decisional perspective, the atoms of the freight transport system. The objective of this work is thus to investigate the difficulties and potential benefits of representing explicitly the choice of shipment size in spatialised freight transport models.

First of all, we proceed to a systems analysis of freight transport, in order to identify decision-makers, their options, preferences and relationships, with the objective to improve the economic realism of freight transport models. This system analysis is based both on a qualitative analysis of freight transport and on a review of recent freight transport modelling advances.

As a result of this analysis, a layered representation of the freight transport system is designed. At the center of this representation is the door-to-door shipment transport operation, which is what is exchanged in the freight transport market. Subsequently, carriers are defined as the producers of these operations, and shippers as the consumers of these

operations. Before analysing with more detail the behaviour of carriers and shippers, it should be noted, at this stage, that classic freight transport surveys, which generally observe vehicle movements, are unable to observe the freight transport market directly, since door-to-door transport operations often involve one transshipment or more.

As producers, carriers obey to a classic microeconomic logic. Their objective is to produce cost-efficiently the transport operations they commit themselves to do, using their resources (*e.g.* vehicles, drivers, fuel, cross-docking platforms, etc.) To do so, they combine elementary transport operations (*i.e.* vehicle loading something somewhere and unloading something else somewhere else) in order to produce door-to-door shipment transport operations. Economies of scale, due for example to the capacity constraint of vehicles, and of scope, due to the possibility to carry together shipments going to distinct origins and destinations (hence which should, in a context of spatial economics, be considered as distinct products), are prominent in the transport technology, and explain the generalised use of various vehicles and modes and break-bulk transport operations by carriers. As a consequence, the linkage between vehicle movements and shipment movements is complicated. This linkage is defined as the "logistics of carriers". The issue of the freight transport market structure is not addressed, although it may play a significant role for some transport modes.

The microeconomic logic of shippers is more complicated. Indeed, the freight transport demand of shippers derives from their own logistic imperatives. The logistic imperatives of shippers stem from the preferences of the customers of shippers. Indeed, a given good provides utility to an economic agent only if it is available to this agent. In general, the agent is asked an effort to have the product he buys at his disposal. This effort can be a trip to a retail center, a delivery lead time, a risk of stock shortage. It is a user cost, formally similar to the generalised cost of passenger transportation. Decreasing the effort of its customers is costly to a shipper: it implies a series of logistic costs, such as more retail centers, faster deliveries, larger safety stocks, etc. Shippers must first find a trade off between their own logistic costs and the satisfaction of their customers, second organise their commodity flows efficiently. This process is defined as the "logistics of shippers", and bridges the gap between production-consumption flows as they are identified in input-output tables, and shipment door-to-door operations.

In classic freight transport models, the freight transport supply is generally represented by networks of links and nodes characterised by travel times, costs, and transport modes. Taking into account the shipment size variable in this representation is not straightforward. This issue is addressed microeconomically, in the case of road freight transport.

The linkage between shipment size and transport price between a given origin-destination pair is not simple. Freight transport price schedules, which give transport prices as functions of shipment sizes, are not linear: they are positive in zero and tend to be flat for big shipment sizes. Our objective is to explain microeconomically these phenomena; in other words, to explain how transport prices result from the trucking technology and of the costs of the resources of motor carriers. Perfect competition is assumed on the road freight transport market.

The trucking industry is characterised by a relatively simple technology, compared to other transport modes. Assume a transport operation consists of an access movement on a relatively short distance towards the origin, where the shipment is loaded, then a haulage movement over a long distance, then an access movement followed by an unloading operation at destination; assume these operations are made using vehicles of limited capacity. If several shipments are loaded at the same place, access movements are assumed specific to each shipment, whereas the haulage movement is shared among shipments. Other spatial constraints such as empty return or trip chaining are disregarded. Shipment sizes are endogenous.

There is a difficulty when deriving the cost function: a crucial parameter of this function is the loading factor, *i.e.* the average share of vehicle capacity actually occupied by freight. Optimising the average loading factor of a fleet, *i.e.* minimising the number of vehicles necessary to carry a set of shipments of given sizes, is the well known bin-packing problem, which is NP-hard. To overcome this difficulty, we make a somewhat oversimplifying hypothesis, and assume shipments are of size one, two or three, with a vehicle capacity of three. It is thus possible to derive the cost function explicitly, hence the marginal costs and market prices.

Despite this simplification, this approach efficiently explains the qualitative properties of freight transport price schedules. In particular, it appears the capacity constraint plays a complicated role. Some shipment sizes, which are big but not full truck loads, constitute a difficulty for carriers, who must find small shipments, or run their vehicles partially empty. As a consequence, depending on the difficulty to find small shipments, carriers may charge big shipments almost the price of truckload shipments. This tends to deter shippers from sending shipments of these sizes. Symmetrically, carriers may charge small shipments a very low price, because these shipments will be consolidated with other, bigger shipments. On the whole, market prices behave so as to deter shippers from choosing shipment sizes which decrease the loading factors of shippers. This explains why so many shipments are full truck loads.

Information availability plays an important role in this analysis. One qualitative property of the freight transport system is that information is not easily obtained. As a consequence, carriers have difficulties to consolidate shipments of which they ignore the existence. To represent this phenomenon, freight transport demand is assumed stochastic from the perspective of shippers. Quite intuitively, we find that the higher the variability of demand, the lower the average loading factors of carriers. Incidentally, this provides a strong rationale for the simultaneous existence of spot markets and long term contracts on the freight transport market. Indeed, shippers who can guarantee regular flows to carriers are able to negociate more competitive rates, because predictable flows mean lower costs for these carriers.

Finally, this approach brings a new argument to an old discussion about the market structure of the road freight transport market. One of the main statements according to which the road freight transport market is not under perfect competition is that perfect competition implies marginal cost pricing, which is inconsistent with the observed complex structure and variability of prices. Other papers indicate that the trucking technology is complex enough for complex price structures to appear under marginal cost pricing. This is particularly true here, where transport operations of shipment of distinct sizes are considered as distinct services, so that marginal cost pricing yields non-linear price schedules. As such, this study strengthens the position according to which complex price structure does not constitute a piece of evidence of market distortions.

Accounting for shipment size in freight transport modelling also requires a description of how it is determined by shippers. Some econometric models of choice of shipment size have been developed, in particular to explain modal choice. However, these models lack an underlying microeconomic sense.

According to the inventory theory, the choice of shipment size derives microeconomically from a trade off between transport costs, inventory costs (e.g. depreciation, capital opportunity cost, etc.), and other logistic costs (e.g. customer dissatisfaction), notably deriving from demand uncertainty, and notwithstanding any interaction between production and transport decisions. In order to investigate the microeconomic drivers underlying this decision, inventory theory models are analysed.

The most classic shipment size model, the Economic Order Quantity model, which is almost a century old, models the choice of shipment size as resulting from a trade off between transport costs and inventory costs. For a regular commodity flow between a given origin and destination, given a fixed transport cost per shipment, and a per ton commodity value of time, the EOQ model yields an optimal shipment size, increasing with the rate of the commodity flow and decreasing with the commodity value of time. It also yields a total cost, which increases less than proportionately with the commodity flow rate.

However, this model has not been estimated on a large scale of heterogeneous firms yet. Indeed, estimating this model requires a database not only describing shipments, by their sizes and unit values of time, and the way they have been carried, but also describing the shipper-receiver relationship, in particular the commodity flow rate. Shipment size databases are scarce, and, to our knowledge, the only shipment size database describing the shipper-receiver relationship is the French ECHO survey. The ECHO survey describes 10,000 shipments with many details. The size and value of shipments are observed, as well as the total annual commodity flow between the shipper and the receiver, which is not exactly the variable we are looking for, but which is a good proxy for a number of reasons. The estimation of the EOQ model gives satisfying results, which confirms the validity of this model on a large population of firms.

Another classic inventory model, the newsvendor problem, considers a flow of commodities from an origin where they are produced to a destination where they are sold. The demand at destination is stochastic, and there is a positive travel time between the origin and destination, so that the shipper must anticipate the demand. If the shipper has overestimated the demand, a certain amount of commodity will remain unsold at destination, with related inventory costs, whereas if the demand has been underestimated, some customers will have to wait before being delivered. An optimal logistic policy is easily derived in this classic inventory theory model. This policy is designed so that the expected level of the destination inventory is a trade off between inventory costs and customer dissatisfaction.

This model yields a total logistic cost, including customer costs, which depends, among other parameters, from the transport lead time and from the demand uncertainty. The derivative of this total logistic cost with respect to the transport lead time gives the value of time of the shipper with respect to freight transport. If the demand is highly variable, then a reduction in travel time is very valuable for the shipper, who can reduce both his inventory cost and the dissatisfaction of customers. In other words, a faster transport mode is prefered by shippers for the improved flexibility it allows. This provides some microeconomic signification to the concept of supply chain reactivity, and a step forward towards incorporating logistics into freight transport economics.

This model can be extended to explain why a given shipper would use two transport modes simultaneously for a given commodity flow, which is impossible for classic microeconomic mode choice models, where mode alternatives are generally assumed mutually exclusive. Operation research models of the use of two transport modes (usually a basic one and emergency one) are not uncommon. However, as usual in operations research, they yield optimal policies, but not the aggregate indices which are necessary to a microeconomic analysis (such as an average total cost and a average demand for each mode). In this study, a model has been developed which accounts for the use of two transport modes, a fast and expensive one and a slow and less onerous one, based on a basic heuristic logistic policy. This model explains that in some cases, the shipper takes advantage both from the low cost of the slow one and the reactivity allowed by the fast one.

On the whole, inventory theory makes it possible to model microeconomically the preferences of shippers with respect to freight transport, together with their logistic imperatives. It creates a clear linkage between freight transport demand and final consumer preferences. It should be noted, however, that inventory theory models are often difficult to use in a microeconomic framework.

Accounting for the choice of shipment size when modelling the freight transport system is a fruitful approach. It allows a more detailed analysis of the freight transport market, as well as of constraints of carriers and shippers, and of their preferences.

From a theoretical standpoint, taking into account shipments offers new insights on the structure of the freight transport market. This is true for carriers, whose technology can be analysed with much more details, and it is also true for shippers, whose logistic constraints can be represented explicitly – as such, this approach constitutes a first step towards the microeconomic modelling of logistics.

However, before shipments can be represented explicitly in spatialised freight transport models, a series of issues must be addressed. First, if shipments are represented, the consolidation process of shippers cannot be ignored. But it cannot be represented explicitly, the data and computational requirements would be prohibitive. A strategy must be designed to represent freight transport supply in a concise way, and to model how door-to-door shipment transport yields vehicle traffic. Second, the behaviour of shippers must be modelled. As an outcome of this study, the EOQ model constitutes a promising first step. However, the scope of this model is limited. Other decisions such as, for example, the number and locations of warehouses are much more complex. They are yet to be modelled satisfyingly.

Résumé

Ce travail a pour objectif d'améliorer la modélisation du transport de fret, par l'identification de ses caractéristiques qualitatives, par l'étude statistique des régularités qu'il présente, et par leur modélisation microéconomique. Il vise notamment à une meilleure compréhension du comportement des chargeurs et des transporteurs, et de leurs relations. Une attention toute particulière est accordée à l'examen du rôle de la logistique dans le système de transport de fret.

Actuellement, la modélisation du transport de passagers comme de marchandises repose majoritairement sur la représentation classique dite à quatre étapes des systèmes de transport. En ce qui concerne le transport de fret, cette représentation présente un certain nombre de lacunes majeures, qui ont été repérées et partiellement comblées par une série de récents progrès méthodologiques. Premièrement, la définition des origines et destinations des déplacements est ambiguë en transport de fret, notamment quands les opérations de transport impliquent des ruptures de charge, c'est-à-dire quand la marchandise est transbordée d'un véhicule à un autre au cours d'une même opération de transport. Deuxièmement, les décideurs, et leurs comportements microéconomiques sont mal identifiés. Troisièmement, la nature discrète des opérations de transport de fret est absente.

Une des pistes d'amélioration de la modélisation du transport de fret consiste à tenter de représenter explicitement les envois qui sont, d'un point de vue organisationnel, les atomes du système de transport de fret. Représenter explicitement les envois dans un modèle spatialisé de transport pose un certain nombre de difficultés, et offre des perspectives d'amélioration de la modélisation. Ce travail a pour objectif d'étudier ces difficultés et perspectives.

Afin de disposer d'une base de travail qualitative, et de faire face aux dégaurs de la représentation classique des systèmes de transport, nous procédons en premier lieu à une analyse systémique du transport de fret. Il s'agit d'identifier les décideurs, leurs options, préférences, et

les relations qu'ils lient les uns avec les autres, tout cela dans l'objectif d'accroître le réalisme économique des modèles de transport de fret. Cette analyse systémique se fonde sur une étude qualitative du transport de fret, ainsi que sur une revue des progrès récents de la modélisation du transport de fret.

Le résultat de cette analyse est une représentation du système de transport de fret sous forme de couches superposées, représentant diverses catégories de décisions liées à la formation des flux de marchandises et de véhicules. Au centre de cette représentation, on place ce qui est effectivement produit per le système de transport de fret, à savoir les opérations de transport porte-à-porte d'envois de marchandises. Subséquemment, on définit les transporteurs comme étant les producteurs de ces opérations de transport, et les chargeurs comme en étant les consommateurs. Avant de décrire plus précisément le comportement des chargeurs et des transporteurs, il faut noter que les enquêtes classiques d'observation du transport de fret, qui décrivent généralement les mouvements opérés par les véhicules, ne permettent pas de reconstituer les opérations de transport porte-à-porte des envois, dès qu'il y a une rupture de charge.

Les transporteurs, en tant que producteurs, obéissent à une logique microéconomique classique. Leur objectif est de produire les opérations de transport qu'ils se sont engagés à effectuer avec les ressources dont ils disposent (par exemple les véhicules, les personnels, le carburant, les plateformes de groupage-dégroupage, etc.), et ce au moindre coût. Pour ce faire, ils combinent des opérations de transport élémentaires (chargement par un véhicule d'une certaine quantité de marchandises à un endroit donné, et déchargement ailleurs) pour produire des opérations de transport porte-à-porte. Les technologies de transport sont fortement caractérisées pas la présence d'économies d'échelle (dues, par exemple, à la capacité fixe des véhicules), et d'envergure (provenant du fait qu'il peut être plus facile de transporter ensemble, sur une certaine partie de leurs parcours, des envois provenant de et allant à des lieux distincts). De ce fait, les organisations des transporteurs impliquent généralement l'usage de véhicules et modes variés, et le recours courant aux opérations de massification et d'éclatement. Cela a pour conséquence notable que le lien entre mouvements des véhicules et mouvements des marchandises est complexe. Nous donnons à ce lien une importance particulière, en le définissant comme la "logistique des transporteurs". Le problème de la structure des marchés de transport de fret n'est pas abordé.

La logique microéconomique du comportement des chargeurs est plus complexe. Les préférences des chargeurs vis-à-vis des possibilités de transport sont fortement déterminées par leurs propres impératifs logistiques, qu'il est donc utile d'identifier et de comprendre. Essentiellement, ces impératifs logistiques sont fondés sur les préférences des clients des chargeurs vis-à-vis de la disponibilité des produits que ces clients achètent. En effet, un bien ne produit de l'utilité au client qui l'achète que si ce bien est à la disposition immédiate de ce client. En général, en plus de la dépense correspondante, un client achetant un bien doit fournir un certain effort non monétaire; cet effort peur consister en un déplacement vers le point de vente corrrespondant, un temps d'attente de livraison, un risque de rupture de stock, etc. Il s'agit d'un coût usager, similaire en nature à la composante temporelle dans le coût généralisé en transport de passagers. Le chargeur peut diminuer l'effort de ses clients, par exemple en multipliant le nombre des points de vente, en diminuant les temps de livraison grâce à l'utilisation de modes de transport plus rapides, en augmentant les stocks de sécurité, etc. Cela vient à un coût, et le chargeur doit donc trouver le compromis idéal entre ses propres coûts logistiques et les efforts que doivent fournir ses clients, puis organiser ses flux de biens en conséquence. Ce processus constitue le lien complexe entre les flux production-consommation des tableaux entrée-sortie des économies et les opérations de transport porte-à-porte. Nous le définissons comme la "logistique des transporteurs".

Dans les modèles classiques de transport de fret, l'offre de transport est généralement représentée par des réseaux d'arcs et de noeuds caractérisés par des temps de trajet, des coûts, et des modes de transport. Il n'est pas facile de prendre en compte la taille d'envoi dans cette représentation. Une approche microéconomique est proposée, afin de fournir quelques éclairages sur cette question.

Dans le cas du transport routier, entre une origine et une destination, le lien entre la taille d'un envoi et le prix de l'opération de transport correspondante n'est pas simple ; en particulier, il n'est pas linéaire. En effet, les grilles tarifaires de transport de fret, qui donnent les prix de transport en fonction de la taille des envois, ne tendent pas vers zéro pour les petits envois, et tendent à s'aplatir pour les tailles d'envoi proches de la capacité des véhicules. Notre objectif est de proposer une explication microéconomique de ces propriétés. Plus précisément, on tentera de voir comment elles résultent de la technologie de transport et des coûts des ressources des transporteurs. On suppose que les transporteurs routiers sont en concurrence parfaite.

L'avantage du transport routier du point de vue de cette approche est que la technologie du transport routier est plus simple que celle des autres modes de transport de marchandise. On suppose qu'une opération de transport d'un envoi d'une origine à une destination consiste en une opération d'enlèvement dans la zone d'origine, avec un déplacement d'accès et une opération de chargement, puis une opération de traction pour rejoindre la zone de destination depuis la zone d'origine, enfin une opération d'accès et de déchargement dans la zone de destination. On suppose que ces opérations sont effectuées avec des véhicules de capacité limitée. Si plusieurs envois sont chargés dans une même zone d'origine et transportés dans une même zone de destination, on suppose que les coûts d'accès, de chargement et de déchargement sont spécifiques à chaque envoi, tandis que le coût de traction est commun aux envois transportés dans un même véhicule. Les autres contraintes spatiales, telles que le retour à vide, et l'enchaînement de déplacement, sont laissées de côté. Les tailles des envois sont endogènes.

Une difficulté majeure se pose quand, à partir de la description de la technologie ci-dessus, on essaie d'écrire la fonction de coût. En effet, celle-ci dépend du taux de chargement, c'est-à-dire de la fraction de la capacité des véhicules occupée par de la marchandise. Les chargeurs optimisent ce taux de chargement, afin de réduire leurs coûts. Or ce processus de minimisation, connu en recherche opérationnelle comme le problème de bin-packing, est NP-difficile. Pour surmonter cette difficulté, nous posons une hypothèse simpliste: les envois ne peuvent prendre des tailles que de 1, 2 ou 3 unités, avec une capacité des véhicules de 3. Il est alors possible d'obtenir une fonction de coût explicite, donc des coûts marginaux et des prix de marché.

Malgré cette simplification, cette approche permet d'expliquer les propriétés qualitatives des grilles tarifaires de transport de fret. En particulier, le rôle de la contrainte de capacité est mis en lumière. Ce rôle est assez complexe. Il semble que certains envois de grande taille mais ne remplissant pas les véhicules sont compliqués pour les transporteurs qui doivent trouver de petits envois pour compléter les véhicules, ou bien faire rouler ceux-ci partiellement vides. En conséquence, en fonction de l'abondance relative des petits envois, les transporteurs peuvent faire payer, pour les envois de grande taille, un prix proche du prix demandé pour les camions complets. De fait, cela dissuade les chargeurs d'utiliser ces tailles d'envoi, ils se reportent soit vers des envois plus petits, soit vers des camions complets. Ceci explique notamment pourquoi autant d'envois ont exactement la taille des véhicules.

La qualité de l'information que les différents agents ont les uns sur les autres joue un rôle fondamental dans cette approche. Une propriété forte du système de transport de fret est qu'il est assez difficile pour les chargeurs d'obtenir de l'information sur les coûts, et il est également difficile pour les transporteurs de prévoir la demande. Or, moins les transporteurs ont d'information sur les envois qu'ils devront transporter, moins il leur est facile d'en organiser le transport, et en particulier, le cas échéant, la massification. Pour représenter ce phénomène on suppose dans notre approche que les transporteurs font face à une demande stochastique, caractérisée par une certaine variabilité. On observe effectivement que plus la variabilité de la demande est grande, plus le taux moyen de chargement des transporteurs est bas. Incidemment, cela constitue une explication satisfaisante de l'existence simultanée de marchés spots et de contrats de long terme sur le marché de transport de fret. En effet, un chargeur capable de garantir des flux réguliers à un transporteur peut, en échange, négocier des tarifs intéressants, car des flux prévisibles impliquent des coûts moindres pour les transporteurs ; ce n'est pas le cas d'un chargeur qui n'a pas lui-même les moyens de prévoir sa demande de transport.

Il faut également noter que cette appproche apporte un nouvel élément à un problème déjà ancien d'économie des transports, celui de la structure de marché du transport routier de marchandises. Sur cette question, deux points de vue s'opposent. Le premier consiste à dire que le transport routier de marchandises n'est pas parfaitement concurrentiel, car la concurrence parfaite implique la tarification au coût marginal, avec laquelle la variabilité des prix observés sur les différents marchés de transport de fret est manifestement incompatible. A ceci, le second point de vue répond que la technologie de transport routier est plus complexe qu'il n'y paraît, et que c'est ce que les prix reflètent. Leur variabilité ne permet donc pas de présumer de la présence d'économies d'échelle. Dans notre approche, la prise en compte des tailles d'envoi et la description détaillée de la technologie de transport routier permet de montrer que de la tarification au coût marginal résultent bien des grilles tarifaires complexe. En ce sens elle tend à confirmer le second point de vue.

Pour prendre en compte la taille des envois dans la modélisation du transport de fret, il faut également décrire la manière dont les chargeurs la déterminent. Certains modèles économétriques de choix de type d'envoi existent, mais ils manquent de fondations microéconomiques.

Selon la théorie de l'inventaire, le choix de type d'envoi résulte d'un compromis microéconomique entre les coûts de transport, les coûts d'inventaire (par exemple : la dépréciation, le coût d'opportunité du capital, etc.), et d'autres coûts logistiques (tels que l'instarisfaction des clients vis-à-vis de la disponibilité des produits), dus notamment à la variabilité de la demande. De manière à mieux comprendre les décisions prises par les chargeurs, les modèles d'inventaire sont étudiés.

Le modèle d'inventaire le plus classique, âgé de presque un siècle, est le modèle Economic Order Quantity. Selon ce modèle, la taille d'envoi choise par le chargeur résulte d'un compromis entre les coûts de transport et les coûts d'inventaire. Plus précisément, pour un flux régulier de marchandises entre une origine et une destination, pour un coût de transport fixe par envoi, et pour une valeur du temps par tonne de la marchandise, le modèle EOQ propose une taille d'envoi optimale, qui augmente avec le débit du flux et qui décroît avec la valeur du temps. Ce modèle donne également un coût total, qui croît moins que proportionnellement avec le débit du flux de marchandises.

Cependant, la validité empirique de ce modèle sur une grande population de chargeurs hétérogènes n'a pas été testée jusqu'à présent. Ceci s'explique par la difficulté d'obtention des données requises pour l'estimation d'un tel modèle. Pour ce faire, il faut une base de données décrivant non seulement des envois par leurs tailles, et valeurs du temps, et la manière dont ils ont été transportés, mais aussi une description de la relation expéditeur-destinataire, notamment du débit du flux de marchandises. Les bases de données d'envois sont rares, et, à notre connaissance, la seule base contenant des variables proches de celles requises pour l'estimation du modèle EOQ est la base française ECHO. Cette base décrit 10000 envois avec beaucoup de détails. En particulier, la valeur des envois, et le flux total de marchandises (tous types confondus) échangés entre l'expéditeur et le destinataire sont mesurés ; ils constituent des proxys acceptables des variables nécessaires. L'estimation du modèle EOQ montre qu'il est assez performant, ce qui conforme la validité empirique de ce modèle pour décrire une population de chargeurs.

Un autre modèle classique de la théorie d'inventaire est le modèle du vendeur de journaux. Ce modèle concerne un flux de biens produits à un endroit donné puis transportés à un autre endroit où ils sont vendus. La demande à destination est stochastique, et il y a un temps de transport positif entre l'origine et la destination, ce pourquoi le chargeur doit anticiper la demande. S'il l'a surestimée, il lui reste une certaine quantité d'invendus. Ils pourront être vendus plus tard, mais cela engendre un coût d'inventaire. Par contre, si la demande a été sous-estimée, certains clients devront attendre avant d'être servis. Il existe une politique de production optimale, s'adaptant à la demande. Cette politique est conçue de manière à ce que le niveau moyen de l'inventaire à destination soit un compromis entre les coûts d'inventaire et les coûts d'attente des clients.

Ce modèle permet également de calculer le coût total logistique optimum, incluant les coûts des clients. Ce coût total dépend, entre autres paramètres, du temps de transport et de la variabilité de la demande. La dérivée de ce coût total par rapport au temps de transport donne la valeur du temps du chargeur. Cette valeur du temps est plus élevée quand la demande est plus variable, car dans ce cas, une réduction du temps de transport a des avantages plus grands en réduction des coûts d'inventaire et des coûts supportés par les clients. En d'autres termes, un mode de transport plus rapide permet à un chargeur d'améliorer la flexibilité de sa chaîne d'approvisionnement. En tant que tel, ce modèle donne donc une interprétation microéconomique du concept de réactivité de la chaîne d'approvisionnement, et permet, dans une certaine mesure, de faire un premier lien entre logistique et transport de fret, dans un cadre microéconomique. De plus, ce modèle peut être amélioré pour étudier la valeur de la régularité des temps de transport.

Ce modèle peut également être étendu pour tenter d'expliquer pourquoi un chargeur peut, dans certains cas, utiliser deux modes de transport en parallèle pour un flux de marchandises donné. Ce phénomène, observé dans certains cas, est inexplicable par les modèles désagrégés classiques de choix modal. Un certain nombre de modèles de recherche opérationnelle représentent l'usage simultané de deux modes de transport. Cependant, si ce type de modèles propose généralement une marche à suivre optimale, il est difficile d'en tirer des indicateurs agrégés nécessaires à l'analyse microéconomique, tels que les coûts moyens, ou la demande moyenne pour chaque mode de transport. Un modèle a donc été développé pour représenter l'usage de deux modes en parallèle, un rapide et onéreux, l'autre plus lent et moins cher. Ce modèle repose sur une heuristique très simple, certainement pas optimale, mais, dans certains cas, plus efficace que l'utilisation du meilleur des deux modes seul. Dans ces cas, le chargeur profite à la fois, dans une certaine mesure, du faible coût du mode lent, et de la flexibilité qu'offre le mode rapide. Cette approche tend toutefois à confirmer le fait observé que peu d'entreprises utilisent deux modes en parallèle pour un flux de marchandises donné.

Dans l'ensemble, la théorie d'inventaire permet la modélisation microéconomique des préférences des chargeurs en ce qui concernt le transport de fret, en lien avec leurs impératifs logistiques. Cela crée un lien explicite entre le fonctionnement du système de transport de fret et les préférences des consommateurs finaux. Il faut cependant noter que les modèles d'inventaire issus de la recherche opérationnelle sont généralement difficiles à exploiter dans le cadre d'une analyse microéconomique.

Tenter de représenter le choix de la taille d'nevoi dans la odélisation du transport de fret s'avère être une approche fertile. Elle permet une analyse plus détaillée du marché du transport de fret, ainsi que du comportement des chargeurs et des transporteurs.

Cette approche permet des avancées théoriques sur la structure du marché du transport de fret, que ce soit pour les transporteurs, dans la mesure où elle permet une analyse plus détaillée de leurs technologies, et donc de leurs coûts, et pour les chargeurs, dont les contraintes logistiques peuvent être, dans une certaine mesure, représentés explicitement. Elle constitue notamment un premier pas vers la modélisation microéconomique de la logistique.

Cependant, avant que la taille d'envoi soit explicitement présente dans les modèles spatialisés de transport de fret, un certain nombre de problèmes doivent être résolus. Premièrement, il n'est pas possible de représenter les envois sans représenter les processus de massification. Or ces processus de massification ne peuvent pas être représentés en détail, cela exigerait beaucoup trop de données et de temps de calcul. Il faut donc trouver le moyen de les représenter de manière résumée, peut-être par le biais d'une approche statistique du lien entre coûts des ressources des transporteurs et prix de marché du transport en fonction de la taille d'envoi. Ceci implique également de modéliser le lien entre les opérations de transport d'envois porte-à-porte et les mouvements effectivement réalisés par les véhicules. Deuxièmement, il faut modéliser le comportement des chargeurs. De ce point de vue, le modèle EOQ semble constituer un bon candidat dans un premier temps. Mais la portée du modèle EOQ est limitée. De nombreuses décisions de nature logistique lui échappent totalement, telles que la localisation et le nombre d'entrepôts. Introduire ce type de décisions dans un modèle spatialisé de transport de fret est certainement une tâche ardue.

Introduction

This work is focused on the freight transport system. It is aimed at understanding the behaviour of both carriers and shippers, and their relationships. Particular attention is paid to how the logistics of shippers determine their demand for freight transport. Qualitative analysis is a major component of this work; its role is to identify facts and regularities which characterise the freight transport system.

The purpose of this work is also to improve freight transport modelling. Freight transport modeling requires that the freight transport system be described in a concise, mathematical way, which is, as much as possible, microeconomically consistent. Under these conditions, freight transport models can be useful as decision support tools[1]. Therefore, we aim at developing microeconomic models which capture the regularities of the freight transport system, including the linkage between logistic and freight transport demand, at the center of which is the joint decision of shipment size and transport mode.

Hopefully, these microeconomic models will constitute possible ways to improve freight transport modelling, in particular if they can be included in spatialised simulation models, which currently only account for the logistic dimension of freight transport demand.

0.0.0.1 The situation of spatialised freight transport models

Freight transport is not a recent problematic of transport economics and policy. Some questions, in particular pertaining to technical regulation and market structure, but not only, have been discussed for decades, and have motivated many academic works; but few of these works actually required to take into account explicitly the topographic conformation of freight transport networks. The current need for spatialised freight transport demand models is mainly motivated by environmental, economic and social issues which, if they are now considered crucial, were

[1] On the usefulness of transport models, and, more generally, of cost-benefit analysis as decision support tools, see Maurice and Crozet (2007).

pretty much ignored a few decades ago. Example of these problems are global warming, rarefaction of fossil fuels, discomfort caused by trucks to car drivers, a renewed interest into alternative transport modes such as railroad transport (Académie des Technologies, 2009). As such, this need is relatively new, and much more recent than the need for spatialised passenger transport models.

Quite naturally, the first attempts to model spatially the freight transport system have been directly inspired from the methodologies which have been continuously developed and applied over many years to model passenger transport. Indeed, at first sight, the freight transport system seems to be correctly represented by a four stage representation almost similar to the passenger transport four stage representation. Household location corresponds to firm location, destination choice is correctly replaced by supplier or customer choice, mode choice and route choice keep relevant. Similarly, the microeconomic theory underlying passenger transport models, including the representations of the preferences of the agents and of the transport alternatives, applies more or less smoothly to freight transport. Finally, data is available to observe freight transport decisions at each of these stages: in many countries, there are databases describing production and consumption of commodities by regions, trade matrices, modal shares, and traffic.

However, spatialised passenger transport models soon proved to fit only imperfectly to the freight transport system. Furthermore, while efforts were made to try to adapt these models to freight transport, it appeared always more clearly that applying the four stages representation to the freight transport system was not that natural. Indeed, it leaves a number of question unanswered, some of which are briefly described below.

- *The definition of the origin and destination of commodity trips is ambiguous.* For example, in the case where commodities are handled through a port to be transshipped, the port is clearly neither an origin nor a destination from the perspective of the shipper (*i.e.* the consumer of the transport operation), whereas it can be the destination of a motor carrier, and the origin for a shipowner. The same ambiguity appears in urban transport, where the vehicles often make rounds during which they proceed to many pickups and deliveries.

- *Decision makers are not explicitly identified.* In the case of passenger transport, supply and demand are clearly distinguished. Supply is represented by the exogenous[2] characteristics of transport alternatives, while demand makes a series of decisions which are explicitly represented. In the case of freight transport, shippers buy door-to-door transport operations with given characteristics, but without necessarily paying the same attention as carriers to how transport operations are actually realised. The four stages representation, by aggregating all commodity flows of a given nature on each origin destination pair, as is usually done, ignores this distinction.

- *The microeconomic description of the preferences of agents is weak.* As a direct consequence of the previous point, the microeconomic principles underlying the preferences of the shippers and the carriers are not clear. In particular, there is no direct linkage between the preferences of the shippers and their logistical imperatives, which are, either way, not defined.

- *The discrete nature of freight transport operations is absent.* Except for some rare cases, such as pipeline transport, transport operations are of discrete nature: a commodity flow is not transported in a continuous, seamless manner, but as a series of shipments, of given sizes and characteristics, which depend as much on the logistic imperatives of shippers as on the characteristics of the available transport options.

This list, not exhaustive, is sufficient to indicate that many crucial features of the freight transport system are either imperfectly represented by, or absent from, the four stages representation. This suggests that this representation may lack one or more important mechanisms constituent of the freight transport system.

0.0.0.2 Problem statement

One option for improving spatialised freight transport modelling, which has the advantage to address some of the points listed above, is to enhance the classic four stages representation by including a decision specific to freight transport: the choice of shipment size. The purpose of this work is thus to investigate *the difficulties and potential benefits of representing explicitly the choice of shipment size in spatialised freight transport models.* This wide field will not be fully addressed; this work is rather focused on some of the questions it raises. These questions are detailed below.

[2]Except for the case where congestion phenomena are taken into account.

Selected issues The choice of shipment size by a given firm optimising its logistics is a long known problem of inventory theory. It has also drawn some interest from the perspective of freight transport modelling, so that shipment databases and microeconomic optimal shipment size models have been available for quite some time. Nevertheless, there are currently no fully-fledged operational spatialised freight transport models in which shipment sizes are explicitly represented. This can be explained by the many issues raised by the introduction of a new decision stage in a model on a microeconomic basis. Some of them are hereby listed:

- *A systemic representation of the place and role of the shipment size decision.* Before aiming at a formal model of the choice of shipment size, one should determine the role of the shipment size decision in the freight transport system, how it results from the preferences and constraints of the freight transport supply and demand, and how it interacts with other decisions, such as mode and itinerary choice, but not only. Without such a representation, designing a model is difficult, except maybe in the case of a purely statistical approach.

- *A widely valid optimal shipment size model.* Many models of optimal shipment size have been developed in the inventory theory. They are generally based on relatively limiting hypotheses, adapted to some firms but not to others, so that their relevancy for a large population of heterogeneous firms is questioned, and should be econometrically assessed. Subsequently, relevant databases are necessary.

- *A reasonable complexity.* It is obviously impossible to design a numerical model where all the shipments sent and carried are explicitly represented on an accurate, individual basis. In particular, representing the consolidation process, by which carriers optimise the use of their fleets, is out of reach of a spatialised freight transport model, because of both the computing and data requirements. Strategies must be designed to keep the description of the freight transport system concise.

This list is certainly far from complete, and merely illustrates the complexity of designing a spatialised model of freight transport.

Selected perspectives However, this approach is also endowed with opportunities. Some of them, of theoretical nature, are presented below:

- *Logistic-based modelling of shipper preferences* The inventory theory contains many models of optimal shipment size, mainly built using operations research tools. Some of these models, which are simple enough (for example when the cost or input demands at the optimum are given by a closed formula), can be used as a basis for a microeconomic analysis. This has been done in some cases, notably to derive microeconomic mode choice models based on an explicit representation of the logistic constraints of shippers. Some of these models insist on the role of the structure of freight rates, others on the influence of the variability of the demand, etc. The explicit representation of shipments is a fruitful way to investigate the microeconomics of freight transport demand, and to link it to logistic issues.

- *Simultaneous use of two modes for a given commodity flow* In the classic four stages representation, as well as in most models of joint choice of shipment size and transport mode, all shippers theoretically use only one transport mode, the optimal one, for a given commodity flow. However, some shippers have been observed to use two transport modes simultaneously for a single commodity flow. Some inventory models describe such logistic protocols. These models can be adapted to design microeconomic models for the simultaneous use of two modes, thus providing a deeper insight into the drivers of the demand for freight transport.

These issues and opportunities are studied in the sequel of this work, thus providing insights on the freight transport system, and on how the shipment size variable can be included in spatialised freight transport models. The objective is to provide relatively simple models which capture their specificities and help to understand their regularities. Hopefully, this work also yields some partial answers concerning how should be designed a spatialised freight transport demand model encompassing an explicit representation of the shipment size.

0.0.0.3 Methodology

The questions which have just been raised can be addressed from various standpoints. In this study, four approaches are used.

As expected, the first one is the bibliographic review; extensive reviews, mainly of academic works, but not only, constitute the first stage of this study. Second, this work is based on systemic analysis, which yields the qualitative description of the freight transport system which is indispensable to further formal analysis. Third, statistics, understood both as data collection and econometric analysis, are called on both to suggest a better observation of the freight transport system and to discuss the empirical validity of some theoretical models. Fourth, given that numerical models are exclusively based on mathematical formulations, some qualitative properties of the freight transport system are represented in mathematical models, based on common microeconomic principles.

0.0.0.4 Outline

This study consists of seven chapters. These chapters are grouped into three parts, which present respectively bibliographic reviews, systemic and metrologic analyses, and microeconomic models.

Part 1: Framework and bibliography This study begins with a wide review of the state of the art of spatialised freight transport modelling, as well as from an analysis of the role of logistics in the freight transport system.

Chapter 1, *Advances in freight transport demand modelling: a microeconomic perspective*, reviews the recent advances in spatialised freight transport modelling. This is not a review of the models currently in use, but rather of the recent modelling methodological advances which are specific to freight transport. Some of these advances consist in enhancing the four stages representation, in order to represent a feature of the freight transport system which is absent from classic models. Others are fundamentally new approaches, based on dedicated model architectures (this is the case, in particular, of a urban freight transport model). This review identifies the state of the art of spatialised freight transport modelling, and also indicates quite clearly that the four stages representation is imperfectly adapted to the freight transport system.

Chapter 2, *Logistic problematics: an analysis of the determinants of freight transportation demand*, is aimed at investigating the role of logistics in the demand for freight transport. To do so, a large range of academic studies pertaining to logistic problems have been examined. This is notably the occasion to discuss the definition of logistics, and to examine which kind of logistic problems are addressed in the academic literature. The approaches reviewed are very heterogeneous. They have been classified according to their perimeters (a single firm, a set of vertically linked firms, a market, a territory); specific attention has been brought to microeconomic models of choice of shipment size. This review notably shows that it is difficult to tell whether any of these works can be used in large scale models; their validity for large, heterogeneous populations of firms is not confirmed.

Part 2: Systemic analysis and metrology The second part addresses directly the limitations of the four stages representation of the freight transport system, by proposing a novel systemic representation of the freight transport system. Some metrologic and econometric results are also presented.

Chapter 3, *A systemic representation of the freight transportation system* discusses precisely the limitations of the four stages representation of the freight transport system. Once these limitations are identified, a new systemic representation describing the agents implied in the freight transport system, their options, preferences, and interactions, is presented. This systemic analysis insists on the pivotal place of shipments in the freight transport system, both from the supply and the demand perspectives. It also clarifies the logistic objectives of shippers on one hand, and the technical constraints of carriers on the other hand. This systemic representation consists of superimposed layers standing for the hierarchy of freight transport decisions, and, as such, should hopefully constitute an appropriate basis for freight transport modelling.

Chapter 4, *Measuring the motor carriers activity: improvement of a road-side survey protocol*, is focused on improving the observation and knowledge of the road freight transport technology. This chapter presents some propositions on how a classic french road-side survey protocol, usually employed to observe the origin and destination of vehicles taking a given road, can be improved to yield a more accurate knowledge of road freight transport. New questions have been added to the classic questionnaire; the enhanced questionnaire has then been tested in the frame of two surveys. These new questions concern currently unobserved variables such as the volume occupied in vehicles, or the organisations (double crew, relays) used by motor carriers, the existence of specific logistic imperatives, or the breaks drivers must take.

Chapter 5, *A note on the econometric validity of the EOQ model*, presents succinctly how the French shipment database ECHO can be used to assess econometrically microeconomic models of choice of shipment size based on inventory theory. A useful particularity of the ECHO database is that it contains variables which are seldom observed, yet crucial for the estimation of even the simplest inventory based models of choice of shipment size. This database is thus used to estimate the simple Economic Order Quantity model, well known in the literature and presented in Chapter 2. An extended specification is also estimated, to measure the influence of variables which are external to the basic model.

Part 3: Microeconomic analysis Taking into account the shipment size offers new possibilities to design microeconomic models of shippers and carriers. This part investigates some of these possibilities.

Chapter 6, *On shipment size and freight rates: technical constraints and equilibrium price schedules*, addresses the issue of freight rates. Indeed, taking into account shipment size immediately raises the question of how freight rates depend on this variable. The linkage between shipment size and freight rates is not trivial. In particular, are distinguished the influences of distance-dependent costs (such as fuel and drivers), of distance-independent costs (such as loading, unloading, or transshipment costs), and of the capacity constraint of vehicles. These influences are analysed in the frame of a most simplified partial equilibrium model, where shippers decide the sizes of the shipment they send, and carriers their rates. The model demonstrates the possibility to model jointly the logistic decisions of shippers and of carriers, and the way the vehicle capacity constraint influences shippers through equilibrium prices. It provides new insights on the structure of costs in the road freight transport market. It also reveals the complexity of representing everything explicitly; some simplifications that could be used in a large-scale model are suggested.

Chapter 7, *Logistic imperatives and modal choice*, adapts a classical model of the inventory theory with two objectives. First, a logistic interpretation of the preference of shippers for shorter travel times is provided. Indeed, a positive travel time means shippers must anticipate the demands of their own customers before knowing their exact amount. The longer the travel time, the more difficult it is for them do to so precisely. This lack of precision implies a series of issues, in other words, costs. The concept of supply chain's reactivity is thus represented formally, and explicitly related to shippers' preferences for faster freight transport. Second, this model is adapted to explain microeconomically why a given shipper would use two transport modes simultaneously on a given supply chain, something which is inconsistent with classic mode choice models. To keep the results analytically tractable, some rather strong hypotheses are made. However, it is possible to show that two modes can be used together by a shipper when the reactivity allowed by a fast transport mode and the transport cost savings allowed by a slow transport mode are complementary.

Part I

Framework and bibliography

Chapter 1

Advances in freight transport demand modelling: a microeconomic perspective

1.1 Introduction

1.1.1 Background and objectives

From a political point of view, freight transport is a complicated sector of the economy to tackle with. It is indeed closely interdependent with both the whole industry, which requires goods to be taken from places to others under a set of tight constraints, and people's everyday life, their need for transport capacity, and their reluctance to bear the negative impacts of freight traffic. This sector is also subject to a number of market distortions and externalities (most notably, scale economies, congestion and environmental impacts), which implies that the market has little chance to reach a socially optimal situation by itself - meaning that regulation is in order. Such a regulation is all the more efficient as it is backed by a thorough understanding of, first, the operation of the freight transport sector; second, the potential influence of a range of regulation policies. Freight transport modelling is aimed to contribute to that understanding, by providing firstly a systemic or economic representation of the freight transport sector, secondly tools to estimate and forecast quantitative indicators of its activity.

The methods applied in the field of freight transport modelling have been, up to recent times, largely inspired from those designed for and used in passenger transport modelling. This is particularly true with respect to the models architectures and the underlying economic and behavioural hypotheses. Their adaptation to the particularities of freight transport has thus been deemed necessary, and has been undertaken by several research teams throughout the world.

A number of papers were published recently to contribute to this adaptation effort. This motivated some review papers: notably that by Regan and Garrido (2002), who lists a series of approaches and theoretical models for freight transport and discusses some issues that remain to be addressed, particularly with respect to shippers' behaviours. Another state-of-the-art review was undertaken by ME&P and WSP (2002), in a diagnosis or prospective perspective. Let us also mention some more specific reviews such as those by Ambrosini and Routhier (2004) and Russo and Comi (2004), which are focused on modelling freight transport in an urban environment. These reviews provide firm ground for understanding how freight transport may be modelled, and which methods are used in operational environments.

Although the quoted reviews may seem quite recent to the reader, innovation in freight transport modelling has kept a fast pace since, which motivated our own effort to provide an up-to-date review. Moreover we shall take an economic perspective on the modelling advances, and highlight those which address issues specific to freight transport in its supply or demand component.

The paper proceeds in three steps that address in turn each of the following three objectives: first, the presentation of some recent, advanced freight transport demand models; then, the analysis and logical organisation of these advanced models; lastly, the identification of some fields that remain to be investigated.

1.1.2 Scope and structure

Our review is focused mainly on spatialised freight transport models. We shall consider the model for freight concentration and intermodal transport designed by Groothedde (2003), the strategic railway simulation model in NEMO and its interaction with the microscopic simulation model RailSys, the NODUS model, the simulation models designed in the frame of the European projects EUNET2.0 and SCENES, the urban freight transport model FRETURB, the ECHO French shipment survey, and finally the REDEFINE European project. Each model will be described in terms of principles and behavioural assumptions and of logical architecture, before we focus on its innovation. We will also compare the methodological choices to the claimed objectives.

As we also aim to identify the economic issues and specifically the demand-related issues, we shall assess the modelling developments in an economic perspective along the following items:

- The representation of demand in terms of volume, behaviour and agent heterogeneity.
- The representation of supply in terms of networks and layers, transport and logistic features, macroscopic relationships including scale economies and congestion effects.
- Relationships between agents and external impacts.
- Policy assessment.

We shall establish a typology to identify and classify the advances, distinguishing both the approach and the scope of the modelling. The typology allows us to suggest potential linkages between the models under review. Our conclusion is devoted to the identification of some under-investigated issues.

1.2 A review of advanced models

A number of works were selected from a large set of recent contributions, on the basis of their innovative characteristics in the field of freight transport modelling. Although they vary widely in both scope and approach, we chose to bring together those works which deal with similar subjects.

1.2.1 Refined modelling of service supply in railway transport

The representation of costs, and particularly of congestion costs (or, equivalently, of capacity limitations) is often over-simplified. This is particularly true with respect to the non-road modes, of which the operating processes make it difficult to model the capacities. We present here a strategy which has been designed to provide an accurate representation of supply costs in railway transport, within a framework which also considers passenger transport.

The Institute for Building and Operating Railway of the University of Hannover has developed an architecture combining two railway network models (Kettner and Sewcyk, 2002; Kettner et al., 2003; Sewcyk et al., 2007). The first one, NEMO (Network Evaluation Model), is a macroscopic, strategic simulation model developed since 1999 at the

IVE (Institute of Transport, Railway Construction and Operation), in collaboration with the ÖBB (Austrian Railways). It is based on a macroscopic network containing the access point of passengers and freight, and the junctions, intersections, and marshalling yards for freight transport. The demand for passenger transport is described as OD flows per time slice, on the basis of an average day according to the transport. The demand for freight transport is described as flows per commodity group. Both demands are inputs to the model. The transport supply consists, for the passenger transport, of a set of services, delivered by trains of given characteristics operated on given lines and serving given sets of stations. Passengers are assigned on these services. The freight demand is handled in a somewhat different way: block trains are operated when possible, whereas the residual freight is carried in single wagons, which follow routes between marshalling yards, according to a given routing protocol. Optimal empty wagons flows are computed using the DISPO software, also developed by IVE, to address potential imbalances. Apart from the block trains, all services are determined i.e. specified as exogenous. The demand, both freight and passengers, is assigned on this set of services. The volume of trains required on each line is then determined, so as to meet the demand on the most heavily loaded section between two neighbouring stations. Thus the number of trains required for each service is endogenous. Consequently the model yields the network loads due to freight and passenger transport. NEMO enables its user to analyse such issues, for example in order to evaluate the economic efficiency of the infrastructure.

RailSys is a microscopic operational simulation model, developed since the 1980s by IVE, mainly for the German Federal Railways. In RailSys, the infrastructure is represented as a highly detailed network, taking into account physical characteristics (radii, gradients, etc.), the signalling system (overlaps, release contacts, etc.), and the operation process (prioritisation strategies and railway operation process). A detailed database is thus necessary for the functioning of RailSys. RailSys then calculates, with respect to the trains characteristics, the running times and minimum headways, using a point mass system and the potential safety running time margins. Using all these data, and a defined train operation, RailSys simulates timetables describing accurately the movements of all trains.

NEMO and RailSys have been interfaced respectively as client and server. The first reason for integration is that it allows both models to rely on a unique database, which ensures that they share consistent data. The second reason is that the outputs of the NEMO model are based on extremely accurate data. The model integration proceeds as follows: first, the microscopic network representation of the infrastructure contained in RailSys is used to automatically generate a macroscopic network in NEMO (Figure 1.1, taken from Kettner et al., 2003). RailSys also transfers the timetables, from which NEMO defines available services and the number of trains for each service. Then, NEMO compares the demand to the supply. If there are more trains than needed, the superfluous trains are removed. On the contrary, if the demand exceeds the supply, then NEMO sends RailSys a request to add a train in the timetable. RailSys then searches for an additional train path. Passenger trains are processed first, then the additional freight trains. It may be impossible to find a path. Finally, capacity limitations may be investigated. We will not get into the detail of this evaluation, which consists roughly in determining the ratio of time during which the infrastructure is occupied. A high ratio indicates a potential bottleneck.

The interface built between NEMO and Railway has two main advantages. First, it allows the use of already existing, accurate data, thus saving the cost of further data collection, and providing a representation of the infrastructure based on highly realistic data. The representation of capacity problems is all the more accurate (although the capacity limitation identification criterion may be discussed). Second, it is very interesting to note that this architecture allows to measure quantitatively, at a strategic, macroscopic level, the impact of a change in operating modes (such as the signalisation, speed limits, or priority rules).

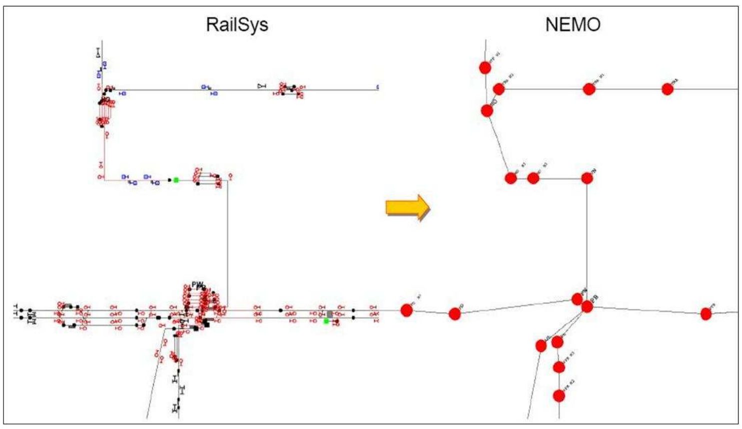

Figure 1.1: Automatic NEMO-RailSys integration (Kettner et al. 2003)

1.2.2 The formation of scale economies

Scale economies are one of the main drivers of the organisation of freight transport, and, on a larger scale, of companies' logistic choices. They characterise a production process such that the marginal production cost decreases when the quantity produced increases. Such a situation is frequent in freight transportation under at least two forms: first, decreasing unit cost with the loaded volume (e.g. due to the fixed capacity of the vehicles, or the presence of fixed assets - logistics platforms, railways, etc.); second, decreasing unit cost when the frequency is increased (e.g. due to lower detention costs). Scale economies are of particular importance in transport networks relying on large organisations, such as hub and spoke networks, or non-road modes. An additional user of such a transport network may result in a decrease of the transport cost for all customers, through a headways decrease or a better use of vehicles' capacities, for example. As a consequence, these economies of scale constitute externalities, and thus have particular implications for regulation. Anyway, their formation is seldom explicitly represented in freight transport models.

In the model of Groothedde (2003), scale economies are represented explicitly, on the basis of the frequency and size of shipments, together with demand grouping by the transport supplier. The model is applied to the design of a profitable, reliable, inland waterway freight transport network, in the highly demanding field of palletized fast moving consumer goods (FCMG). As explained by Groothedde, a number of companies tried to set up such an organisation for their own use, but all of them failed. The project Feasibility of Inland Shipping Networks: "Distrivaart" was aimed at developing an intermodal hub-and-spoke network that would comply with the requirements of FCMG transport. Groothedde's work showed that owing to the presence of scale economies, a profitable network could be designed.

The proposed transport scheme consists in a set of inland waterway services whose frequency is determined (a service being defined by an itinerary and a set of served ports), provided by dedicated pallet barges with capacity of about 20 truckloads. These services are organised in a hub and spoke pattern. Each freight shipper makes the decision to have its shipment carried on by either the truck only, or the waterway service, which necessitates specific truck movements and transshipment operations (Figure 1.2, taken from Groothedde, 2003). As the transport time by waterway usually exceeds the order lead-time, demand has to be anticipated before it is sent by waterway. This is only possible up to a certain point, and the residual demand has to be accommodated by direct road transport.

An extensive market analysis was performed to identify the potential market. Manufacturers and retailers were identified, yielding a market of 26.6M pallets a year. The search of the optimal set of services was tackled as an optimisation problem, consisting in the minimisation of total costs (including operating costs, detention costs, and handling costs.) Note that some costs were not considered, notably the potential costs of the cooperation between agents pertaining e.g. to the harmonisation of time-windows between manufacturers. Once defined, the problem was solved using the simulated annealing method, which provided both an optimal organisation, and a development path, made up of stages of increasing profitability. This methodology may be hard to extend to the scale of macroscopic simulation, due to the data and computational resource requirements. Nevertheless, it can be applied to design intermodal transport services in an operational context: if it is successful to yield profitability on a financial account, then its potential for social welfare is even greater since it also incurs environmental pollution and congestion savings. To sum up, Groothedde's work exemplifies the scale economies achieved in the transport costs by the cooperation between agents.

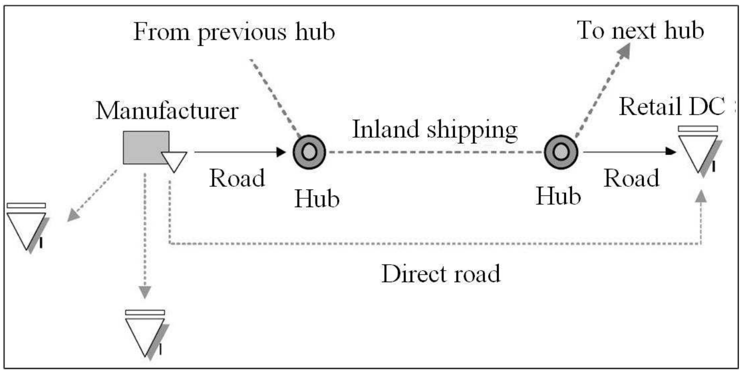

Figure 1.2: The transport possibilities (Groothedde, 2003)

1.2.3 Integrating logistic features in transport chain and generalised cost

In spatialised freight transport models, the choice of the transport mode for freight demand is generally modelled as the minimisation of the per ton generalised cost provided by each mode. This is evaluated by taking into account the physical attributes of the mode path such as carriage time, money cost, need for handling, reliability, etc. In addition the generalized costs may be cast into the framework of random utility and discrete choice models, in which the principle of utility maximization yields a choice probability for each alternative. The possibility to combine two transport modes or more in parallel is usually disregarded, or addressed in a generic, implicit way (e.g. using access-egress links, of which the transport mode is not made explicit).

Owing to model calibration, the method may prove efficient even when the drivers of mode choice are not thoroughly understood – which is certainly the case when logistic features are neglected, or when the contract terms between the shipper and the transport supplier are specific.

The NODUS model (Beuthe et al., 2002; Jourquin and Beuthe, 2000, 2005), has been designed to cope for the first flaw, by way of refined representation of the transport operation for each shipment from origin to destination, including transshipment operations. In NODUS both the transshipment and transport operations are modelled as transitions from state to state, a state being characterized by location and type of conditioning. A transition induces a change in location and/or conditioning, together with a time expense plus a money cost which depend on the logistic or transport technique utilized. The transport techniques are distinguished by infrastructure type and vehicle type e.g. several types of barges. The network of states and transitions makes up the supply network in NODUS (Figure 1.3, taken from Jourquin and Beuthe, 2000); hence a network path from origin state to destination state models a transport chain including logistic operations.

Precisely, the transition attributes associated with a network arc induce a generalized cost which accounts for the following items:

- *movement costs:* implied by physical operations (capital cost, labour, fuel, insurance, maintenance, tariffs), be it on board or during a loading, unloading, or transshipment operation,

- *inventory costs:* implied by the detention of goods (opportunity cost, potential depreciation cost) and the storage costs,

- *residual costs:* (e.g. general administrative costs).

These costs should depend on the shipper and shipment type and size. This is approximated by distinguishing commodity groups. Costs are evaluated on the basis of a unit cost per ton, which implies that the shipment size is not taken into account. Costs pertaining to information availability, reliability or freight safety are not made explicit. The network paths from an origin to a destination thus represent alternative transport chains that may make use of all available modes, eventually in sequential combinations i.e. intermodality. Each commodity flow is faced with a set of transport chains, and it is assigned to the chain with minimal generalized cost. Congestion effects can also be modelled at the arc level.

To sum up, the NODUS approach enables one to represent explicitly the various components within the generalised cost of carrying a shipment through a logistic and transport chain. In connection to the search for scale economies, the approach is suitable to predict the demand choices between barges of different capacities.

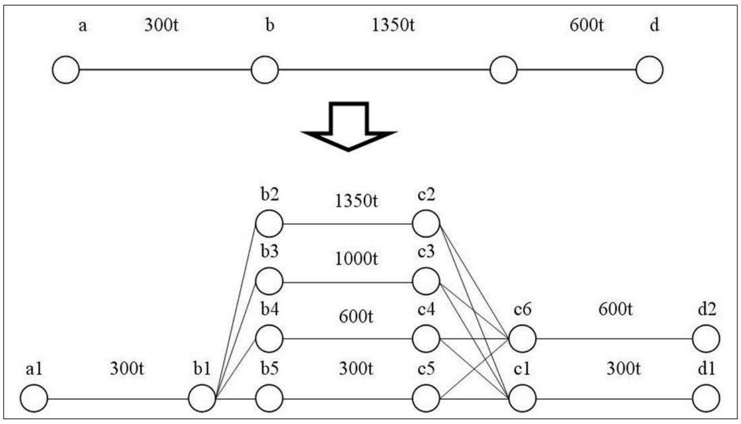

Figure 1.3: Formation of the supply network in NODUS (Jourquin and Beuthe, 2000)

1.2.4 Modelling the spatial and technical structure of the industry

The demand for freight transport arises from the fact that the demand and supply of commodities are spatialised and temporalised, i.e. goods are not produced at the times and places they are needed. A description of the commodity demand and supply therefore provides a strong basis for the generation of freight transport demand, especially its linkage to the rest of the economy, which can be particularly relevant in a forecasting objective.

The class of Spatialised Input-Output (SIO) models, as reviewed by Marzano and Papola (2004), is purported to capture the spatial and economic relationships between the local demand and supply of commodities. They constitute a generalisation of the Input-Output (IO) class of models, initially designed by Leontief (1936). Leontief's original idea is to represent, for each sector, the inputs necessary to produce a given amount of output. This linkage between inputs and outputs constitutes a set of technical relationships, which make up the sector production structure. Under their simplest form, the technical relationships are modelled as linear combinations, the coefficients of which are known as the technical coefficients.

In an SIO model, the spatial structure is also addressed by distinguishing among regions of production - hence of intermediary consumption. This requires to model:

- *the structure of production:* by industry sector and region, namely under the shape of an IO matrix, region by region, and the size of each sector in each region,

- *the trade between regions:* for each type of input i.e. each commodity group. This trade gives rise to the freight flows between regions. It may be modelled by trade coefficients which indicate how the demand of a given region in a given input is fulfilled by other regions. In turn the trade coefficients may be modelled on the basis of either economic principles such as utility maximization (taking into account both the input price in the region of production and the cost to transport the input from the region of production to the region of consumption), or of statistical principles such as inference by entropy maximization.

The last issue in an SIO model pertains to the final consumption by region and sector. This makes a specific model input which, together with the industry intermediary consumption, sets the need for and level of economic production and thus of freight transport activity.

An SIO model enables one to assess first the impact on freight transport demand of a change in the final demand of a given good, and second the impact on the economy of a change in transport costs. The two main limitations to the SIO class of models pertain to, first, their large data requirement; second, that they cannot address issues such as the firms' behaviours, or the migration behaviour of the workers that provide the labour force, among others.

The class of Spatial Computable General Equilibrium (SCGE) models is aimed at the spatial and economic relationships between the local demand and supply of commodities, as in SIO models, and also at the labour market (with the subsequent demand for passenger transport) and above all the firms' behaviour in the production of commodities, in relation to the goods' regional prices and the market structure in each sector. A wide range of specifications are available, see e.g. Harker (1986) for a set of available assumptions about the nature of competition in each region. Harker describes how one can represent either a monopoly or a Cournot-type oligopoly competition in each region, and whether monopolistic firms control the transport industry or not.

The specification of an SCGE model covers a wide range of issues. As illustration, let us consider an SCGE model for the Netherlands called Regional Applied general Equilibrium Model (RAEM 3.0), from Thissen (2003) and Ivanova (2007). By region the following issues are modelled:

- *Products:* sectors and varieties are considered, each sector providing a given number of varieties.

- *Production:* it is described as a two-tier production function. The upper tier is a Cobb-Douglas function setting the trade-off between labour and sector intermediate inputs. The intermediate inputs nest is modelled along a Dixit-Stiglitz monopolistic competition framework (see e.g. Fujita et al., 1999). There is no lower tier for labour, which is considered a uniform input.

- *Consumption:* total income in each region is fully spent on consumption. The utility function is two-tier. The upper tier, which yields the consumptions of each sector, is specified as a Stone-Geary utility function. The lower tier, which concerns the consumption of each variety in a given sector, is specified as in the Dixit-Stiglitz model.

- *Prices:* the price of a variety in a given region is equal to the price of this variety in the region where it is produced plus transport costs if that region is different.

- *Labour:* the labour market is represented on the basis of a search model, which yields wages, employment rates and the commuting matrix, thus providing a major driver of passenger transport demand.

- *Transport:* is modelled under the assumption of perfect competition.

- *Government:* the government raises income taxes so as to achieve social transfers. It consumes a number of commodities, under a budget constraint.

- *International trade:* it is modelled on the basis of an Armington assumption, i.e. that the goods produced for domestic consumption or for exportation have different specifications; the converse applies to importations. Switching between the specifications is possible but not totally flexible.

- *Migration:* based on the differentials in the regional utilities of living, people may migrate in order to improve their situation.

This shows the complexity of SCGE models. They are purported to capture a wide range of phenomena, on the basis of a wide range of microeconomic, behavioural models. Some technical assumptions are required to ensure the existence and uniqueness of equilibrium, which limit the model outreach. An SCGE model represents explicitly the linkage between freight transport demand and the rest of the economy — at the expense of a large data requirement and heavy specification work.

1.2.5 Identifying the logistic stages within the trade relationship

A model of production and consumption may be coupled with a model like NODUS so as to integrate accurate generalised transport costs within the trade relationships between regions. However such a treatment would miss other important logistic issues, such as for example the number, location and function of warehouses or cross-docking platforms. Indeed a warehouse, through the storage of one or several commodities, is purported to facilitate the matching of disaggregate demands and supplies over time and space.

The identification of logistic facilities, such as warehouses and break-bulk platforms along logistic chains from production to consumption region necessitates a specific model. Let us now introduce two methods designed to tackle this issue.

The first approach has been designed in the frame of the SCENES European project[1] (ME&P, 2002). This project had a number of objectives, among which building a European spatialized transport model, by improving the STREAMS model (ME&P, 2000). The SCENES model includes an SIO freight transport demand model, which yields Production-Consumption (PC) matrices by commodity group. The PC to OD issue is

[1]At the origin of this project is a pioneering initiative, called SMILE (Strategic Model for Integrated Logistic Evaluations), by the Transport Research Centre of the Ministry of Transport and the research organisations NEI (Netherlands Economic Institute) and TNO Inro. See e.g. Tavasszy et al. (1998).

addressed by a specific logistical module called SLAM (Spatial Logistics Appended Module), which works in three steps as follows:

1. For each PC pair a small set of candidate regions for a logistic operation in a distribution centre is selected. The selection is based on three factors of, respectively, economic activity, centrality (meaning the presence of the region in the chains for serving the PC pair), and accessibility to the various infrastructures networks.

2. A number of candidate logistic chains are generated, whereby the commodity travels through zero, one or two of the previously selected distribution centres. The generalised cost of each kind of chain is evaluated on the basis of transport costs (the arc cost being a weighted average of the costs of the various modes), inventory costs (uncertainty of the demand, safety), and logistic costs (detention, handling).

3. The PC volume is assigned to the candidate chains according to a nested logit model, where the logistic chain type is determined at the upper tier, then its geographic location at the lower tier. The model yields OD matrices between zones of production, consumption or logistic stage. These include the usage of logistic facilities by freight flows, and the amount of usage at a given place is related to the amount of logistic facilities supplied there.

The second approach, chosen for the EUNET2.0 model (Jin and Williams, 2005), is also based on an SIO model of production and trade, in association to a model of logistic chains. However the two features are embedded in a single, extended SIO framework, wherein significant logistic stages are addressed as additional industry sectors.

In the EUNET model, by commodity group a set of candidate logistic chains is defined (Figure 1.4, taken from Jin and Williams, 2005). Each chain consists in a series of transitions called logistic legs such as factory towards depot, or distribution centre towards local wholesale. Commodities are therefore also distinguished by the logistic leg they comply with, consequently making up what we could call virtual commodity groups. The trade between regions is assumed to derive from a utility maximisation behaviour, taking into account the generalised transport cost and the scale factors that characterize the regions' industrial and logistic structure. As a consequence, the regional productions, consumptions, interregional exchanges and logistic chains are simultaneously calculated. Some logistic specific variables which could be observed, like the handling factor (i.e. the number of time the freight is touched between the production and the consumption places), are used for model calibration. In EUNET the PC matrix integrates the logistic stages, which need not be modelled in an companion logistic module such as SLAM.

Both the SCENES and EUNET approaches take into account the logistic organisation as a determinant of the demand for freight transport. Their application requires a reasonable amount of data, disregarding the details of the logistic choices made by the various companies. A common, major advantage is to address mode choice in the context of the choice of an integrated transport and logistic chain. However, some economic hypotheses in the Input-Output models are questionable (particularly the linearity of the interdependence between sectors, and the non-substitutability of production factors - a limitation which is relaxed in SCGE models). Furthermore the logistic choices are not modelled in a micro-economic, behavioural way, which puts at risk any forecasting application.

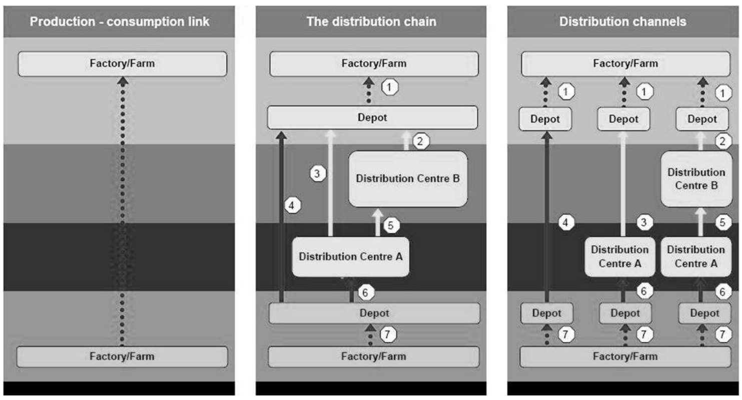

Figure 1.4: An example of logistic chain in EUNET (Jin and Williams, 2005)

1.2.6 Modelling the carrier behaviour

Urban freight transport is a particular problem for modelling since it features specific shipment endpoints, small distances, small commercial speeds, small shipment sizes and the need to organize efficient rounds. It is therefore difficult to address urban freight transport using the classical four-stage architecture of transport demand models.

The FRETURB model (Routhier et al., 2002; Ambrosini et al., 2004) developed in the frame of the French research program TMV[2] is aimed

[2] *Transport de Marchandises en Ville*, *i.e.* Urban Freight Transport.

at analysing freight flows in an urban area on the basis of limited data requirement; a related objective is to assist local authorities in their transport policies.

Within the TMV program, three large surveys were conducted in the French cities of Bordeaux, Dijon and Marseille: strong statistical regularities were identified concerning, first, the organization of truck rounds in terms of stop number and duration, and of average distance between two stops (Figure 1.5, taken from Routhier et al., 2002) etc.; second, the infrastructure impact of urban freight transport in terms of parking type and duration, of road traffic due to trucks etc. These regularities were embedded in FRETURB to model the movements of commercial vehicles on a given urban area on the basis of a set of socio-economic variables.

Two aspects of the architecture of FRETURB are particularly interesting. First, as a spatialised model it is focused on a set of zones that represent the studied area, rather than a network of nodes and arcs. As a consequence the model does not yield indicators by nodes and arcs, but indicators aggregated by zone: total distance covered (in vehicle.km), road occupancy (in vehicle.hours distinguished by the type of occupancy: moving, parking, illegal parking, etc.) Second, FRETURB has a specific architecture which combines the following modules:

- *Movements' generation:* in each zone, the number of movements (delivery and pick up operations) is estimated from the number and characteristics of the economic activities within the zone described by the type of activity, the type of settlement (offices, warehouse, etc.), the number of employees, etc.

- *Distance covered by commercial vehicles:* knowing the number of movements, the way rounds are organised, and a number of descriptive variables such as the zone's density, the distance covered by the commercial vehicles between each operation is calculated, and as a consequence the total distance covered is estimated.

- *Road occupancy calculation:* this is performed along similar principles.

- *Road occupancy with respect to time of day:* it is computed on the basis of observed timetables.

As a conclusion, the FRETURB model is an elaborate way to extrapolate results of surveys conducted in a limited set of cities to any city of similar socio-economic development by using a limited amount of descriptive data. It is all the more relevant as urban freight transport is a particularly difficult context to address by classical demand modelling methods, both for theoretical (companies behaviour, organisation of rounds) and practical reasons (need for data). However, the framework is not suited to cost-benefit analysis. Indeed, the model does not represent how carriers adapt to a change in the economic environment.

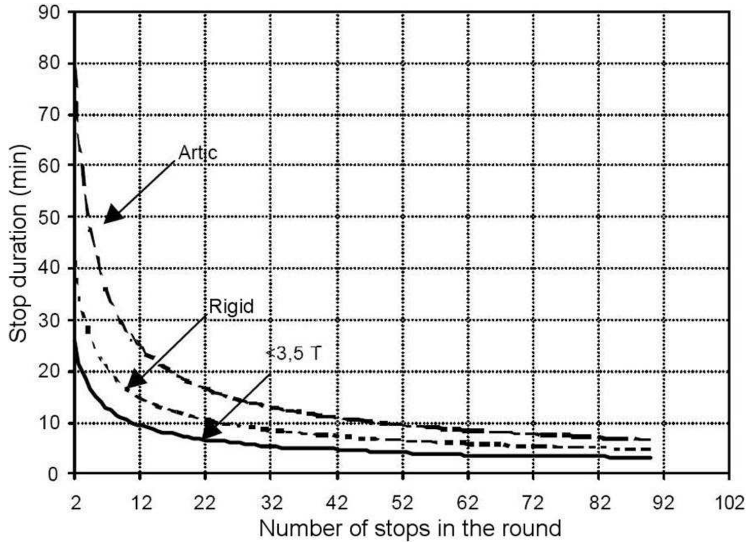

Figure 1.5: Average duration of a stop in a round (Routhier et al., 2002)

1.2.7 About shipment size: a survey and a model under construction

Among the issues which are not addressed in the modelling methodologies reviewed so far, a major one is the representation of shipment characteristics and of the cooperation between the various agents involved in a transport chain. Both issues are particularly relevant in the context of an individual freight carrier: however little microeconomic theory of this specific area is available yet; even if it were, its application would require critical data about shipments. A shipment survey is appropriate to yield such data and to provide insight into then behaviour of both shippers and carriers.

Let us now introduce the ECHO survey, which was conducted in France from 2002 to 2004, of which the first results have been made available by Guilbault et al. (2008) (see also Guilbault and Gouvernal, 2008). The ECHO survey is based on a sample of about 3,000 businesses and

10,000 shipments (a shipment being defined as a given amount of freight of a given kind, made available at given place and time, by a unique shipper in order to be moved as a whole towards a given, unique receiver). Shipments using non-road modes have been over-sampled in order to improve statistical significance. Each shipment is accurately described, with particular emphasis on the relationships between the various agents involved in the shipment transport chain. The shipper business is also described, with emphasis on the way it organises its production.

This survey already yielded noticeable results about the structure of shipments and the use of transport modes within transport chains. It can certainly be used for much deeper analyses, notably concerning the shipment size and the relationships between freight agents (see Chapter 2). Similar surveys exist in other countries, though not with the same level of accuracy (e.g. the Commodity Flow Surveys in the U.S.A. or in Sweden.)

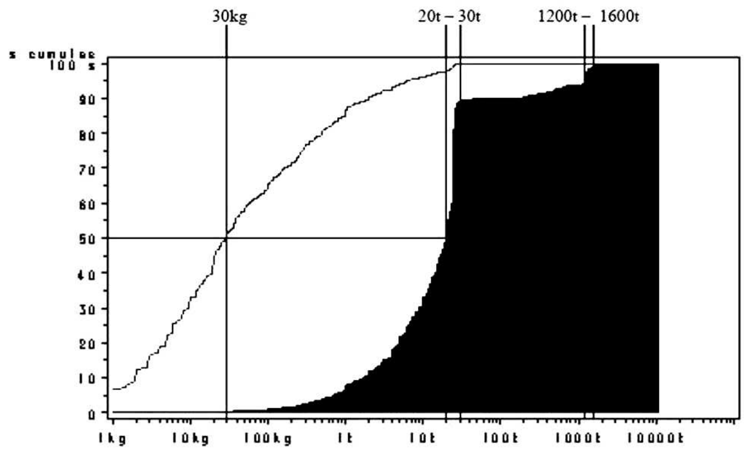

Figure 1.6: Distribution of shipments (Guilbault et al., 2006)

Let us also mention the work of de Jong et al. (2005), continued in de Jong et al. (2007) and de Jong and Ben-Akiva (2007). This work aims at relating the shipment size to logistic costs at the disaggregate level of a shipment, and conversely to relate the transport costs to the consolidation of shipments in vehicle loads. The architecture of the model is made up of three steps:

- *Disaggregation:* P(W)C flow matrices (W standing for an intermediate wholesale sector) are disaggregated into firm-to-firm flows;

- *Logistic choices:* the decision of shipment size, transport mode and routing are made in a disaggregate way, on the basis of a generalised logistic cost function;

- *Aggregation:* the resulting shipment flows are aggregated so as to yield vehicle flows. Consolidation should be taken into account at this step.

Although the model is still under construction, it deserves much attention since it explicitly takes into account the shipment size.

1.2.8 Identification of the long-run drivers of freight transport demand

Beside the spatial structure of freight transport demand, the long-run drivers in its temporal evolution make a major issue in the understanding of the transport system. In the last decades, in-depth changes have taken place in the logistic organization of firms, the organisation of transport (optimisation, subcontracting), the inventory strategy (including pooling and concentration), the production organisation (outsourcing, specialisation, postponement), market strategies (product diversification, short order lead-times, short life-cycles), etc. (see e.g. Dornier and Fender, 2007). In parallel, the sector of freight transport and logistic has been under consolidation, which also implies specialisation, externalisation, merges. All these evolutions impact the freight supply and demand, and therefore vehicle flows, and it is a tempting issue to identify the causalities and put them in a hierarchy.

The European project REDEFINE (NEI et al., 1999) was conducted under the 4th Framework Program to analyse the logistic drivers of the transportation demand in five countries (France, Germany, the Netherlands, Sweden and the United Kingdom), using both quantitative and qualitative approaches. The quantitative approach, in which we are interested, consisted in relating the overall industrial production, expressed in monetary unit, to the overall distance covered by commercial vehicles throughout a series of stages characterized by key ratios.

The ordered set of key ratios provides a framework for the distinction and analysis of the various trends in the evolution of freight transport demand. These are listed hereafter:

- *Density of value:* the ratio of the value of a produced good to its mass. This ratio is useful to convert the money value of a production into its mass value, generally measured in tons in freight transport.

- *Modal split:* the proportion of goods mass which is carried by the road.

- *Handling factor:* the ratio between the mass lifted and the mass produced, which stands for the average number of times the goods are handled.

- *Average length of haul.*

- *Vehicle carrying capacity.*

- *Loading factor:* this ratio defines the average filling rate of the vehicles.

- *Empty return:* to yield the share of those vehicles running empty.

On the basis of this analytical frame, an overall trend may be split into a set of simpler phenomena (Figure 1.7, taken from NEI et al. (1999)). As a consequence, it is possible to identify the critical drivers and, up to a certain point, the scope for policy. It is also possible to use this framework for extrapolation. The main disadvantage of the approach is that the key ratios are still very aggregate indicators. Their respective trends may well hide strongly varied evolutions at the disaggregate level of microeconomic agents. Thus the REDEFINE model provides a first step towards understanding the system evolution, rather than a definitive explanation.

Breakdown	France	Germany	Netherlands	Sweden	United Kingdom
Value of production and imports	+28%	+14%	+17%	+82%	-4%
Value density	*+23%*	*-2%*	*-3%*	*+51%*	*-32%*
Weight of produced and imported goods	+4%	+16%	+21%	+21%	-7%
Modal split	*+10%*	*+20%*	*0%*	*+11%*	*+1%*
Products transported by road	+14%	+33%	+21%	+34%	+1%
Handling factor	*+2%*	*-2%*	*+3%*	*-20%*	*+18%*
Road tonnes-lifted	+16%	+31%	+25%	+8%	+18%
Average length of haul	*+36%*	*+4%*	*+29%*	*+37%*	*+24%*
Tonne-kilometres	+57%	+33%	+60%	+48%	+46%
Vehicle carrying capacity	*+15%*	*N.A.*	*+24%*	*+28%*	*+9%*
Load factor	*+7%*	*N.A.*	*-3%*	*-4%*	*-4%*
Average payload	+23%	N.A.	+20%	+22%	+4%
Empty running	*-21%*	*N.A.*	*-7%*	*-7%*	*-5%*
Vehicle-kilometres	+28%	N.A.	+30%	+18%	+37%

Italic cells are ratios, others are aggregates.

Figure 1.7: Economic activity and road freight transport 1985-1995

1.3 Assessment

Having reviewed some advanced models for freight transport demand, we are now in a position to assess their respective outreach with respect to the following set of issues: first, the demand model in terms of volume, behaviour and heterogeneity; then, the supply model in terms of networks and services; next, the relationships between agents and the market externalities; lastly, the potential outreach for policy assessment.

Our criteria are targeted mainly to supply-demand models, wherein the demand and supply are modelled separately prior to being faced with each other, which yields the activity of the related sector in freight transport. Out of the models reviewed, those in REDEFINE and, up to a certain point, FRETURB, are focused on freight transport activity in a straightforward way, with no attempt to model the supply and demand components and to put them into balance.

1.3.1 Quality of the demand's representation

The demand for freight transport derives from the need in given places of commodities available elsewhere. Given the transport services available and their characteristics, shipments will be transported in order to fulfill this need. The main questions from a modelling perspective are: how to quantify the need for transport? On which grounds and in which way are the transport services selected? To which extent can the shipments be aggregated in the model?

1.3.1.1 Demand volume

On the basis of much empirical evidence (see e.g. Guilbault et al., 2008), shipments are known to be widely varied in size and characteristics. Besides, few freight databases include such a level of detail. As a consequence, the demand for freight transport is often stated in flows of a given intensity, expressed in tons per period of time, from origin to destination zones, by type of commodity. Overall, the demand is generally represented by a set of origin-destination flow matrices. Network assignment models focus on how the transport demand, taken as an input, results in vehicle flows on the infrastructure networks. This category includes NODUS and NEMO. In these models, OD flow matrices are exogenous.

Other models focus on the formation of freight transport demand: notably those in the SCENES and EUNET research projects, which both include an SIO component. They proceed jointly to the generation and the distribution of the freight transport demand. This architecture enables to derive the need for freight transport from the spatial and industrial structure in the area of scope. Here the two recent achievements are, first, the description of the industrial structure hence the population of shippers taken in an aggregate way; second, the inclusion of important logistic stages related to warehouses and platforms, into the process of production and distribution. Thus the Production-Consumption matrices, as yielded by the SIO model, are turned into origin-destination matrices, which are more appropriate to perform the next step: to model the choice of transport services.

We outlined the strategy used in the SCENES model, which consists in appending a specific module (SLAM), in order to turn the P-C matrices into O-D matrices. Such a module can be designated as an LIO stage, for Logistic Input-Output. The EUNET approach takes a different way: logistic organisations were categorized and commodity groups were further disaggregated to distinguish the logistic leg that pertains to the commodity, thus yielding virtual commodity groups. Both approaches allow for logistic imperatives to appear in the formation of freight transport demand, although the drivers in the logistic choices are not modelled in a microeconomic way.

Among the models we reviewed, FRETURB is the closest to considering shipments explicitly. The demand is predicted on the basis of variables describing the economic and industrial base of the area of scope. Furthermore, it is not expressed in flow but in operations (pickup or delivery). Shipments are not considered explicitly, but the level of detail is high, and the demand formation is closely linked to the industrial basis, which is very appealing in a forecasting perspective. To sum up, the demand for freight transport differs from the demand for passenger transport at least for its specific linkage to the rest of the economy. The projects of making explicit the logistic imperatives and of taking the shipments as the decision unit have been undertaken, but the recent advances still leave much room for development.

1.3.1.2 Description of the behaviour of transport demand

Assuming that the needs for moving goods have been modelled, the next step is to predict which transport services will be used, i.e. how the freight flows will spread on the networks. Along a partition which has already been proposed (see e.g. Marzano and Papola, 2004), two main strategies can be identified throughout the works reviewed. The first is mainly statistical: a set of descriptive variables are used to predict the flows, building on a number of statistic regularities. The second approach is based on the description of the agents' behaviour, including notably the paradigm of utility maximisation.

Statistic approach The statistic approach is best illustrated in FRETURB wherein, from the large amount of data collected in a sample of three cities, statistic regularities were observed considering the derivation of freight transport from industrial and economic activity. Thus econometric relationships were estimated to link the intensity of freight transport to variables describing the industrial and economic basis. The way shippers and carriers organise themselves is not explicit.

The SLAM module in the SCENES model is also based on a statistic approach. In order to turn P-C flows into O-D flows, the module yields the probabilities for a break-bulk operation and the region where it would take place. Three indicators of economic activity, centrality and infrastructure accessibility are used as explanatory variables, with no underlying economic model.

Both approaches are readily operational and yield useful results; either one may be instrumental in a forecasting perspective but, as will be shown later, they are not as appropriate for policy testing.

Behavioural approach In a behavioural approach, some agents are explicitly modelled as decision-makers involved in a choice process to select one option among a set of alternatives. In most of the models that we reviewed, the choice of the network route for a shipment is modelled as a discrete choice, of which the decision-maker may be the shipper or a carrier. No indication is provided to distinguish between the two economic positions; in the real world they can indeed be integrated, in the case of own account transport.

Then comes the issue of which choice alternatives may be considered by the decision-maker. In the area of freight transport the significant advance is to identify alternatives that integrate transshipment options: this is achieved in NODUS.

The next step in a behavioural approach is to model the perception and evaluation of the choice alternatives by the decision-maker. In conjunction with the identification of logistic features in the transport chain, the significant progress here is to identify the economic drivers of this behaviour, including operation costs, detention costs, inventory costs and handling costs: this is also achieved in the NODUS model[3]. A related

[3] the approach is restricted to facilities that are available to any customer, with no

advance is to take into account the temporal requirement on the shipment: this is achieved in Groothedde's model with the segmentation of shipments with respect to the requirement of disposal either with some delay or as soon as possible.

The last issue in the behavioural approach pertains to the economic preferences of the decision-maker, and to his trade-off between the attributes of an alternative. In general, the assessment of an alternative is modelled through a generalized cost function (or disutility function), which takes into account a set of attributes, each of which is weighted by a coefficient of trade-off against money. Here the advance would be to make two separate accounts of, respectively, the time and money expenses: thus a time versus cost trade-off could be modelled at the level of the integrated alternative; and the decision-makers could be distinguished by their relative preference of time to cost (i.e. their unit value of time). This approach has two main advantages: first, in terms of economic outreach, this enables to model non linearity in the utility functions and also to make explicit the influence of the agent's money budget as well as that of his time budget; second, it is instrumental in that this increases the modeller's control over the calibration process and provides more flexibility in the specification of a statistical distribution for agents' trade-offs between time and money. This advance has been achieved in passenger mode choice models since the 1960s, especially in price-time models (Quandt, 1968, Marche, 1973). However in the freight models that we reviewed the time and cost expenses are not accounted for separately at the level of the choice alternative. To our knowledge the distinction is only achieved in the truck network assignment of the French Department for transport (Danzanvilliers et al., 2005), thus being restricted to route choice on a road network by OD pair with no consideration of logistic features.

An additional issue pertains to the shipment size. Obviously the choice of the shipment size depends on the commodity group and its logistic requirements, the distance to overcome, and also the available transport services and their characteristics (see e.g. Hall, 1985). This issue is particularly relevant to understand correctly the activity of freight transport.

Addressing demand heterogeneity The demand consists in a set of agents (or a set of shipments sent by or routed by these agents), who need to use transport services. These agents or flows may have very different requirements on those services, and considering them as a uniform

distinction of firm

population may lead to large biases in freight transport modelling.

There are two classical strategies to tackle demand heterogeneity. The first one is to split the demand into classes (also called segments), which is very much constrained by the level of detail in the available data. Flows are generally categorised into commodity groups (as in almost all the works reviewed in this paper), but some groups may still be very heterogeneous. The segmentation in EUNET is noteworthy since the commodities are distinguished by logistic family. Demand segmentation is purported to improve the model relevance by grouping the similar components of demand; however there is the issue of which criterion would be relevant to characterize this similarity. The notion of a logistic family is still to be defined clearly; if some commodity groups had similar logistic imperatives, their transportation would probably be organised in similar fashions, which would be amenable to a unified model. Demand segmentation has been taken in an original way in FRETURB, in which the businesses are distinguished along a number of characteristics.

A related strategy, which was not used in the models we reviewed, consists in modelling the drivers of demand heterogeneity in a probabilistic way, by associating a random variable with given statistical distribution to each driver and also a joint distribution to the vector of drivers. This is an explicit, probabilistic approach to demand heterogeneity - in fact very much the same as demand segmentation into classes.

The second broad strategy is to address demand heterogeneity in an implicit way, in the framework of random utility theory, by incorporating an error term in the utility functions that the agents associate to choice alternatives, as in the logit choice model. This error term stands for unobserved characteristics or idiosyncratic choice criteria, among other features (see e.g. Ben-Akiva and Lerman, 1985; Anderson et al., 1992). This is instrumental when the drivers of some choices are not fully understood, particularly so in the case of modal choice. This strategy is used in SLAM and EUNET.

The two strategies are integrated in the framework of discrete choice models, by making the random utility functions depend on the segment characteristics (such as in random utility with random coefficients).

1.3.2 On the supply representation

Let us come to the supply-side and consider it in a demand-oriented perspective: our aim is to assess the supply features that are relevant in the demand behaviour and choices. The detailed models of supply operations fall out of our scope, except in their connections to the demand, as in the interaction of Railsys and NEMO.

Let us consider in turn the three issues of, respectively, the spatial and layered representation of supply; the modelling of transport and logistic features; and the formation of scale economies.

On the spatial and layered representation of supply

Throughout our review, except for ECHO and REDEFINE, the spatial extension of the supply is modelled, either in a zone-based approach in the case of FRETURB, or in a network-based approach. A network model of nodes and arcs is most appropriate for the infrastructure layer, be it the transport mode road, railway, inland waterway, maritime or air.

In connection to the infrastructure layer, two other layers may be modelled. The distinction of vehicle types is achieved in NODUS. The distinction of both vehicle types and services is achieved in NEMO for the railway mode, and in Groothedde's model about the road and inland waterway modes.

However there is no model of integrated transport and logistic operations as delivered by some logistic providers in the real world: yet this is related to the issue of making explicit the shipment size.

The modeling of transport and logistic features

Nevertheless, significant advances have been achieved to model logistic features and operations within a transport chain: in NODUS these are modelled by dedicated network arcs, whereas in SLAM and EUNET the significant logistic stages are made explicit and used to turn PC relationships into OD relationships.

On macroscopic relationships: scale economies and congestion

Scale economies are of particular importance at every layer in the supply of freight transport, then also in the transport and logistic choices of the demand: in the shipment size, shipment frequency, vehicle and mode choice, and service choice. The classical way to represent scale economies in freight transport demand models is based on a mass unitary cost associated to any operation of a transport or logistic kind, together with transition costs associated to any transfer from one operational stage within the transport chain to the next. This approach enables to simulate how the demand agents can benefit from the presence of scale economies; it provides no indication of the formation of scale economies, of the underlying rationale of the transport and logistic providers. To achieve that purpose, a refined model of supply is in order, as in Groothedde's work. An intermediary step is taken in NODUS, in which the unit costs are closely related to the characteristics of the facilities. Taking a further step towards realism may be possible by taking explicitly into account the shipment size, and then account for consolidation, as is being done by de Jong and Ben-Akiva.

Another macroscopic relationship between the flow intensity and the level of service is that of congestion. This issue has been much more observed and well understood than the formation of scale economies, and it is often modelled on the basis of a speed-flow relationship at the level of a given transport arc. This requires to define a flowing capacity for that arc: models of roadway capacity have been well-developed, whereas the capacity of non-road modes is a more complex issue that must be addressed at several layers (in terms of vehicles, shipments, services...). The impact of congestion on reliability is also a topic for further research, for road and non-road modes.

1.3.3 On agents' relationships and market externalities

On the relationships between sector agents

The formation of scale economies is closely related to the industrial organization of transport supply, hence to the relationships between the transport and logistic providers. Another such relationship lies in the complementariness of the various facilities and services, which can be used in an integrated way by the agents constituting the demand. No other feature of supply relationship was detected in our review.

On the demand side, in every model it is assumed that shipments are independent of each other: little attention is paid to the issue of the shipment size, which pertains to the internal organization of the demand agent and is obviously more important than any relationship between demand agents. A noticeable exception to that point is the real-world context of Groothedde's work, where shippers joined together so as to stimulate the design of a multi-client service: in our classification this is rather an issue of scale economy.

Lastly, the matching between providers and users was indicated in the French shipment survey, ECHO: this provides some insight into the commercial relationships in the freight transport sector, especially about the temporal requirement and the price.

On market externalities

The macroscopic relationships of scale economies and congestion constitute market imperfections (or distortions) in the framework of the neoclassical theory of microeconomics.

Other market imperfections lie in the external impacts that the freight system exerts on the socio-economy and the environment: these include positive impacts such as the achievement of scale economies in any sector of the economy, as well as negative impacts - from the emission of pollutants and noise, to the risks of accidents and the degradation of the residents' living conditions. These external impacts fall outside the topic of freight demand models: it is easy to address the negative impacts by using dedicated models by impact type, taking the flow and level of service results of the demand model as input to evaluate the impact.

1.3.4 On the ability to assess regulation policies

This is a twofold issue: first, is a modelling framework appropriate to take into account a given policy? Second, is it possible to perform a Cost Benefit Analysis, and what would be its outreach?

To answer these questions one has to face the policy targets and instruments with the model scope in terms of (1) supply representation, so as to effectively accommodate the implementation of the policy; and (2) demand representation, so as to effectively simulate the demand response to the policy. From our earlier observations, it is obvious that the NODUS, SLAM and EUNET models provide the widest frameworks for policy assessment. The FRETURB model may be appropriate to simulate some of the effects of a change in commercial speed in an urban setting.

If the model is policy-responsive, then it may be used to perform a cost benefit analysis of the policy under investigation.

1.4 Conclusion

1.4.1 Synthesis based on a two-dimensional typology

In this paper, advances in freight transport demand models were reviewed and assessed. Let us sum up our analysis by putting forward a classification framework, in which each model is assessed along the two dimensions of the focus and the behavioural content.

The focus axis includes three categories as follows:

- *Supply side orientation:* this indicates advances in the representation of transport and logistic features.

- *Demand side orientation:* this indicates advances in the formation of the demand and/or choices of the demand agents.

- *Sector activity orientation:* the focus is on deriving the intensity of the freight transport sector activity, in a direct way rather than through a demand-supply model.

The behavioural axis includes three levels of analysis, presented hereafter by order of increasing depth and outreach[4]:

- *Descriptive approach:* this pertains to the works that provide more data, or some trend analysis, without further treatment. Such works provide a sound basis for further studies and, up to a certain point, the understanding to freight transport.

- *Statistical method:* this category refers to works in which statistic regularities are identified between various variables, but with no underlying microeconomic model.

- *Behavioural method:* this pertains to explicit models of agent behaviour, often on the basis of utility maximisation. This category is most appropriate for realistic simulation, project evaluation and policy assessment.

1.4.2 Some research perspectives

Our typology provides insights into how to combine the recently achieved modelling advances in order to take an in-depth approach to topics from among demand, supply and activity: hence to improve the behavioural basis. As the field of freight transport makes indeed a large area for study and research, several directions for development were suggested along our review. Let us emphasize here two issues which we believe particularly relevant for further investigation:

[4]This typology axis was suggested in other researchs (see e.g. Catalani, 2003)

Approach / Orientation	Descriptive	Statistical	Behavioural
Supply side			NEMO-RailSys Groothedde NODUS de Jong & Ben-Akiva
Demand side		SIO SLAM	EUNET 2.0 RAEM 3.0 de Jong & Ben-Akiva
Sector activity	ECHO REDEFINE	FRETURB	

Figure 1.8: Modelling advances by focus and behavioural content

Market structure the relationships between the various agents of a transport chain, and particularly the contracts they link, certainly exert a strong influence on the elaboration of transport services and costs, and therefore on the shippers' decisions. For instance, large flows may imply strong competition on some links and some modes, low prices, and as a consequence, high availability and flexibility for shippers. Conversely, scale economies are achieved through capital-intensive methods, meaning that some cooperation between suppliers is required to benefit fully from them — which probably requires in turn some coordination mechanism. Overall, the market structure plays an important role in freight transport, particularly so in the intensity of demand and its modal choices.

The choice of the shipment type the size, frequency, conditioning of shipments result from demand decisions made under supply conditions. In the theory of logistics, there are models for the choice of shipment size and frequency; they still have to be included into the demand-side of freight transport demand models.

Chapter 2

Logistic issues and their modelling

2.1 Introduction

Freight transportation demand derives from the spatial and temporal nature of the economy. More precisely, it derives from the fact that consumers are located at places and times which are not the ones where commodities are efficiently produced. This discrepancy constitutes a gap, calling for a bridge: freight transport flows make this bridge.

Understanding freight transportation demand thus proceeds from understanding this gap and the way it is bridged by the agents implied, either firms or end-consumers. This immediately raises a set of questions of two natures. The first type of questions pertains to the agents behaviours, and their drivers: how do firms address the spatial and temporal gap between production and consumption? How do their interactions influence this gap, and conversely? How do these more or less atomic behaviours build to form macroscopic phenomena, such as the ones a freight transport modeller is interested in?

From the perspective of freight transport modelling, which is ours, this first set of questions is followed by a second one, pertaining to the macroscopic modelling of a whole population of agents of heterogeneous natures and behaviours.

We translate these questions into two objectives, which guide our approach in this chapter. First, we aim at understanding freight transportation demand as the result of the behaviour of the agents facing the production-consumption gap. Second and subsequently, we aim at proposing methodological progresses or room for progress in freight transport demand methodology.

To address these two objectives, we proceed to a bibliographic review. The aim of this review is to identify the approaches used in addressing logistic problematics, to understand more closely the drivers of freight transportation demand, and finally to identify methodological tools or models that could be used from a freight transportation demand modelling standpoint. The perimeter of this review is defined by the following criterion: does the work considered address a logistic issue? The word logistic is used here in its widest meaning: each work pertaining to the behaviour of firms in addressing the spatial and/or temporal dimension of the economy is of potential interest. Naturally, these works are far too numerous for us to review them all. This review thus keeps indicative, and only presents a subset of the works of interest.

The sources of information on which this review is based are of varied natures, as a consequence they will be presented in a organized way in subsection 2.1.1. This review reveals that our field of investigation is characterised by a large heterogeneity of problem natures and scales. Our findings have thus been organised accordingly, along rules we present in subsection 2.1.2.

2.1.1 Information sources

Our sources of information are of varied natures, and we present them here from the most specialised to the least specialised.

The main source of information this review is based upon is the very large range of academic reviews, technical reports proceeding from administrations or from design offices, and books. These works belong to such varied fields as transport, transport economics, industrial economics, spatial economics, industrial management, logistics, operation research, etc.

The most important non-academic source of information is the special press, for it provides useful elements of information about the current state of the logistic sector. The logistic sector is large and consists of many companies, of various sizes and businesses, who communicate a lot through a large range of newspapers, congresses, groups, etc. A non-comprehensive list of these newspapers comprises *L'Officiel des Transporteurs*, primarily oriented towards freight transport but which tackles always more with logistics, as carriers extend their perimeters towards this domain. *Logistique Magazine* is rather focused on companies whose core businesses pertain to the sector of logistics. It appears to the reader of these newspapers that the frontier between logistics and freight transport is becoming less tangible. Similar newspapers exist in other countries.

If we get towards less specialised media, the technical press is a rich source of information. In particular, the French economic newspaper *Les Echos* has proved particularly instructive. Aside from financial markets and firms' strategies, this newspaper pays a significant attention to the industry, which is not the case of other financial newspapers, such as *La Tribune* for example. Matters such as raw materials, the consequences of global warming in terms of regulation are often mentioned, as well as the general conditions of freight transportation. The reader often meets featured articles focused on logistic organisations, facilities, and markets.

Finally, we do not want to forget the mass media as one of our sources of information, and particularly the press[1] as well as the television. However, transportation seems to be seldom considered as a process on its own by these media; on the contrary it is often addressed from the perspective of one of its social impacts, such as congestion, accidents, social conflicts or, with increasing importance, global warming. In this context, the specific sector of freight transportation raises always more interest nowadays. Logistics is little present apart from common mentions to the increasing importance of just-in-time organisations in the industry and their impact on the freight transport market.

2.1.2 Methodology

Through parsing the wide range of information sources presented in the previous section, we have been able to identify a large set of approaches addressing a wide range of varied logistic issues. Despite the heterogeneity of these approaches, we managed to derive one classification, consisting of three categories.

It soon appeared that the proposition of a series of definitions as well as the delimitation of the perimeter of analysis was the necessary preliminary stage of this review. Therefore, Section 2.2 aims at addressing this objective.

Once these basic elements given, the second part of the review is structured by the fact that first, the approaches identified address problematics of highly heterogeneous scales, second that scale can be used as a way to organise this presentation. As a consequence, the presentation proceeds along increasing scales of concern and issues. Section 2.3 first presents works related to the logistic behaviour of a single firm. In such approaches, the universe in which the firm has to take its decisions is

[1] Let us quote, without claiming exhaustiveness, the following French newspapers of large audience and diverse leanings: *Le Figaro*, *Le Monde*, *Libération*

generally taken as given, *i.e.* exogeneous. It appears these issues are mainly of an engineering nature, and rely accordingly on engineering approaches. Section 2.4 is then devoted to examine relations between firms. These issues take on a strategic nature, and economic approaches are more adapted to addressing them. These approaches allow to relax the hypothesis of an exogeneous universe. We then present macroscopic approaches in Section 2.5. These approaches may involve macroscopic aggregate indicators, such as traffic or mode use, or large scale territories, such as regional or national areas. These methodologies pertain more to political concerns than to engineering concern, and the methodologies used include geography, statistics and economics. Such approaches are theoretically able to account for all kinds of relationships between a large set of varied actors.

As the shipment size constitutes a particular problematic of freight transportation demand modelling, as well as one of the focal points of this work, we have chosen to devote an independent section to this specific issue. The diverse approaches related to the choice of shipment size are thus regrouped and presented together in Section 2.6.

Our approach is summarised in Figure (2.1).

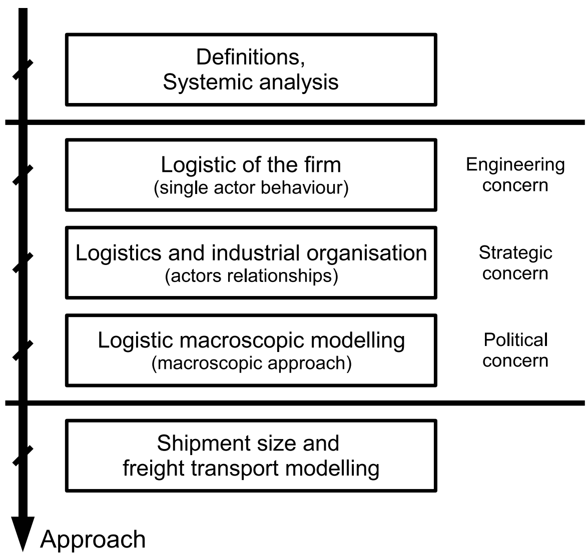

Figure 2.1: Roadmap of the approach.

2.2 Definitions, sector, agents

Finding a definition of logistics is complex. It can refer equally to a sector of activity, a profession, a function of the firm, or an academic field. We will provide highlights that should help deciding for a definition useful in the frame of this work, *i.e.* modelling-oriented. These highlights come from different perspectives.

A set of fundamental definitions are first presented in Section 2.2.1. We then focus on the particular concept of Supply Chain Management in Section 2.2.2, first on its nature then on the subsequent changes in the organisation of the logistic function in firms. We eventually show examples of analyses of the behaviour of agents of the logistic sector, in Section 2.2.3.

2.2.1 The logistic function in the firm and the supply chain: definitions

Goods or services[2] provide utility to consumers insofar as they are *available* to them. This availability pertains both to the location in space and time. Consumers can contribute to this availability, for example by fetching the goods or services they want directly where those are located. For example, they can grab a fruit on a tree they have in their garden, or they can go to a forest nearby and hunt to get some meat, provided it is legal to do so, as most of the economy had functioned for quite a while. In a somewhat more sophisticated way, they may also go to the supermarket and buy fruits and meat.

In the latter case, what consumers do is that they produce themselves the final stage of a possibly very sophisticated chain of transport and logistic operations, which will eventually allow them to dispose of the service or good they desire. This chain starts from a possibly large set of raw materials, comprises a possibly large number of transformation stages, together with a potentially even larger number of waiting, moving and conditioning stages. The former stages are of industrial nature, they may be referred to as *operations*. On the opposite, the latter ones can be referred to as to logistic stages, or *inter-operations*. As a consequence, one definition of logistics is: "the management of inter-operations" (Dornier and Fender, 2007).

Some authors (e.g. Carbone, 2004; Tatineni and Demetsky, 2005*a*)

[2]At first sight, it can seem irrelevant to refer to the logistics of services. Nevertheless, providing a service generally involves physical operations taking place at given place and time, and not anywhere.

use another definition, which may seem too wide at first sight[3]: "Logistics is the function of the firm in charge of managing physical streams and related[4] information and financial streams. This function consists in three tasks: physical tasks, administration, planning." This definition efficiently identifies the object of the logistic function of firms, *i.e.* physical streams. But it disregards the objective of the logistic function.

The French Association for Logistics ASLOG[5] gives the following definition:

Definition 2.1 *Logistics is the art and manner to provide a given commodity at the right time, right place, at the lowest cost and with the best quality.*

This definition clearly indicates the object of logistics, which is to ensure products are available to customers, and that the logistic process is cost efficient. By listing clearly the four criteria of time, place, cost, and quality, this definition also states implicitly the nature of the decisions which constitute the logistic function. Therefore, this definition constitutes a sound first step towards a microeconomic modelling of the logistic function. It will be retained for the sequel of this work.

From this definition of logistics, logistic operations can be defined from a top-down approach:

Definition 2.2 *A* logistic operation *is an operation pertaining to the logistic function of the firm.*

This definition is unambiguous and quite large. For example, an industrial operation such as the transformation of one good into another is not a logistic operation, but scheduling this transformation is a logistic operation.

However, this definition of logistics should be clearly distinguished from other concepts very commonly met such as supply chains and supply chain management. These concepts are almost always closely related to the strategic interaction of firms working together to deliver final products to end consumers. Indeed, before final products become available to customers, they are the object of many operations proceeded by many firms, which have to cooperate in a more or less intense way in order to provide a service. This set of firms constitutes the supply chain, which has been defined as follows by Christopher (1992) (in Carbone, 2004):

[3]This definition is at the opposite of narrow-scope definitions such as the one used in Daganzo (2005), who understands logistics as the science which studies how to convey items from production to consumption in cost-effective ways.

[4]"related" is ours

[5]*Association Française pour la Logistique.*

Definition 2.3 *The* supply chain *of a (set of) good(s) or service(s) is the network of organisations that are involved, through upstream and downstream linkages, in the different processes and activities that produce value in the form of this product or service in the hand of the ultimate consumer.*

It thus appears that the supply chain may constitute a useful basis to segment the market from a modelling perspective, particularly to understand how the final consumer's demands impact the whole supply chain upwards, and the resulting logistic requirements and freight transport demand.

Two remarks should be done about these definitions. First, the definition of logistics provided in this section makes no clear distinction between shippers, *i.e.* the agents who consume transport operations, and carriers, *i.e.* the agents who produce transport operations. This distinction is necessary to understand the complicated linkage between logistics and transport demand, and how this linkage should be taken into account in freight transport modelling. This deeper analysis is the object of Chapter 3.

Second, the notions of supply chain and supply chain management should be distinguished. While the concept of supply chain refers, as stated above, to an organisation of firms, the concept of supply chain management is quite different.

2.2.2 Supply Chain Management, integration and segmentation

As logistic processes and methods are key to the economic performance of the firm, a large body of literature, both professional and academic, has been targeted at searching for, and promoting, efficient logistics. Our main information source in this section is the work of Dornier and Fender (2007), as well as on a course given by Pr. M. Fender provided at the *Ecole Nationale des Ponts et Chaussées*, entitled "Supply Chain Management".

Supply Chain Management

The supply chain is of strategic importance for a firm, for the following reasons. First, goods and services provide utility to a consumer only if they are available to him. To the notion of availability corresponds the professional notion of *quality of service*, which accounts for the delivery time, the risk of stock-outs, etc.[6] Firms do not compete only in price and quality and other marketing-related variables, but also in quality of service. A firm which improves the availability of the goods or services it sells increases its competitiveness. This advantage depends on the willingness of its customers to pay for a better availability of the goods it sells (see Chapter 3).

Second, the profitability of the whole supply chain depends on the final consumer market. As a consequence, the quality of service provided by the most downwards firm of the supply chain is important not only to this firm, but to all the other ones.

Third, the cost to reach a given quality of service for a given product depends not only on the way each firm manages the logistics of this product in the frame of its own perimeter, but also on the way these firms cooperate. Indeed, in many instances, if each firm of a supply chain optimises its logistic without coordination with the other firms of the supply chain, the overall result is sub-optimal. For example, unexpected (or unpredictable) variations in the final demand, which are most common on several markets, may imply a lack of accuracy in the production plans of the ultimate firm of the supply chain. As these variations flow upwards the supply chain, their amplitude may increase, potentially resulting in huge variations, and even disruptions, in production plans at the lower levels of the supply chain, with the following undesired consequences such as shortages, over-dimensioned production units, uncontrolled costs, finally resulting in the whole supply chain being out of the market. This phenomenon, due to a lack of coordination and information sharing[7] between firms of a supply chain, is well known and often referred to as the "bullwhip effect" (Forrester, 1961)[8].

Fourth, firms are usually organised into departments (e.g. production, purchases, research and development, etc.), each with its own function. The optimisation of the logistic function, and of the supply chain requires that these distinct departments cooperate. This need for cooperation may be crucial, for example, between the transport department, the purchases department, and the production department, as the optimisation

[6]We set the hypothesis that a consumer is always able to decide which one he prefers between two firms which propose similar services except for the quality of service. It implies that the quality of service can be represented by a variable.

[7]Information sharing is a strategic question for a firm, and the trade-off between information sharing and lack of coordination is certainly not trivial. Besides, dampening the bullwhip effect does not necessarily imply more information sharing, as we will see later.

[8]This microscopic effect should be distinguished from the macroscopic effect of variations of stocks described in Metzler (1941), as will be explained in Section 2.5.2

of physical streams pertains to decisions which are traditionally taken by these departments, sometimes without coordination at all. The research and development department may also be associated, to take into account logistic imperatives when designing new products.

Fifth, the firm has to consider the supply chain it belongs to from a strategic perspective. The diverse firms in a supply chain may have asymmetric relationships, some of them dominating others. The position in the supply chain, as well as information availability, may play a strategic role. Indeed, all the firms of a supply chain not only consider how they can cooperate so that the supply chain is competitive as a whole, but also try to determine how they can extract the largest share of the added value yielded by the whole supply chain. The example of retail distribution is self-speaking: by taking control of the ultimate consumer market, distributors have reached a powerful position on the consumer goods supply chain. The internet crisis of 2001 is another good example of the strategic importance of supply chains: several arising firms neglected the necessity of an efficient logistic coordination with providers to provide competitive services, observed their logistic costs increase faster than their turnover during their full lifetime, until failure.

These issues, which were before under-addressed by firms, have been granted progressively always more attention for the last decades. Taking them into account is the object of a relatively young management field, called *Supply Chain Management* (SCM).

Definition 2.4 Supply Chain Management *is a management field devoted to taking into account the logistic and supply chain issues from a strategic perspective.*

As we will show in the next section, the Supply Chain Management has the particularity to focus on the way to take into account logistic issues in the frame of a firm whose historical architecture is not necessarily adapted to them. This imply first to focus heavily on the interfaces in the firm and between firms, second a lot of integration.

Integration

Historically, the logistic function was performed locally: either no optimisation was proceeded to, or these optimisations rested on local perimeters. There was not necessarily a dedicated logistic department in the firm. Those optimisation perimeters where often limited by a set of other frontiers, such as the various departments of a firm, even its units, or countries, and limits between firms. Reserves of productivity were left unexploited, since the logistic function often involves interdependencies between different departments, units, or even firms.

Today, not all the firms totally apply the SCM principles. Nevertheless, the management of the logistic function have been undergoing a series of changes for some decades. The optimisation perimeters have got larger, they have aggregated and have yielded larger perimeters, in a process that may be referred to as integration. Three types of integration can be distinguished:

- *Functional integration*: The classical segmentation of the functions of firms, *i.e.* the departments, now cooperate more than before in order to increase the overall efficiency of supply chains and, as a consequence, the profitability of firms.

- *Sectoral integration*: It is necessary for firms to cooperate along a supply chain to optimise its efficiency. In particular, firms need to share informations of possibly strategic importance. This exchange may happen in the frame of a partnership (e.g. vendor managed inventory[9] agreement between a producer and a distributor), or be imposed by the most powerful organisation on the supply chain.

- *Geographical integration*: international trade is less and less dampened by economic frontiers. As a consequence, firms that needed before to organise themselves country by country are now able to approach their markets on a regional, continental, or world basis. This has obvious consequences on freight transport demand.

These trends are changing deeply the way the logistic function is managed in firms, with large consequences on markets, on the way firms organise their production, their transports, etc. They may be further studied by getting into the detail of the agents' behaviours, and their relationships.

Segmentation

As a set of integrations allow firms to manage their flows more efficiently, with respect to the quality of service provided on the end consumer service, the better understanding of logistic issues also led to new, more relevant ways to segment and regroup issues of similar characteristics.

[9]In such a configuration, the vendor manages the purchases on behalf of the purchaser. This situation may be efficient when the demand for the product is highly variable, and when the marketing strategy of the vendor heavily relies on commercial actions such as sales promotion. This cooperation may yield significant inventory reductions, which is profitable to both parties. Such an organisation goes against deeply rooted habits.

One of the notions which appeared and is now always more widely used is the notion of logistic family. As firms get a better knowledge of their logistic imperatives[10], they also discover that some products are similar with respect to these imperatives. These similarities constitute the basis for a logistic segmentation, which allows to apply similar solutions to products which have similar requirements, and thus to fully benefit from increasing returns to scale of logistic systems (these increasing returns to scale pertain to the characteristics of logistic assets such as warehouses or vehicle fleets, or to the relatively higher simplicity and reliability of systems processing similar tasks, etc.)

The notion of logistic family constitutes a sound basis for segmenting the demand for freight transportation. Indeed, products which present the same logistic imperatives and, as such, are processed similarly in logistic systems, are likely to present similar characteristics when considered from a freight transportation modelling perspective. As a consequence, we propose the following definition.

Definition 2.5 *A* logistic family *is a set of products which present similar logistic imperatives.*

Logistic families are now widely used as a segmentation criterion in firms. For example, it is common to hear about ABC classification, where A refer to a set of homogeneous, cheap, and much sold products, whereas C refers to a set of heterogeneous, expensive products with a small turnover, and B refers to intermediary products. In a paint firm, A may contain the white paints, B the most common paints, and C some exotic paints such as waxes, limes, etc. Family A products make most of the turnover (but not necessarily most of the margin), and the logistic system associated with them is efficient and not expensive. On the contrary, family C products constitute only a small part of the turnover, and their logistic is expensive. But from a commercial point of view, the presence of the whole range of products is necessary. As a counterpart, the customers are ready to pay more and to wait more to buy C-kind products.

2.2.3 Logistic agents

As suggested by the definition of Christopher (1992), the supply chain of a given good consists of a set of agents. The supply chain is a complicated

[10] These imperatives, or constraints, can be the transport technologies which can be used to carry a given type of good, warehousing requirements (temperature, safety, etc.), lifetime, customer preferences with respect to the availability of goods, demand variability, etc.

concept from the perspective of freight transport modelling because it is transversal to classical delimitations. It implies several functions in the firm, and several firms of various sizes, kinds, and locations which coordinate their activities in order to provide a single given output. An approach towards understanding the supply chain is to examine the types and behaviours of the agents intervening on it.

Methods of systems analysis, such as employed in Savy (2006*b*), or of industrial economics and management sciences, such as used in Carbone (2004), are particularly relevant tools to operate such an approach. Savy investigates the roles of all the agents intervening in the frame of a logistic platform. His work focuses on their various time terms, their various objectives, their relationships.

Carbone's work is focused on the specific role of logistic providers. The strategy of the twenty first European logistic providers is studied, then a specific survey is led in Italy. Their objectives are examined, as well as their market segments and the relationships between them. Some conclusions are drawn, such as the heterogeneity of the logistic landscape between European countries, the specialisation of logistic providers according to their clients, in parallel to their strategies of growth, the trend towards an asymptotic non-asset based logistic provider business model [11], etc.

This analysis also reveals the diversity of relationships observed between logistic providers and their customers, and shows that their relationships may be described by their types and magnitudes. The type refers to a set of similar traits (activities, expectations, duration) whereas the magnitude of the relationship refers to its closeness or intensity. This is confirmed by the work of Golicic et al. (2003), who conducted a series of interviews under the form of round tables with experts. According to their conclusions, the relationships between the various agents of a supply chain are well characterised by their types and magnitudes[12].

[11] The notions of 3PL – *Third Party Logistic Providers* – and 4PL – *Fourth Party Logistic Providers* – are often met in logistic-related works. Whereas the 3PL concept refers to a company providing a logistic service, including transport and warehousing services, and potentially more sophisticated logistic services such as conditioning, co-manufacturing or inventory management, the less consensual 4PL concept refers to a theoretically non-asset based company, that would provide its clients an integrated logistic service by coordinating carriers and 3PL actors. Although the existence of the latter business model is questioned, a clear trend of the main European logistic providers towards supplying a larger range of more integrated logistic services, to increase their own profitability and to decrease their amount of assets has been observed between 2000 and 2004 (Carbone, 2004). This is why we consider that the 4PL notion makes sense, as a kind of asymptotic business model.

[12] This is an important issue of the theories of organisations, which try to explain the

However, getting into the detail of the relationships between agents is often a too complex and sophisticated approach for firms from the perspective of many issues. Even before thinking about optimising the supply chain they belong to, many firms try to optimise their own logistic.

2.3 Logistic problems of the firm

Before getting into the complexity of their strategic interactions with other agents, firms first have to organise efficiently their own logistics, while considering their environments as exogenous. In doing so, they meet a number of concrete issues of various time terms such as the location and movement of raw materials, intermediate products, final products, mobile resources (such as pallets, trucks, etc.) and fixed resources (such as plants, warehouses and machines)

To address these issues, firms can use a series of models of various natures, which are more or less relevant depending on the situation. The objective of this section is to present these models, and to discuss them from the perspective of freight transport demand modelling.

Some of the models firms use are very detailed: they operate on the basis of a technical, detailed description of the problem, and provide detailed, numerical, potentially immediately applicable solutions; they are often based on operation research theory. These models are discussed in Subsection 2.3.1. In some cases, approaches based on system dynamics, presented in Subsection 2.3.2, can be used. They are intermediary between the low-level, fully detailed models of operations research, and high-level, less accurate microeconomic models. This last category of models, presented in Subsection 2.3.3, is generally based on a simplified representation, and tries to allow for an intuitive interpretation of its outcome. After a brief discussion on how these methods are applied to take into account externalities in Subsection 2.3.4, Subsection 2.3.5 concludes this presentation of the modelling of logistic issues of the firm.

---

existence of firms and markets as modes of regulation. These theories also investigate the existence of hybrid forms, such as alliances. It was for example observed that alliances of various types existed. The two works just quoted brought evidence that it is not sufficient to consider only the type to fully describe an alliance. A second variable – magnitude – has been deemed necessary to do so. The question of whether it is sufficient or not remains open.

2.3.1 Low-level problems and operations research

Operation research (OR) is a field of mathematics dedicated to providing tools for decision support. OR problems generally imply an accurate (quantitative) depiction of the problem, and yield quantitative results, which optimise a given objective function. Typical OR problems in logistics include scheduling, planning, packing, routing, etc. and are often of combinatorial nature. In order to illustrate these problems and their corresponding methods, we present in this section a set of typical problems from which methods addressing various realistic situations derive. The considerations hereafter are mainly based on Korte and Vygen (2008). Examples of problems are taken from softwares' documentation and other sources.

The Traveling Salesman Problem (TSP)

A classical example of the *Traveling Salesman Problem (TSP)* is the case where an agent has to visit a set of destinations. The distance between each pair destination is known, and the objective is to find an itinerary minimising the total distance covered[13].

This problem is NP-hard, which means that when the instance is large, heuristics have to be used in order to find efficient solutions in a reasonable computation time. Such algorithms do not guarantee that the solution provided is optimal. According to Korte and Vygen (2008), the most successful way to obtain a good solution for the TSP in a reasonable time is local search, *i.e.* starting from a given tour then improving it with local modifications, such as cutting the tour into two pieces and joining them differently.

If the so-called agent is a vehicle (including a driver), the problem may be referred to as the *Vehicle Routing Problem* (VRP.) Complementary constraints may be added, in order to address varied problems:

- *Multiple-depot problem:* observe that constraining the starting position of the agent does not change the solution of the TSP. On the contrary, the problem may be enlarged to two or more agents

[13]The TSP is stated as follows:

Instance: A complete graph K_n and weights $c : E(K_n) \to \mathbb{R}_+$,

Task: Find a Hamiltonian circuit T whose weight $\Sigma_{e \in E(T)} c(e)$ is minimum,

where a complete graph is a set of vertices and undirected edges such that any pair of vertices is connected by an edge ($E(G)$ is the set of the edges of G, whereas $V(G)$ is the set of vertices of G) and a Hamiltonian circuit is a circuit with an itinerary that passes once and only once through each vertice.

starting from two determined points, having to visit all destinations once, at a minimum cost. This can be applied to choose the tour of a set of trucks starting from different depots to deliver an identical product to a set of locations.

- *Time constraints:* the travel time may be taken into account, and time constraints may be considered for the visits (e.g. stores opened at certain times or deliveries allowed during a given time-window) and for the depots. The total tour duration may be limited, and the operation time at each stop may be taken into account as well.

- *Ordered stops:* the order in which the stops are visited may be constrained. For example, there may be backhaul stops, where empty containers have to be picked up, which can be visited only after all the delivery stops have been made. There may be pickup and delivery stops, so that the vehicle capacity has to be considered. In the last case, one can furthermore assume that the vehicle is last-in-first-out (LIFO).

- *Dynamic routing:* the route may be optimised dynamically as information arrives on the fly. The corresponding algorithms are called *online algorithms*.

Softwares that propose routing algorithms include TransCAD[14], LogiX-central[15], ILOG[16] TransportPowerOps, the Outbound Transportation and Containerisation Optimisation solution of i2[17].

The Chinese Postman Problem

A classical example of the *Chinese Postman Problem* is the case where a postman has to deliver the mail within a given district, so that he has to walk along each street of this district, starting from and finally returning to the post office[18].

[14]www.caliper.com/tcovu.htm

[15]www.logixcentral.com

[16]www.ilog.com

[17]www.i2.com

[18]Under its general form, the problem is stated as follows:

Instance: An undirected connex graph G and weights $c : E(G) \to \mathbb{R}_+$,

Task: Find a function $k : E(G) \to \mathbb{N}$ such that G', the graph which arise from G by taking $k(e)$ copies of each edge $e \in E(G)$ is Eulerian[19] and $\Sigma_{e\in E(G)} k(e) \cdot c(e)$ is minimum, The Chinese Postman is in fact a particular instance of a more general combinatorial problem called the *Minimum Weight T-Join Problem*, which can be solved in $O(m^3)$ time, where m is the number of edges in the graph. The problem may be generalised to the consideration of directed edges, minimum number of passes through each edge, and so on.

Realistic situations described by this problem are not as commonly met as for the TSP; we can quote for the sake of illustration the example of door-to-door delivery such as mail, flyers, phone books, but also collecting trash or snow clearing. However, softwares do not advertise much this feature; an algorithm is implemented in TransCAD.

The Bin-Packing Problem

A classical example of the *Bin-Packing Problem* is the situation where one has to ship a set of shipments of given sizes. The objective is to ship them using a minimal number of vehicles of given capacity. Another case corresponding to the Bin-Packing Problem arises when one has a set of beams of given sizes and want to cut them into another set of given sizes; the objective is then to make the desired set using a minimal length of beams[20].

This problem is NP-hard, approximation algorithms are therefore necessary. There is a various set of such algorithms. Classical approximation algorithms include the *Next-Fit Algorithm (NF)* in which each piece a_i is packed in the current bin if it is possible, in the next one if not. The *First-Fit Algorithm (FF)* packs sequentially each piece a_i in the first bin where there is room left to do so. Both these algorithms behave well if the pieces are small relatively to the bin's capacity. The *First-Fit-Decreasing Algorithm (FFD)* is similar to FF apart that the pieces are sorted in decreasing order before proceeding.

Some generalisations of this problem include (see e.g. Hall, 1989):

- *More than one dimension:* the bin-packing problem may need to be extended to represent some situations, such as packing pieces

---

where an undirected graph G is an Eulerian graph if each vertex is connected to an even number of edges. It is therefore possible to find an itinerary in the graph covering each edge once and only once, in a time linear in the number of edges and vertices.

[20] The Bin-Packing Problem is stated as follows:

Instance: A list of non-negative real numbers $a_1, \ldots, a_n \in [0; 1]$.

Task: Find a $k \in \mathbb{N}$ and an assignment $f : \{1, \ldots, n\} \leftarrow \{1, \ldots, k\}$ with $\Sigma_{i:f(i)=j}\ a_i \leq 1$ for all $j \in 1, \ldots, k$ such that k is minimum.

in a vehicle for example. It may be necessary to consider 2D bin-packing, or even 3D bin-packing.

- *Online bin-packing:* it may be necessary to represent the case where the pieces to be packed arrive sequentially, without information on the subsequent items.

Among the softwares offering bin-packing algorithms, we find the ILOG software CP Optimizer, the Outbound Transportation and Containerisation Optimisation solution of i2.

The Minimum Cost Flow Problem

A classical example of the *Minimum Cost Flow Problem* is the situation where a set of uniform shipments must be transported from a set of origins (e.g. warehouses) to a set of destinations (e.g. final customers) to a minimum cost; or where empty rail cars have to be moved from their locations to the places they are required[21].

The solution is found by linear programming. A known particular case of the Minimum Cost Flow Problem is the *Hitchcock Problem*, where the capacities of the edges are infinite. A generalisation of the problem exists where the costs of the edges depend on the flow on the edges.

The following softwares propose minimum cost flow algorithms or solutions based on this kind of algorithms: TransCAD, ILOG LogicNet Plus.

The Facility Location Problem

There are numerous examples of the *Facility Location Problem*, which all are of strategic importance, as facility location decisions are usually not easily reversible. The decision may imply the location of distribution

[21] The Minimum Cost Flow Problem is stated as follows:

Instance: A directed graph G, capacities $u : E(G) \to \mathbb{R}_+$, numbers $b : V(G) \to \mathbb{R}$ with $\Sigma_{v \in V(G)}\ b(v) = 0$, and weights $c : E(G) \to \mathbb{R}$,

Task: Find a b-flow f whose cost $c(f) = \Sigma_{e \in E(G)}\ f(e) \cdot c(e)$ is minimum (or decide that none exists),

where a b-flow defines flows over the edges such that the flows entering a vertex are equal to the flows getting out of it – the flows are balanced, apart for the source vertices where the balance is positive and for the sinks for which the balance is negative. The function b defines the value of the balance at each vertex. The formal definition is: a function $f : E(G) \to \mathbb{R}_+$ with $f(e) \leq u(e)$ for all $e \in E(G)$ and $\Sigma_{e \in \delta^+(v)}\ f(e) - \Sigma_{e \in \delta^-(v)}\ f(e) = b(v)$ for all $v \in V(G)$, where $\delta^+(v)$ denotes the edges getting out of vertex v whereas $\delta^-(v)$ denotes the edges arriving in v.

centers making the interface between an industry and final customers, the location of break-bulk platforms in a distribution network, the location of a new plant, etc[22].

Under some hypotheses which are generally verified in realistic facility location configurations, the Facility Location Problem is NP-hard. Despite strong interest on this question since the 1960s, the first approximation algorithm has been proposed in 1997 by Shmoys et al. (1997). Local search techniques work well for facility location.

Algorithms solving this problem and generalisations constitute the basis of a series of softwares; basic versions of these algorithms may be found e.g. in TransCAD whereas softwares such as LogicNet PlusXE of ILOG, i2 Strategic Supply Chain Design solution, JD Edwards EnterpriseOne Supply Chain Management of Oracle[23] offer integrated strategic network design facilities which build on these kinds of algorithms.

General comments

First of all, note that the set of problems we presented is far from complete. Operation research is used to address a wide range of concrete issues such as production and transport planning and scheduling, inventory management, etc.

Second, these tools are seldom applied under their canonical form in an operational environment. Concrete situations imply accurate specifications, and the logistic decision support softwares are significantly dedicated to an important related task: database management. Historically, apart for some exceptions such as ILOG, of which the products have been designed specifically to identify feasible solutions in constrained problems and to improve such solutions towards optimality, the objective of this kind of softwares (called *Enterprise Resource Planning — ERP*), was to improve data availability and quality. More sophisticated features were then developed, and the softwares which offer them are generally referred to as *Advanced Planning and Scheduling Systems — APS*.

[22]The Facility Location Problem is stated as follows:

Instance: A finite set $\mathcal{D}$ of customers, a finite set $\mathcal{F}$ of potential facilities, a fixed cost $f_i \in \mathbb{R}_+$ for opening each facility $i \in \mathcal{F}$, and a service cost $c_{ij} \in \mathbb{R}_+$ for each $i \in \mathcal{F}$ and $j \in \mathcal{D}$.

Task: Find a subset X of facilities (called *open*) and an assignment $\sigma : \mathcal{D} \to X$ of the customers to open facilities, such that the sum of facility cost and service cost $\sum_{i \in X} f_i + \sum_{j \in \mathcal{D}} c_{\sigma(j)j}$ is minimum.

[23]www.oracle.com

Third, these problems also draw a significant attention from the academic world, and papers concerning them are regularly published in a wide set of reviews such as Transportation Research Part E., the Journal of Business Logistics, Logistics Information Management, the International Journal of Manufacturing System Design, etc.

Comments from a freight transport demand modelling perspective.

At first sight, the approaches and results of operation research seem far from spatialised freight transport demand modelling theoretic and practical concerns, for several reasons. We will present these reasons and explain how and where we may nevertheless find room for synergies.

- *Perimeter:* the question of perimeter is prominent when comparing firm approaches and freight transport demand modelling approaches. We address on one side problems of which the perimeter is limited to the firm, with the rest of the economic environment considered exogenous, whereas on the other side we are interested in large scale effects, which are, by essence, out of the scope of operation research methods, as much for data availability as for the low level of detail of the desired output of such models[24].

  It is however possible, under some circumstances, to use operation research tools to address medium-scale issues. For example, Groothedde (2003) has designed an intermodal freight transport network using both trucks and inland waterway transport which would satisfy the needs of a set of firms while achieving significant transport cost savings (see Chapter 1 for details). There are drawbacks though, in particular this research was made possible through provision of a large amount of proprietary data by a significant set of firms. Furthermore, we do not know how these results can be generalised.

  It is interesting to note that in the specific field of urban freight transportation modelling, operation research methods such as VRP algorithms seem more adequate than classical transportation methods to address urban freight transport modelling. Indeed, urban freight transport heavily relies on truck tours, along which the vehicles deserve a sequence of stops, which imply that the vehicles' origin and destination are somewhat disconnected from the goods'

[24] We obviously push back from our discussion the set of network algorithms which are equally useful in both contexts, such as shortest path algorithms.

ones. Spatialised transportation models, on the contrary, generally translate directly good flows into vehicle flows, using simple transformations based on factors such as average payload. Such an approach is therefore irrelevant for urban freight transportation modelling (Routhier et al., 2002).

- *Robustness:* the main piece of criticism towards operation research tools is, according to Daganzo (2005), that they deliver results of an accuracy which may be out of proportion with the accuracy of the inputs, especially when those inputs pertain to costs (either because these costs are not observable – this is the case of opportunity costs, or because they may vary significantly – such as market prices.) This is particularly problematic for freight transport demand modelling, both for the lack of accurate data and for the usually large time interval considered, on which costs may vary a lot and are famously hard to forecast.

  Daganzo (2005) presents a methodology devoted to produce nearly optimal, robust results based on models using few parameters, as we will see in subsection 2.3.3. However, his criticism is not completely relevant, as a set of methods of operation research address the problem of result robustness (*i.e.* results which remain good even if the situation does not correspond to the inputs) and inputs lack of precision (e.g. stochastic inputs.) Let us quote for the sake of illustration the work of Lee et al. (2007), which proposes a two-step optimisation of a joint forward-backward[25] distribution network with stochastic demand and prices. The proposed algorithm contains a nest optimising the flows for a given facility location, price and demand scenario, then proposes an optimal facility location on the basis of the total expected distribution cost.

- *Ease of use:* this question is twofold. It pertains first to data requirements, second to whether the results are intuitive or not.

  It is common that a realistic depiction needs a lot of accurate data for an operational research tool to yield satisfying results. However, freight transport demand modelling is a field of transportation sciences renowned for the difficulty to access to detailed data. The data usually available at regional, national or international scales is usually of aggregated nature, and as such is not immediately in the scope of the type of problems presented in this subsection.

[25] *i.e.* which handles both the forward flows and bakward flows, due to customers sending their deliveries back for example.

> The second drawback of OR tools is the lack of intuition on the results. The algorithms yield specific solutions, with little information pertaining to the linkage between the inputs and the result. As a consequence, the identification of strategic parameters and causalities is not straightforward, which makes the results hard to generalise.

The algorithms presented in this subsection are useful to support decision in the frame of a firm, or to optimise a collaboration between a limited set of firms. However, freight transport demand modelling is concerned with public policy decision support. This means that the modeller faces a situation of a much larger scale, and of a potentially extraordinary larger complexity. Because of the lack of data and of the need to identify the most strategic phenomena and the causality relationships between them, freight transport demand models have to yield clear results and clear intuitions of the linkage between results and inputs. As this is not the first objective of OR tools used in logistics, those would need some adaptation before they can be used in freight transport demand models.

2.3.2 Medium-level problems and system dynamics models

This subsection is devoted to two methodologies which may be used to draw particular highlights on some specific logistic problems: discrete system dynamics, and control engineering. They pertain are based on a more theoretical approach, where the representation of the issue is simplified.

Discrete system dynamics

The discrete system dynamics approach (usually simply referred to as the "system dynamics" approach in the papers we quote) consists in representing the system studied by a set of variables of time $t \in \mathbb{N}$, and a set of discrete equations describing the values of these variables at $t+1$ function of their values in t.

To illustrate this approach, let us present the Automated Pipeline Inventory and Order Based Production Control System (APIOBPCS) model. Designed by John, Naim and Towill (1994), it describes a system where the order decisions are based on the forecast demand (based on exponential smoothing), a fraction of the difference between target and actual inventory levels, and a fraction of the difference between target and actual goods-in-transit levels. Figure (2.2) provides a basic illustration of the model.

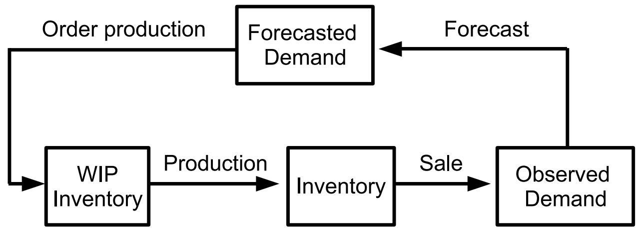

Figure 2.2: The architecture of APIOBPCS.

This framework is able to illustrate the bullwhip phenomenon. Indeed, consider a random demand series, where the daily demand t is i.i.d. of given variance σ_d. Feeding this series into the system results into an order series of given variance σ_o. The ratio of these variances σ_o/σ_d is called the *amplification ratio.* If it is higher than 1, then a bullwhip effect is identified: perturbations amplify upwards the logistic system.

This framework is used by Potter and Lalwani (2007) in order to assess the influence of this effect on transport demand. The question is considered both from the perspective of the fleet size and of the loading factor of the vehicles. The approach provides little information though, apart from intuitive results such as the probable need for a greater fleet if the variability of the shipment's size is greater.

Control engineering

The control engineering approach consists in modelling the behaviour of the supply chain with tools of the theory of signal. Such tools are the use of transfer functions, frequency response curves and spectral analysis. It is thus possible to consider the supply chain as a system transforming the demand input signal into the supply output signal.

This approach has been used in Dejonckheere et al. (2003) in order to give a theoretical explanation of the bullwhip effect. The logistic system is modelled using a set of variables depending on the time period t and a set of linear[26] equations describing the values of these variables in t as functions of their values in $t-1$ and the shipment need at day t. The transfer function of the supply chain is derived, allowing the calculation of the *frequency response plot (FR)*, which gives the ratio of the standard deviations of the output and the input. The FR plot is interesting in that

[26] The linearity of the system ensures the relevancy of the approach.

a value of more than 1 reveals an amplification of variability in the logistic system, *i.e.* a bullwhip effect.

Using this approach, Dejonckheere et al. (2003) are able to show the influence of the order policy rule on the bullwhip effect. They first consider the classical order-up-to policy with exponential forecasting. The order-up-to policy is:

$$O_t = \hat{D}_t + k\hat{\sigma}_t - \text{inventory position}_t$$

where D_t is the demand at day t, $\hat{D}_t$ its estimator, $\hat{\sigma}_t$ the estimator of its variance, k a constant ensuring a given level of service[27]. The exponential forecasting formula forecasts future demand based on passed observations is:

$$\hat{D}_t = \alpha.D_{t-1} + (1-\alpha).D_{t-2}$$

The second ordering policy is the smoothing policy, first introduced by Towill (1982). It is not an order-up-to policy. It is rather based on forecasting the demand and controlling simultaneously the net stock level and the work-in-progress level[28]

The FR plots of both these policies illustrated by Figures (2.3) and (2.4) show that the second policy is globally efficient in dampening the oscillations inside the logistic system, apart from low frequencies. On the contrary, the first policy systematically implies an amplification of the variations. However, although we tend to think that less variability is a good thing, the work of Dejonckheere et al. (2003) also indicates that the smoothing policy probably implies higher inventory costs. The next step, *i.e.* making a microeconomic trade off between these two outcomes, is not done.

These approaches are interesting as they allow for the use of a large set of powerful analysis tools. Furthermore, implications of findings such as the smoothing rule of Towill (1982) may be significant in an operational environment.

However, the transferability of these results is questionable, and the various trade-offs are not explicited. In particular, the influence of some parameters which are both basic and strategic, such as costs, is not clear.

[27] Assuming the demands of each day are identically distributed random variables, provided that the estimator $\hat{\sigma}_t$ is accurate enough, a given constant k is equivalent to a given probability that there is no stock-outs.

[28] The exact formula is the following: $O_t = \hat{D}_t + 1/T_n \cdot (DNS_t - NS_t) + 1/T_w \cdot (DWIP_t - WIP_t)$, where the demand is still exponentially forecasted, NS_t is the net stock and TNS_t the desired net stock level, and WIP_t it the work-in-progress stock and $DWIP_t$ the desired work-in-progress level. T_n and T_w are parameters.

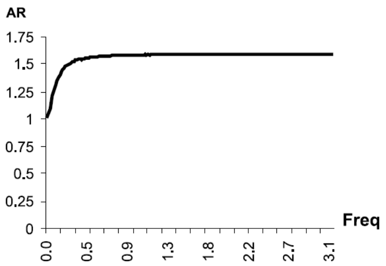

Figure 2.3: FR plot in case of an order-up-to policy.

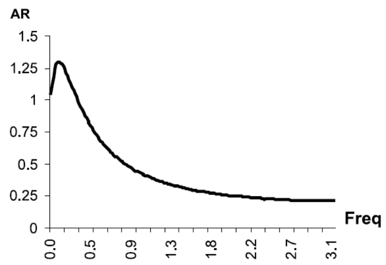

Figure 2.4: FR plot in case of a smooth ordering policy.

The next part is devoted to presenting methodologies focused on deriving near-optimal solutions depending on few parameters.

2.3.3 High-level models

We present here a set of methodologies which address operational logistic issues with light models, based on generic descriptions and few parameters. These models do not aim to provide an detailed optimal solution to a given realistic and accurately described situation, such as it is done in subsection 2.3.1. They aim both at providing a correct solution to a realistic situation, and at giving insights on the linkage between the solution provided and some strategic parameters. A wide range of issues can be addressed this way, from small-scale logistic operating problems to large-scale logistic network design problems. This subsection rests heavily on Daganzo (2005).

The Economic Order Quantity model

The first model to present is the simple Economic Order Quantity (EOQ) model, a classic in the inventory control litterature. Originally developed by Harris (1913), the credit for its first in-depth analysis usually goes to Wilson (1934).

The EOQ model can be applied to a wide range of issues, among which to find the optimal shipment size in the case of a simple supply chain. Consider a continuous production of commodities at a given origin, at rate Q. These commodities must be shipped to a given destination, where they are also consumed at rate Q, using a vehicle of given operating costs and capacity S. We assume that travel time are reliable, therefore there is no need for a safety stock.

Denote b the cost of dispatching a vehicle, and t the travel time between the origin and the destination. Denote a the value of travel time savings by unit of commodity. If the goods are shipped by bundles of size s (note that $s \leq S$), the average time a unit of commodity waits in the origin inventory before being shipped is $s/2$. The average pipeline inventory level (the amount of commodities being carried at a given instant) does not depend on s. Therefore, the total cost per unit of commodity, denoted $g(s)$, is:

$$g(s) = \frac{Qb}{s} + \frac{1}{2}as. \tag{2.1}$$

Note that $as/2$ is replaced by as if the commodities are consumed regularly at the destination. In that case, each commodity waits on average $s/2$ in the destination inventory. The qualitative properties of the model remain.

The economic shipment size is obtained by minimising $g(s)$ with respect to s:

$$s^* = \min\left\{\sqrt{\frac{2bQ}{a}}; S\right\}. \tag{2.2}$$

It is interesting to note that the total cost per unit function is robust with respect to s: a variation of 10% in the shipment size implies a increase of about 4% of the total cost. The converse consequence is that even if the parameters are known with uncertainty, or are liable to evolve, the solution provided by the EOQ model remains reasonably efficient. Furthermore, the formula is easy to handle and gives insights on the linkage between the optimal shipment size and the cost parameters. However, some of the hypotheses stated above reduce the generality of the model.

Continuous approximation

Models such as the previous one have a limited generality if they rely on hypotheses of uniformity of some of their parameters. This uniformity, which is generally necessary for the analytic resolution of the equations, is particularly irrelevant when one considers the variability of parameters such as demand (in time) and density (in space).

It is in fact possible to overcome this limitation, using the *continuous approximation* method. This methodology, first applied to transportation problems by Newell (1973), has been extensively used, notably by Daganzo (2005), to address a wide range of logistic problems. It applies when one searches a discrete solution (schedule, facility location) to a problem where one or more parameter varies smoothly.

For the sake of illustration, we will apply this methodology to extend the EOQ model. Assume $q(t)$ varies over a given time window I, the solution with uniform shipments has no reason to be optimal. One thus looks for the times $\{t_i\}_{i\in\mathbb{N}}$ at which vehicles should be dispatched, which is equivalent to deciding the headways $H(t_i) = t_{i+1} - t_i$. Consider the total cost per unit $C(t)$. Assume q is continuously derivable, then for all i, t_i^* exists such that $q'(t_i^*)$ is equal to the average slope $(q(t_{i+1}) - q(t_i))/(t_{i+1} - t_i)$. Then, the total cost per unit of good between two shipments is equal to:

$$\int_{t_i}^{t_{i+1}} \left(\frac{c_v}{H(t_i)} + \frac{c_w H(t_i)}{2} q'(t_i^*) \right) dt.$$

This is where the continuous approximation is performed. First, we replace $q'(t_i^*)$ by $q'(t)$ in the formula; this is valid if q varies smoothly, as assumed. Second, we consider the continuous function $H(t)$ instead of the discrete set of headways $\{H(t_i)\}$. As a consequence, we solve the following extremely simple variational problem:

$$\min_{H:I\to\mathbb{R}} \int_I \left(\frac{c_v}{H(t)} + \frac{c_w H(t)}{2} q'(t) \right) dt,$$

which yields:

$$H(t) = \sqrt{\frac{2c_v}{c_w \cdot q'(t)}}.$$

This continuous function is then discretised, and it is shown on some examples that the solution performs very well, provided q' does not vary too quickly.

This methodology is powerful in that the intuitive nature of the results remains, whereas it provides near-optimal solutions in realistic frameworks. Such an approach may be a good trade-off between accuracy and robustness, especially when there is uncertainty on some parameters.

Interface with OR models: statistical regularities

Continuous approximations may be applied in a wide range of situations. However, some problems such as distribution in a spatialised framework still cannot be addressed without using OR models. As a consequence, a method has been designed in order to interface these models with continuous approximation methods. The idea is to divide the main problem into a set of micro problems, all of which can be addressed using combinatorial algorithms, and then to use known statistical regularities of the solutions of these algorithms in order to build a both easy to handle and realistic approximation.

Consider a complex problem: a firm which delivers at home a large set of customers. Assume the deliveries are small, so that an efficient organisation implies that the vehicles make rounds. Therefore, we are in a combinatorial framework. Furthermore, consider the customers are not uniformly spread on the territory studied. Our objective is to define the best location for our distribution centers, taking into account installation and distribution costs.

We first consider the distribution costs of a single depot delivering a given set of customers. Denote the density of customers on the area deserved δ, s the average shipment size, S the vehicle's capacity, and r the distance of the depot to the area. Then it is possible to prove that the average length of an optimal tour is approximately (Daganzo, 2005):

$$2r + \frac{k \cdot S}{\sqrt{\delta}.s},$$

where k is a constant approximately equal to 0.82. As a consequence, it is quite easily possible to derive analytically the expected distribution cost from a depot to a given region of density δ.

The continuous approximation can then be used on δ in order to decide how many depots should be installed and where, so as to minimise the overall distribution costs.

Nevertheless, statistical regularities are not always easy to identify. Let us quote the work of Hall (1989) as an example. In that paper, vehicles make rounds in order to deliver sets of shipments of stochastic sizes to customers. The objective is to provide insights on the influence of bin-packing rules on the distribution costs, and on the trade-off between the average loads of the vehicles and the average route length. Hall could not derive analytic formulae, thus limiting the outreach of this approach in this case.

Applications

Daganzo (2005) examines a series of frameworks of increasing complexity. For each of these frameworks, a simple model is designed. Then, the limiting hypotheses are relaxed one after the other, as far as possible. The problems thus addressed are:

- *One-to-One Distribution:* in this simple framework, cost-effective ways to ship a flow of goods from a given origin to a given destination are examined. This results in the EOQ model and its generalisation presented previously.

- *One-to-Many Distribution:* the previous framework is enlarged in order to consider several destinations. The problematics of vehicle routing appears. The analytic approach reaches its limits, as a consequence particular cases are presented when some costs are negligible before others. This has the advantage to help identifying the strategic parameters.

- *One-to-Many Distribution with Transshipments:* the framework is enlarged to consider potential transshipments in a break-bulk platform.

- *Many-to-Many Distribution:* it is the most general framework in Daganzo (2005). However, strong hypotheses are necessary to lead an analytic approach.

Similar methods are presented in a number of papers. Let us quote, for example, Blumenfeld et al. (1985) who examines the various networks listed above, and possible simplified ways to derive optimal shipment schedules when there are many interdependencies (such as production of different products, consolidation centers, etc.), and Blumenfeld et al. (1991), who consider a one-to-many framework and examine the amount of cost savings allowed by various synchronisation policies. Similarly, Hill (1997) examines jointed production-shipment policies, in a particular framework where the inventory holding cost is higher for the buyer than for the vendor[29]. The result is interesting in that although the flow considered is of constant rate, the optimal policy implies non-uniform shipments.

Commentary

The methodology is efficient in reaching the objectives announced at the beginning of the subsection: light models involving a reduced set of strategic parameters, yielding robust results, that may be applied in operational contexts, and yet provide valuable insights into the causalities at stake.

[29]This may be the case when, for example, the buyer is a retail seller located in a dense urban area where renting warehouse surface is expensive, whereas the seller is located in an area where renting warehouse is cheap.

The small amount of data necessary to use this methodology, as well as the intuitive nature of its results, let us think that some of its elements can be applied with benefit to large-scale, policy-oriented freight transport demand models.

However, in the examples presented here, the focus is set mainly on the logistic of one firm, or, if several agents are explicitly considered, on the globally optimal setup. The influence of the strategic behaviours of distinct agents within a supply chain is not considered. The influence of the nature and intensity of the competition on the market of the final good neither. The approach is always based on the minimisation of a total logistic cost function, but some of these costs are either taken as is, without explanation (such as opportunity costs), whereas other costs are simply neglected (costs of shortages). The level of service is not considered, which is problematic given its central importance in the whole structuration of logistic systems. After a short discussion on the role of externalities in logistics microscopic modelling, we present in section 2.4 how these questions may be addressed.

2.3.4 Externalities

Logistics microscopic modelling, as we saw in the previous subsections, is mainly focused on fulfilling cost-effectively the requirements of the logistic function of the firm. This is the reason why very few papers use these approaches in order to address issues which are out of this scope, such as externalities (*i.e.* impacts of decisions taken by agents on other agents who are not involved in these decisions).

One example of such works is Anciaux and Yuan (2007). The authors examine the transfer along a various set of sequences of transport operations, involving systematically road transport, and potentially air, rail and inland waterway transport. They assess, for each of these alternatives, the transport costs, transsshipment costs, the total travel time, the environmental pollution, noise pollution, and risks caused. The shipment size corresponding to each of these alternatives is derived endogenously, from an EOQ-type model. They identify the best alternative with respect to each of these criteria, but do not synthesize the analysis.

Examining the question of externalities from this scope is interesting insofar as these externalities can be corrected by incentives, and these incentives influence behaviours. As a consequence, it is necessary to understand behaviours. But the microscopic scale considered in this section neglects the relationships between actors. These relationships may play a major role in the response to incentives, they should be considered with utmost attention from a policy decision support perspective.

2.3.5 Conclusion

The logistic issues of a firm, even when its economic environment is taken as given, are complex and of various natures. This is reflected by the wide range of tools which can be used to address these problems. Depending on the accuracy of data, low-level models of operations research can provide detailed solutions to clearly stated problems. Such models are particularly useful to run complex processes where the costs and requirements are precisely known.

Medium-level models of system dynamics help understanding some non-trivial effects of given logistic policies, in particular the ability of a given system to resist to variations of exogenous parameters such as prices and demand. These approaches have the advantage over low-level approaches to help diagnose the impact of choosing a given logistic policy in terms of inventory levels and inventory level variations, and how these variations propagate in a supply chain.

High-level models can be used for more strategic decisions where data is less accurate and where insights on the relationships between decisions and outcomes are valuable. They both provide satisfying approximate solutions and highlight the trade offs at stake. They represent in a simple and not very data demanding manner the linkage between logistic requirements (the need to proceed to a given set of operations), costs (including transport costs, and inventory costs), and the resulting trade offs between logistic resources such as warehousing and transport. As such, they constitute an interesting basis for the microeconomic modelling of the behaviour of shippers.

2.4 The influence of logistics on industrial organisation

In order to model freight transport demand in a realistic way, one has to consider logistics from the perspective of firms. This motivated the presentation int the previous section of the set of low-level approaches firms apply when they consider their economic environment as exogenous. Complementarily, from a high-level decision support perspective, attention must be paid to the interactions between agents.

These interactions result from the strategic behaviours of firms, and are difficult to model. Both analytical and numerical approaches involve strong assumptions so as to yield meaningful results. In a very simplistic first approach, we can consider that the strategic behaviours of firms have two types of drivers: non-cooperative ones, *i.e.* reasons for firms not to cooperate, such as market competition, or cooperative ones, *i.e.* reasons for firms to cooperate, such as cooperation for the efficiency of a whole supply chain.

Figure 2.5 is a simplified illustration of some of the drivers which structure supply chains. Based on Christopher's definition, the starting point is the end market. Then, we distinguish n much simplified parallel, competing supply chains. Each of these supply chains is simplistically represented as a linear set of k firms, starting from the farthest of the end market, indexed 1, to the nearest one, indexed k. We represent three drivers of strategic behaviours in Figure 2.5: competition between supply chains, which is a non-cooperative driver, cooperation in a supply chain for the quality of service to be the most cost-effective in the final market, which is a cooperative driver in the frame of a supply chain, and finally margin sharing in a given supply chain, which is a non-cooperative driver. Indeed, if firms coordinate their activities so that their supply chain is globally efficient on the final market, they are able to withdraw a higher margin. The remaining question is then: how is this extra margin shared among the firms of the supply chain?

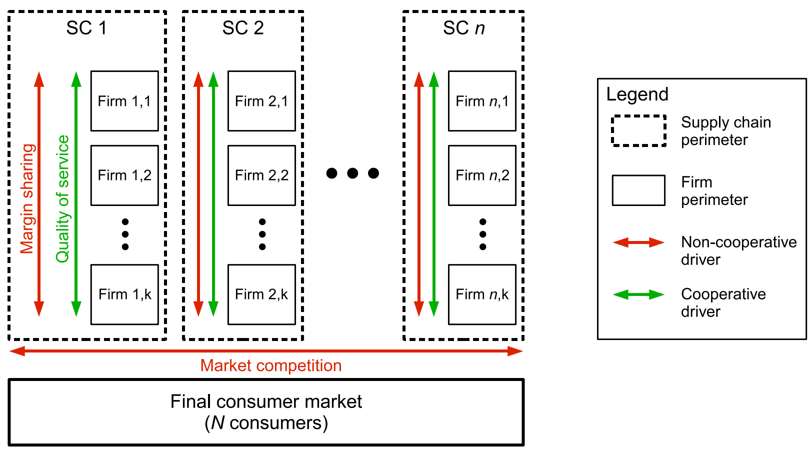

Figure 2.5: Drivers underlying the strategic behaviours of firms.

We present in Subsection 2.4.1 works which focus on competitive, non cooperative behaviours, and in Subsection 2.4.2 works focused on cooperative behaviours.

2.4.1 Competitive behaviours

One of the prominent elements structuring the relationships between agents is the relative position of the final product on its market. Significant attention has been paid to competitive relationships by microeconomists, and have led to classical models such as the perfect competition model, Cournot-Nash oligopolistic competition in quantity, monopolistic competition, competition in price and quality[30], etc. These models have been generalised in many ways. We present here some generalisations pertaining which take into account logistic elements.

Competition in final product availability

In theoretical microeconomics, the basic model of the firm is focused on a firm which faces a demand made up of agents who are only interested in the price of the good the firm sells. This treatment is then extended to account for other variables, typically characteristics of the commodities considered[31].

The availability of a good to the final consumer, *i.e.* its location at the place and time of need, can be considered as a variable of the direct utility function of consumers. As a consequence, firms providing goods compete with each other in the availability of the goods they provide. Ensuring a given level of availability is the object of the logistic function. A better availability means higher logistic costs, but this linkage is not trivial. For a more detailed analysis of this trade-off, see Chapter 3.

This issue was investigated by Chopra (2003), which studied a series of supply chain configurations and compare them with respect to availability variables (e.g. order lead-time, shortages probability, return possibility, customer information) and to costs (e.g. inventory level, number of transport operations). Although there is no quantitative analysis and the trade-offs are not formally stated, this analysis shows that logistic characteristics play a significant role in market competition.

[30]Examples of references concerning these models are Tirole (1988) and Anderson et al. (1992).

[31]Lancaster (1966) was the first to propose a consumer theory in which the direct utility function of consumers does not derive directly from the good(s) the consumers are granted with, but from its (their) characteristics. This approach is based on the idea that the value of this good from the agent's perspective derives from its characteristics. Among other consequences, this theory provides a sound basis for modelling substitution effects.

Supplier-receiver interaction

At first sight, it is legitimate to ignore supplier-receiver interactions in the microeconomic analysis of a supply chain. Indeed, under the assumption of perfect competition, costs for a supplier are identical to market prices for a receiver, so that these interactions do not influence the economic functioning of the supply chain.

Nevertheless, this (often underlying) perfect competition assumption is not always realistic. It is therefore useful to relax this hypothesis, and to investigate more closely supplier-receiver relationships. Assume, for example, that the supplier disregards the overall efficiency of the supply chain it belongs to. Assume also that this supplier is in charge of choosing the transport mode by which the commodities it sends to the receivers will be carried. Then the supplier will probably prefer a cheaper transport mode to a more reliable one, against the preferences of the receiver[32], as the shipper considers it does not bear the costs incurred by this lack of reliability. This is the kind of approach chosen by Winston (1981). Winston considers two cases: either the receiver chooses everything (FOB pricing), in which case he maximises his utility, or the shipper chooses the mode (CIF pricing). In the latter case, it is assumed that both agents bargain according to a Zeuthen-Hicks model (Bishop, 1964), so that their joint utility[33] (*i.e.* the sum of the utilities of the two agents) is maximised.

Winston's work thus reduces to the maximisation of the joint util-

[32]Why should the receiver valuate more reliability than the shipper? A basic argument is that as stock-holders can diversify their portfolios, firms should be risk-neutral. Winston disproves this argument by saying that as risk-neutral as stock-holders may be, managers are not, since their jobs are at stake.

[33]Consider two agents, whose behaviour is characterised by utility or profit functions which depend on their decisions (e.g. two firms competing in a Cournot-Nash framework). A subset of these decisions are Paretian. Nevertheless, there is an undefinitely high number of Pareto decisions. It is therefore assumed that a bargaining process takes place. The Zeuthen bargaining model basically assumes that the agents make rational concessions up to an equilibrium, which maximises the product of their utilities (Picard, 2007). This result is consistent with the analysis of Nash (1950), in a more general framework. Nash further argues that when monetary transfers are possible between agents, then the solution of the bargain game maximises the sum of the utilities of the agents, and each agent has the same utility (note that it is assumed that the agents are equally skillful in negotiation), hence the result used by Winston, which is in fact not present in Bishop (1964). The Hicks model has been developed on another basis; it is an asymmetric model of the negotiation between employers and unions, which has the particularity to take into account a temporal dimension. Bishop introduces this temporal dimension in the Zeuthen model to build his composite model, which he calls the Zeuthen-Hicks bargaining model.

ity (*i.e.* minimisation of global costs), but it has the merit to represent explicitly the agents implied in the transport operations and their interaction. As such, this framework enables one to model some imperfect competition effects, which are known to yield some counter-intuitive results.

Competition in a spatial framework

When addressing logistic issues, the spatial dimension of the economy plays an important role. It appears through the availability of goods, which has been analysed in the previous section, and also through the spatial location of firms. From this latter perspective, generalisations of classical competition models to a spatialised framework are interesting.

Harker (1986) presented some generalisations. This approach consists in taking the hypotheses of the classical models and giving them a spatial dimension. Consider a set of regions, each one described by a supply and a demand function on each commodity market. Consider generalised transport costs[34] between these regions. The equilibrium conditions of these models are classical: first, the supply must equal the demand within each region. Second, the effective prices must be uniform within each region. Formally, if region i imports a good from region j, then $p_i = p_j + t_{ji}$, where p_i denotes the price in region i and t_{ji} the unit transport price from region j to region i; else $p_i \leq p_j + t_{ji}$.

Such models are referred to as *Spatial Computable General Equilibrium (SCGE)* models. This denomination insists on the objective of their designers to develop models which can be solved numerically, and which can address large-scale perimeters. Among other features, these models address explicitly the issue of the structures of the distinct commodity markets.

One of the difficulties in using these models is to dispose of generalised transport costs reflecting realistically their effect on trade. Combes and Lafourcade (2005) studied how these transport costs should be modelled. They took the example of freight transportation by truck, between pairs of regions in France. They based their analysis on the detailed accounting of costs such as fuel, wages, vehicle operating and depreciation, etc. They showed that in general, total distance and total time, and even as the

[34]We knowingly use this classical term of transport economics out of the scope of its usual definition. Our objective is to indicate that the difference between a good available in the region of the consumer at a given time and a good available in another region at the same moment pertains not only to freight transport rates, but also to travel time, amount to be shipped, sensitivity of the consumer toward delivery times, etc.

crow flies distance constitute excellent proxies for transport costs, up to a multiplicative constant. However, transport direct costs constitute but a small part of generalised transport costs, which encompass many other logistic costs. These other cost components are neglected in this analysis.

These models have been build to forecast interregional or international trade flows, notably as to provide inputs to freight transport models. Nevertheless, they currently have two main flaws. First, the representation of freight transport costs and, more generally, of the freight transport market, is generally oversimplified. Second, logistic issues are absent from these models. As explained in more detail in Chapter 3, they play a crucial role in the formation of freight transport demand; freight transport demand cannot be explained alone by the production and consumption of commodities.

2.4.2 Cooperative behaviours

As argued in Section 2.2, the whole supply chain of a given good or service is implied in the quality of service this good is supplied with. This raises two questions. First, how should firms cooperate in order to improve the overall efficiency of the supply chain they belong to? Second, on which basis should firms set the limit up to which they are willing to cooperate?

The first question seems to imply, by its formulation, that the basic situation is characterised by an absolute lack of cooperation between firms. On the contrary, situations of cooperation, implicit and explicit, already exist. For example, delivery and pickup time windows are almost uniform between firms, so as to limit complex coordination operations. Hensher and Puckett (2005) present modelling recommendations focused on taking such interactions into account in urban freight transportation models. Among other features, their proposition has the originality to take into account explicitly concepts such as relationship type and magnitude (see Subsection 2.2.3) in the model's architecture.

The choice of transport mode may also be considered as an interaction. This is the approach of Holguin-Veras et al. (2007), who consider the relation between a shipper and a carrier as a cooperative game (*i.e.* a game where a non-cooperative decision is never advantageous). They led an experimental economic approach, where agents playing respectively the roles of shippers and carriers decide the shipment size and mode, and showed that the market coordination mode was efficient in opting for the most advantageous joint choice[35]. Nevertheless, the outreach of

[35] Experimentally confirming the theoretical analysis of Winston (1981) presented above.

this work is limited by the simplicity of the decisions involved.

Cooperation in a supply chain may indeed be difficult. This is the reason why SCM took as much importance as a management field. This also induced deep thinking on the ways firms should cooperate. One of the methods investigated and used by firms is the Vendor Managed Inventory (VMI) coordination mode. Consider a supplier (e.g. a productive firm) selling to a receiver (e.g. a retail distributor) a product which will ultimately be sold to consumers. If the receiver lacks information on these products or has little control on the commercial policy of the supplier (e.g. if the supplier advertises using mass media), it may be appropriate to set up a VMI organisation, in which it is the supplier who decides the amount of goods to be delivered to the receiver[36].

The classic and VMI organisations have been compared in a number of works, among which Disney et al. (2003). Disney et al. use a dynamic system approach based on the APIOBPCS model (described in subsection 2.3.2) in the frame of a supply chain consisting of a manufacturer and a retail center, in order to show that a VMI organisation allows transport costs savings, compared to a classic organisation.

Cooperation is possible in a number of ways, more or less relevant with respect to the circumstances. Firms may find an overall benefit in sharing information, designing jointly the products they sell, etc. Nevertheless, even if the best strategy was known in each case that may arise, questions would still be left unanswered, in particular: how would the overall benefit be shared among firms? One may forecast, without taking too much risks, that this benefit is generally shared in a way that leaves at least one firm unsatisfied. This leads to the second question we raised at the beginning of this subsection, pertaining to the margin sharing out.

2.5 Logistics macroscopic modelling

Policy issues arise in the logistic sector because of its impacts on the rest of the economy (pollution, transport infrastructure use, land use, but also economic efficiency, etc.) To observe sectors and assess the potential need for regulation, administrations may avail themselves of a set of tools for building strategic diagnosis. Such tools may be applied to the analysis of the logistic sector, although with some methodological difficulties.

Some pre-modelling approaches are presented in Subsection 2.5.1. We

[36] It is needless to say that this idea raised some apprehension before it was finally experimented and deemed successful in a number of supply chains, e.g. between l'Oréal and Carrefour in France for cosmetics.

then give an illustration of a classical macroeconomic approach accounting for the linkage between logistic and economic issues in subsection 2.5.2. Subsection 2.5.3 is dedicated to the spatialised analysis of the logistic activity. Lastly, we present how the logistic imperatives of firms translate into transportation demand in Subsection 2.5.4.

2.5.1 Pre-modelling approaches

Despite the large stream of academic and professional literature pertaining to logistics and SCM, there are striking gaps in the knowledge of this sector of activity. Strongly linked to the economic efficiency of the territory, an efficient logistic sector is a factor of attractiveness in the international economic competition. On the contrary, by the transport flows it generates, it is the source of negative externalities: noise, pollution, risk of accident, congestion, etc. Despite the intensity of these issues, institutions have little knowledge of logistic as an activity on one hand, and as a driver of the structure of a territory on the other hand. We present here some first steps towards a better understanding of both questions.

The logistic sector and the logistic business segment

Our first task should be to provide a definition of the notion of sector. However, this proves difficult and no general answer will be provided to this issue; a relevant and interesting discussion about this matter can be found in Carbone (2004). We only recall that the notion of sector in national statistics is based on the homogeneity of supply (*i.e.* the use of similar production techniques.)

This is a convenient definition to answer a large amount of questions, particularly from a macroeconomic perspective. As a consequence, national statistics usually rely on this definition. On the contrary, it proves weak in addressing the field of logistics. In fact, even the identification of the freight transport sector proves difficult using this approach, since a lot of firms still carry their goods by themselves (own account transport), and do not identify this activity as freight transport. The situation is even worse in the case of logistics, as logistic activities are not much externalised.

This fact was stated by Savy (2006*a*), who used the notion of business segment to assess the importance of the logistic activity in the economy. French firms are committed to state their activity, which is used to define the sector(s) they belong to. In addition, they also have to declare the number of people they employ and their functions[37]. People are assigned to categories along the job they do, and these categories are the business segments. Using this notion, which is transversal to the notion of sector, it is possible to measure how many people work in transport, transport forwarding, warehousing, or material handling-related, but not conditioning-related. 2M people were working in the logistic business segment in 1999, among which two out of three were working in firms whose activity was neither freight transport, nor logistics.

These figures were confirmed by a more recent study by Mariotte (2007), distinguishing clearly the transport and the logistic business segments. The logistic business segment was defined as people working in warehousing and services associated to warehousing. According to this definition, 700k people work in the transport business segment, 800k in the logistic business segment, among whom only 22% worked in the logistic sector, 15% in the wholesale trade sector or as trade intermediaries. These figures confirm that the logistic function is not much externalised yet, at least not as much as the freigh transport. The study also located these jobs, as illustrated in Figure (2.6), taken from Mariotte (2007). This study thus constitutes a first approach towards a geography of logistics.

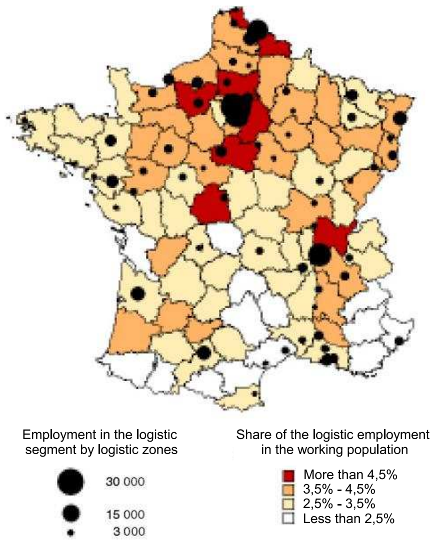

Figure 2.6: Employment in the logistic segment in France in 2007

[37]DADS dataset: *Déclarations Annuelles des Données Sociales* (yearly social data declaration).

Geographic analysis

The above mentioned macroeconomic approach can be usefully complemented by a geographic approach, this quite naturally since geography focuses on space, and space is a crucial dimension of logistic issues.

In a work for the DIACT[38], Savy (2006*b*) provided an analysis of the interaction between logistics and territory. His work includes first a series of interviews with a set of agents; second, a geographic analysis.

This analysis considers three objects. First, a series of maps of the flows of goods in France were drawn by commodity type[39], allowing to identify types of regions and interactions, as well as main flows. Then, the ratio of the goods getting in and out of a region to the internal flows were calculated, and analysed together with the movements of palettized goods per region and per capita. This allows to identify the type of logistic activity of the different regions, and the heterogeneity of the territory with respect to the integration in logistic flows. Finally, a map of the total areas of warehouses and of the number of related workers was drawn, in order to illustrate the specialisation of some regions and the concentration of logistic activities in particularly active zones.

This work allowed to draw conclusions on the logistic attractiveness of the French territory, and the heterogeneity of this attractiveness between French regions. Possible reasons to this heterogeneity were presented, together with a set of recommendations. As a conclusion, this approach is original as it takes into account explicitly and simultaneously the spatial dimension of the territory and the logistic function of the economy.

2.5.2 Macroeconomic analysis

Macroeconomic analysis is based on a high-level, aggregated depiction of the economy, which encompasses many elements pertaining to all the economic sectors. Besides, the spatial dimension is generally not at the center of macroeconomic approaches. From this perspective, logistic issues are thus mainly peripheral.

However, the particular case of inventory levels is interesting. Indeed, in the case of big economic shocks, for example during economic crises,

---

[38] *Direction Interministérielle à l'Aménagement et à la Compétitivité des Territoires*, land planning and territory competitiveness interministry department.

[39] Using the SITRAM database, *Système d'Information du Transport de Marchandises*, database on freight transport. This database is maintained by the *SOES* (formerly *SESP*), statistic service of the *MEEDDM*, French ministry in charge, among other fields, of transports. It describes freight flows in tons, commodity type, origin and destination, transport mode and conditioning

many firms slow down or stop their production as long as their current inventory levels allow it, implying a spectacular decrease of their activities. These kind of effects, related to the logistic function of firms, are discernible at a macroeconomic scale. Due to our limited knowledge of macroeconomic methodologies and literature, the models addressing this phenomenon will not be presented here. Therefore, this presentation is limited to one of these models, for illustration's sake.

When Keynes presented his theory of general equilibrium, he limited his analysis to a static framework. Shortly after, a series of works were conducted to investigate its dynamic properties[40]. It is indeed possible to consider that a variation in the consumers income does not impact immediately consumption, whereas one can symmetrically assume that it is the production side which exhibits inertia. Metzler (1941) takes a step towards accoutning for logistics in his model, by assuming that there are inventories, so that if the production level does not match the sales level, the level of inventories variate correspondingly.

Metzler considers a number of scenarii, which are varied by the assumptions about the sales forecasting policy as well as by the inventory replenishment rules. He identifies that under particular circumstances, a perturbation can lead to an overshoot of the inventory level, oscillations, and even explosion of the system. As a consequence, he identifies, at the macroscopic level, a phenomenon which is similar to the bullwhip effect presented in Section 2.2.2.

The hypotheses underlying this work are strong though, since it is assumed that all the firms have the same replenishment rules and the lag between production and sale is the same for all markets. Furthermore, this particular work is focused on the reaction of a whole sector to a somewhat exogeneous perturbation, before the equilibrium is reached again[41]. As a consequence, we cannot consider this work to represent the bullwhip effect defined by Forrester (1961), which refers to an increase in the variability of the size of orders upward a given logistic system.

This work builds on an aggregate description of the economy, without taking into account the spatial dimension. The next subsection presents works focused on this aspect.

[40] Notably by Samuelson, who also developed the well known IS-LM general equilibrium model (Guerrien, 2002).

[41] The perturbation considered is a step variation of the level of non-induced investment of firms.

2.5.3 Spatial analysis

The location of the fixed assets of firms is, in itself, a logistic problem. Nevertheless, and despite the very large number of them, the works tackling this issue virtually never refer to, or set themselves in the frame of, the field of logistics.

We present in this subsection some of these approaches, and the typical problems they allow to address. They can be classified into two categories, depending on whether they focus on the location of the distribution, or on the location of the production.

Spatial distribution of the demand

One of the first steps one can make towards modelling the influence of logistic considerations on spatial structure is to take into account the distance final consumers have to travel to reach a given good or service. This was done first by Hotelling (1929), who considered a duopoly where two firms compete on a single, linear street, on which consumers are uniformly spread. The objective was to build a model where market shares were continuous variables of prices, by introducing another product characteristics: the distance to the retail shop. Hotelling argued that distance could be replaced by any characteristic liable to vary continuously among customers[42]. Nevertheless, he concluded that firms tend to concentrate, which is of particular importance in understanding the spatial structure of the economy[43]. One can find several generalisations of this problem in Anderson et al. (1992): more than two firms, simultaneous entry or sequential entry, choice of pricing policy (f.o.b. or uniform pricing), etc.

However these models are often doomed with non-existence of an equilibrium. To overcome this problem, they can be generalised to considering price, location and other product characteristics. It is thus shown that firms can choose to regroup or to move apart depending on the relative sensitivity of the customers to the price or to other characteristics, also called preference for heterogeneity.

This phenomenon is illustrated by Figure (2.7), taken from Anderson et al. (1992), in the case of three firms, on a road of length l. z_i^* de-

[42]Such as, for example, the utility for each customer of each of these two products; in which case this model is an instance of address model (Anderson et al., 1992).

[43]It is also important to note that the spontaneous behaviours of the firms in this framework do not guarantee a socially efficient outcome. Their concentration lead to an average distance covered by customers way larger at the equilibrium than at the social optimum. This incapacity of the market to reach a social optimum is a common trait of models considering product differenciation (Anderson et al., 1992)

notes the equilibrium position of firm i, a function of μ, the preference for heterogeneity of consumers, and τ, the unit generalised transport cost. Three situations appear: for small values of $\mu/\tau l$ (*i.e.* when the spatial dimension is relatively more important for customers than the preference for heterogeneity), there is no equilibrium[44]. Then, for higher values of $\mu/\tau l$, an equilibrium exists, where the firms are spread along the road. They get closer as the preference for heterogeneity of the consumers increases compared to the cost of transport, and finally choose to concentrate at the same place when $\mu/\tau l$ is high enough. This means that when the geographic characteristic of goods is of little importance relatively to their idiosyncratic characteristics, the monopolistic market power a firm can withdraw from going far from the other firms does not compensate the customers it loses doing so.

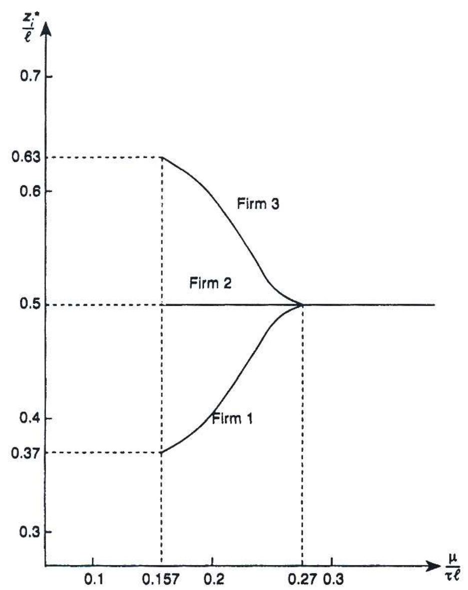

Figure 2.7: Optimal location.

This explanation of aggregation or separation of firms as an effect of the relative importance of the location of a good from the perspective of customers gives a particularly important insight on the linkage between logistic characteristics of the goods and the spatial structure of the economy.

[44]This phenomenon was identified by Hotelling.

Spatial distribution of production factors

However, a firm does not decide the location of its fixed assets on the basis of its customers' locations. At the opposite side of the logistic system, the location of production factors is important as well. This issue is central to many models of international trade.

The first of these models was developed by Adam Smith, who explained trade as the result of the difference in absolute productivities from a region to another. The international trade theory of Ricardo, based on comparative advantages, proved to be more realistic (Krugman and Wells, 2006). According to this theory, regions specialise in the sectors of lowest opportunity cost in their respective economies. The Hecksher-Ohlin-Samuelson model (presented in Helpman and Krugman, 1985) is a general equilibrium model which leads to similar conclusions, but explains them not by exogenous differences in productivities, but by exogeneous differences in production factor endowments, these production factors being immobile.

These theories have been generalised by the fruitful introduction of scale economies in production, and product differentiation (Helpman and Krugman, 1985). Such models are able to explain both intersectorial trade (which was already correctly accounted for by classic foreign trade models) and intrasectorial trade (which was, by construction, irrelevant in classic models). These models represent the complex relationships between firm location (strictly speaking capital location), worker location, prices, and wages. Many hypotheses can be made: workers can be assumed immobile or mobile, or partly mobile (the so-called "agricultural" workers being immobile and the others mobile), capital can be assumed immobile or mobile, etc. Much attention is paid to how the equilibrium evolves when transport costs change, and a feature generally shared by these models is that when transport costs decrease, the economy gets progressively more specialised: a so-called "center-periphery" structure appears. This asymmetry is then reduced when transport costs get even lower. For a review of these models, see Fujita et al. (1999) or Combes et al. (2006).

It should be noted that in these models, the description of the spatial dimension keeps quite basic. In many cases, the economy is described by two points separated by a given distance, and transport costs are described as "iceberg" costs, *i.e.* a fraction of the amount of commodities carried is consumed by the transport operation.

Evans and Harrigan (2005) proceeded to a more realistic approach from a logistic perspective. Their work was focused on the specific role of travel time in a supply chain when the demand is not known with certainty, and on how firms decide their location in such a framework. Evans and Harrigan model the economy as a general equilibrium. Firms sell textiles in the United Stated on two periods. They have the possibility to locate their factories in Asia, or in the Caribbean. If they locate their factories in the Caribbean, these firms can adapt their production: they produce before the first period the goods they will sell during the first period. Then, they observe the sales realised during the first period, and decide the amount they produce for the second period, considering the inventory left from the first period. On the contrary, if the factories are located in Asia, transport times are too large to do so. As a consequence, firms have to produce once and for all the amount that they will sell over the two periods. Therefore, they face a higher demand uncertainty, so that the expected amount of unsold products and the expected amount of unsatisfied customers is higher.

When examining the equilibrium of this model, it appears that the firms which are faced with the most uncertain demand locate at the Caribbean. As a consequence, the wages are higher at the Caribbean than in Asia. But, despite these higher wages, it is beneficial for the firms located in the Caribbean to be close to their market. The conclusions of the theoretical model are confirmed by an empiric analysis of aggregated textile trade data as well as by an econometric analysis of proprietary data of a textile manufacturer[45]. This approach prefigures the specific signification of the value of time in the frame of a supply chain; see Chapter 7 for details.

Comments

These theories, which are instances of the microeconomic approaches used to explain spatial structure, provide insight on elements of the supply chains: the end market, and the issue of retail selling points, or the choice of the best region for production.

Nevertheless, their strong linkage with the logistic issues of firms is seldom advertised. Furthermore, there is no synthesis of these two approaches to our knowledge; such a synthesis would be a second step towards taking logistic requirements into account in economic models.

[45]It is also interesting to note that, as flexible production is made possible thanks to the advances of the technologies of communication, as stated by Evans and Harrigan: "*It turns predictions about the* 'death of distance' *on their head: in our model, improvements in communications technology make distance matter* more *for income, not less.*"

2.5.4 Logistic drivers of transport decisions

In connection with the spatialised freight transport models which are addressed in Chapter 1, let us focus on the linkage between the logistic and economic behaviours behind such decisions as the choice of mode or itinerary.

Mode choice

It has been generally aknowledged that the choice of transport mode for a shipment depends heavily on the logistic requirements of the given shipment. Yet not much has been built on this observation. First, it is hard to identify and to measure these logistic imperatives. Second, the complex linkage between the role and relationships of a firm in a supply chain and the logistic imperatives of its shipments has not yet been much investigated.

Pre-modelling approaches do exist, such as the work by Woodburn (2003). The author interviewed a set of firms. He identified their respective roles in their supply chains, and asked them about the predictable trends for their own supply chains[46] , together with their assessment of their potential for a shift towards rail. He thus identified the following trends expected by the agents within the supply chain: first, the risks associated with the increase of transport costs are of small magnitude, since all competitors bear them equally. Second, the use of rail is small but this may come to change. Third, customers' requirements will get stronger, and maybe more prone to variate. The logistic system will have to be changed accordingly. The interviewees acknowledged that redeploying a logistic system is difficult, and the flexibility of road is an advantage in the prospect of a redeployment. Note that the second and third statements of the actors are contradictory. The recent trend reveals that in the trade-off between end-consumer requirements and transport efficiency, firms shift towards the demand side.

Such qualitative approaches are hard to extrapolate to quantitative modal shift potentialities, but they are very useful in that they allow to assess its drivers and trends, to a certain extent. This work also indicates that methodological innovations can yield a much better understanding of logistics, to a small incremental cost: it was easy to identify the role of the firm in the supply chain, this provided much understanding.

[46]It is interesting to note that the most important trends is the expected increase in the service demands made by the customers, and the shift toward more flexible organisations – particularly just-in-time organisations.

Itinerary choice

The issue of itinerary choice is of central importance in freight transport modelling, itinerary decisions are the microeconomic causes of traffic flows. Understanding the microeconomic drivers of this decision is thus a crucial problem of freight transport modelling. But the linkage between logistic requirements and itinerary choice are not fully understood yet.

Many elements play a role in the decision of itinerary choice: time-dependent operating costs, distance-dependent operating costs, depreciation costs and opportunity costs of the freight transported, and other logistic costs. However, the econometric approach to this issue seldom distinguish these costs. The itinerary choice decision is usually reduced to a trade off between transport price and travel time.

Some works tried to address this simplistic representation by introducing other variables. For example, the agents willingness to pay for a reduction in travel time variability can also be taken into account. This issue is investigated by Fowkes et al. (2004). The authors performed a stated preference survey in order to assess the agents' willingness to pay so as to avoid respectively: a delay in the departure time of the vehicle (*i.e.* scheduled delay, e.g. the vehicle has broken down), a delay in the arrival time of the vehicle once the vehicle has left the origin (*i.e.* unscheduled delay, e.g. the vehicle is stuck in a traffic jam), and finally variability in the arrival time[47]. The results show that the highest willingness to pay is to avoid unscheduled delay. It amounts to about 100€ (end-2000) per hour, whereas the willingness to pay to avoid delay is about 60€ per hour, which is intuitively meaningful since a situation with unscheduled delay is more constrained than with scheduled delay, where the use of another itinerary or another mode is conceivable.)

These results constitute a basis for further modelling refinements, and highlight the complexity of the preferences of shippers towards freight transport alternatives. Furthermore, these results make a first step towards the issue of departure time choice in freight transportation, which is still under-addressed.

[47]Note that the definition and assessment of travel time variability is problematic in itself, and that this is not specific to freight transportation. For example, de Jong et al. (2004) reviewed works about the willingness to pay of passengers to avoid travel time variability. They notably insist on the difference between choosing a definition of variability (e.g. standard deviation versus mean ratio) and proposing options of various variabilities in SP studies. In particular, the standard deviation versus mean ratio is often too complex a notion for the respondents, which means that it has to be translated into more illustrative indicators (e.g. the difference between the 80th and the 90th centiles of arrival times.)

The disadvantage of these methodological elements is that they ignore one of the basic difference between passenger and freight transportation modelling. The load unit in passenger transportation is indeed the passenger: as such, it cannot be modified, and this makes a great difference with freight transportation in which the size of shipments is a decision variable. This degree of freedom, which is specific to freight transport, has not yet been accounted for in freight transport models yet, although some works have already addressed it. The next section is focused on some of these works.

2.6 Shipment size in the frame of transport demand modelling

Freight transport demand is generally represented by a set of flows, expressed as aggregated indicators, e.g. tons per year. And yet, there are many reasons to think that the shipment constitutes the unit decision in freight transportation, and that it should be used as the modelling unit when possible[48], in order to bridge the gap between traffic modelling methodologies and logistic behaviours of firms. Taking into account the shipments characteristics is thus a key direction of improvement of freight transportation demand models.

The shipment size decision is difficult to analyse and to model, first because of the lack of data[49] , second due to the limited knowledge concerning the microeconomic drivers of firms decisions and specifically the linkage between their logistic imperatives and their decisions pertaining to shipment characteristics on the one hand, and macroscopic knowledge of costs, trends, and practices on the other hand.

We present pre-modelling methodologic elements such as database collection and descriptive analysis in Subsection 2.6.1. Subsection 2.6.2 is focused on the use of shipment-related variables in econometric appro-

[48]This question is discussed in more detail in Chapter 3. Besides, Routhier et al. (2002) argue that, faced with the particular organisation of urban freight transportation, it is statistically more efficient to consider the movement (*i.e.* the vehicle movement between two stops, even if these stops are part of a route) as the decision unit than the shipment or the origin-destination flow when modelling urban freight transport.

[49]Data about shipment size are difficult and expensive to collect, particularly due to the difficulty to follow a shipment accross the sequence of transport operations it goes through from its shipper to its receiver. Furthermore, such data are often confidential, due to strategic information they contain with respect to the type, amount and value of the freight exchanged and to the firms implied in the transaction.

aches. We then introduce some research works investigating the micro-economic drivers of shipment choice in Subsection 2.6.3. We eventually present choices and recommendations concerning shipment size choice in freight transportation demand models in Subsection 2.6.4.

2.6.1 Pre-modelling approach

Prior to any quantitative modelling approach, data availability is required, first to allow phenomena identification and investigation, second to calibrate potential subsequent models. We present some examples of databases, and then some trends for France.

Databases

Shipment databases are not available in all countries, and when they are available, they are not systematically renewed every year (see below the French case).

Shipment databases provide more or less accurate data on shipments. The definition of shipment is not straightforward. In the following of this work, we use the following one, adapted from Guilbault et al. (2006) [50]

Definition 2.6 *A* shipment *is a given amount of freight of a given type, handed over at given place and time, by a unique shipper in order to be carried as a whole towards a given, unique receiver.*

Instances of shipment databases include the American *Commodity Flow Survey (CFS)*, which is collected every five years (Holguin-Veras, 2007), and includes about 5 millions of shipments. A CFS is also available in Sweden, with information on about one million of shipments, for years 2001 and 2004/2005 (de Jong and Ben-Akiva, 2007). These databases contain information such as shipment origin and destination, size, value, commodity type, etc.

The French CFS, named *ECHO*[51], includes a unusual amount of information on each shipment, particularly pertaining to the organisation along the shipment movement. As described in Guilbault et al. (2006), the 2003-2004 survey concerns 10,000 shipments emitted by about 3,000 shippers, which is relatively small when compared to other shipment

[50] The original definition is the following : "*la quantité de marchandises remise à un même moment à un même endroit par un chargeur unique, pour être transportée dans sa totalité vers un destinataire unique*". The translation, hopefully accurate, is ours.

[51] "*Envois-CHargeurs-Opérateurs de transport*", that we may translate as: "Shipments-Shippers-Carriers".

databases. This low sample size is compensated by the large amount of information collected with respect to each shipment, including notably:

- *The attributes of the shipment:* commodity, size and weight, distance, value, desired travel time, costs and service level requirements,
- *The sequence of operations:* physical characteristics (origin, destination, conditioning, transport operations, logistic operations, energy consumption, travel cost and time) as well as organisational characteristics (number of agents intervening, decision levels, subcontracting relationships),
- *Shipper and receiver attributes:* economic characteristics (activity, turnover, relationships) as well as production characteristics (organisation of the production, information system) and physic characteristics (accessibility to transport infrastructures, amount shipped per year.) The shipper-receiver relationship is also described.

These databases constitute a basis to understand the linkage between freight transport demand and the logistic function of firms. Furthermore, even the simplest shipment size microeconomic models require the observation of specific variables, which cannot be derived from classic freight transport databases Given the large number of variables observed in the ECHO database, it is conceivable to proceed to the econometric assessment of some of these models, a task which is impossible with other, simpler CFS (see Chapter 5).

On the whole, the snapshots classic freight databases give of freight transport systems are both partial and biased. For example, these databases are often focused on vehicles movements. This is the case of the French database SitraM. As a consequence, if a shipment is moved from its origin to its destination through several vehicle movements with transshipment operations in between, the itinerary of the shipment cannot be observed. For this reason, the amount of freight moving on the French territory from or towards a foreign country is underestimated.

Trends

The considerable amount of detail available in the ECHO database allows an extensive descriptive analysis of the freight transport system, which is useful at a pre-modelling stage. Among other statements, this database allows the identification of a deep trend towards lower shipment weights between 1988 and 2003.

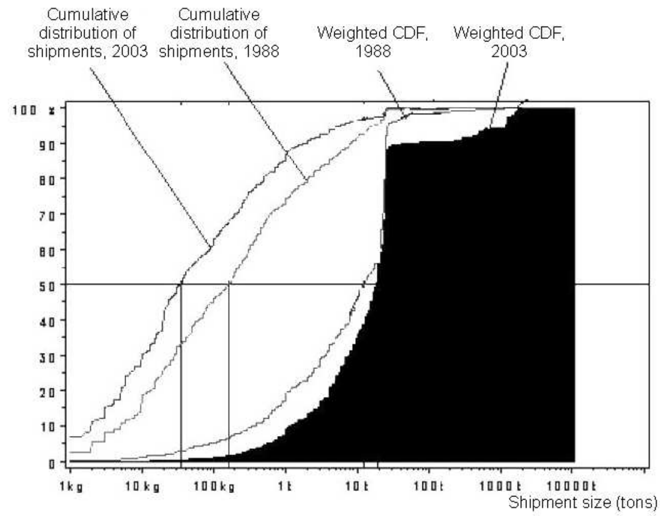

Figure 2.8: Distributions of shipment sizes between 1988 and 2003, taken from Guilbault *et al.* (2006)

Figure 2.8 shows the cumulative distribution of shipment sizes in 1988 and in 2003, and the cumulative distribution of shipment sizes weighted by the size variable (measured in tons). Let us notice that there are thresholds in the weighted cumulative distribution functions, located at the values of maximum vehicle capacities. This indicates that some shipments sizes are very probably constrained by vehicle technologies. The presence of flat, horizontal areas in the weighted cumulative distribution functions also confirms a theoretical result of Hall (1985), see Subsection 2.6.4 for details. Hall (1985) states that given freight rates, some shipment sizes are avoided by shippers. This is the case, in particular, of shipment sizes close, but not equal, to the capacity of vehicles.

The evolution of the distribution of shipment sizes is confirmed by statistical summaries such as the median size, which falls from 160 kg to 35 kg between 1988 and 2003, and the weighted median which increases from 12.6 ton to 19 ton. Such an apparently paradoxical statement is interesting. It is tempting to interpret it, in a first approach, as an overall increase of logistic systems efficiencies, both being able to send smaller shipments when needed and in using the whole capacity of vehicles in order to move freight cheaply as much as possible. This issue is addressed from a more theoretical perspective in Chapter 6.

The analytic stages following the descriptive analysis are generally based on technical tools. Indeed, taking into account the shipment size variable in microeconomic approaches or in large-scale freight transport models proves quite difficult. On the contrary, when this variable is observed, it can prove useful in econometric approaches. As a consequence, we now present how this variable has been used in various econometric studies.

2.6.2 Econometric analysis

Shipment databases constitute a basis for the identification of statistical regularities. In particular, they can be used in order to refine the analysis of issues such as freight transport demand forecasting, modal choice, or shipper-carrier interaction

Freight transport demand forecasting

Using the shipment as the decision unit raises a certain amount of theoretical and practical problems. One of them is that classic demand forecasting methodologies cannot be used straightforwardly to forecast demand at the level of the shipment.

Based on proprietary data provided by a single firm, Garrido and Mahmassani (2000) proposed an econometric model to estimate the number of loads[52] sent from each zone of a given area. They estimated a multinomial probit specification, with spatial and temporal autoregression[53], where the latent variable associated to each zone consists in an alternative-dependent constant and the first factor given by a principal component analysis of the socio-economic attributed of those zones.

The dataset used has been granted by a major U.S. carrier company, containing information on a large amount of loads: origin, destination, earliest/latest pickup time, earliest/latest delivery time, distance travelled, revenue earned. The model is estimated against a subset of this dataset, corresponding to the loads picked up in the state of Texas (16,287 loads). This sample is divided into four subsamples corresponding to the four seasons. Three of these subsamples are used for estimation, whereas the fourth is hold out to assess the estimated model. There are 80 zones and 2 time intervals (0-12am and 12am-12pm.)

[52] Loads are not exactly shipments but constitute a first step towards this unit.

[53] Denote $\varepsilon_t = \{\varepsilon_{it}\}$ where $i \in 1, \ldots, R$ and $t \in 1, \ldots, T$ the error terms pertaining to each zone/time interval alternative. Denote W the contiguity matrice, where w_{ij} is 1 if zone i and zone j are contiguous and 0 else. Consider μ_t the dynamic process defined as $\mu_t = \alpha\mu_{t-1} + \lambda_t$ where λ_t is a vector of independent, centered normally distributed processes λ_{it}. It is therefore assumed that $\varepsilon_t = \rho W \varepsilon_{t-1} + \mu_t$, thus representing the spatial temporal autoregressive nature of the model.

The estimation indicates a negative spatial autocorrelation, thus revealing the presence of high-demand clusters surrounded by low-demand areas. The temporal autoregressive coefficient is significant and lower than 1, indicating the existence of a dynamic effect and the temporal stationarity of the error term. However, the estimated model's performance on the hold-out sample is mitigated. A Chi-squared test rejects the hypothesis that the expected number of loads and the hold-out sample values follow the same multivariate normal distribution. On the contrary, the non-parametric Spearman's rank text rejected the no-agreement hypothesis with more than 99% confidence. This illustrates the difficulties of such an approach.

Modal choice

Shipment characteristics are difficult to forecast when forecasting the whole freight transport demand, as illustrated by the previous example. On the contrary, they are easier to handle when focusing on modal choice for a given demand. A number of works have investigated the possibility to specify an econometric model considering both the shipment size and the transport mode. We present two of these works, which provide an illustration of the methodologies used.

Holguin-Veras (2002) examines a series of model specifications which are summarized here. The objective of the model is to forecast the choices of shipment size and vehicle in the frame of urban freight transportation. To do so, a shipment size submodel is derived:

$$y = \left(\beta_0 + \sum_{i=1}^{n} \beta_i \delta_i\right) ln(D) + \eta,$$

where y is the shipment size, D is the distance travelled[54], δ_i are binary variables representing the commodity group and the activity type, η the error term.

The vehicle choice is then modelled as a discrete choice. The utilities of the vehicle options are:

$$U_i = \alpha_{1i} + \alpha_{2i} C_i^w + \alpha_{3i} XL_i(y) + \varepsilon_i,$$

where C_i^w is the average unit cost per ton, the α are parameters to be estimated, and XL_i denotes a sort of distance between the shipment size

[54] As it will be explained in the sequel of this chapter and in Chapter 5, there are strong theoretical and econometric reasons to believe that shipment size and transport distance are at most loosely related. This may explain why these models do not fit the data perfectly ($R^2 = 0.46$ for the most complete specification.)

and the average payload M_i of the vehicle of type i:

$$XL_i = |M_i - y|,$$

thus representing the linkage between the shipment size and the vehicle mode choice [55]. ε_i is the error term, it has been successively specified as a multinomial logit error term and as a heteroscedastic extreme value error term. No correlation has been assumed between η and the ε_i, which implies that the mode choice/shipment size process is modelled as sequential, *i.e.* that the choice of transport mode is conditional to the choice of shipment size.

The approach developed in Abdelwahab and Sargious (1992) overcomes this limitation, in assuming that these error terms are dependent. The framework set up by the authors is not the same as in the previous work, since they focus on the choice between rail and truck. They consider the choice of transport mode and shipment size are simultaneous, and opt for the following specification:

$$\begin{aligned} I_i^* &= \gamma Z_i + \varepsilon_i, && (2.3)\\ Y_{1i} &= \beta_1 X_{1i} + \varepsilon_{1i} \;\text{ iff }\; I_i^* > 0, && (2.4)\\ Y_{2i} &= \beta_2 X_{2i} + \varepsilon_{2i} \;\text{ iff }\; I_i^* \leq 0. && (2.5) \end{aligned}$$

Equation (2.3) represents the mode chosen (truck if $I_1^* > 0$, rail otherwise), whereas equations (2.4) and (2.5) represent the shipment size conditionally to the choice of respectively truck or rail. X_{1i} and X_{2i} are exogenous variables; Z_i consists of some or all the exogenous variables in X_{1i} and X_{2i}, and also additional exogenous variables. Y_{1i} and Y_{2i} are endogenous. The three error terms ε_i, ε_{1i} and ε_{2i} are centered, serially independent, trivariate normally distributed. Their variance-covariance Σ is:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1\varepsilon} \\ \sigma_{12} & \sigma_2^2 & \sigma_{2\varepsilon} \\ \sigma_{1\varepsilon} & \sigma_{2\varepsilon} & \sigma_\varepsilon^2 \end{pmatrix}.$$

The parameters to estimate are β_1, β_2, γ, σ_1, σ_2, $\sigma_{1\varepsilon}$, and $\sigma_{2\varepsilon}$. Among them, $\sigma_{1\varepsilon}$ and $\sigma_{2\varepsilon}$ draw special interest, as they capture the interdependency between the shipment size decision and the mode choice.

[55] The approach proposed by Holguin-Veras, not used in the following of his work, allows for the use of a distinct shipment size specification for each vehicle type. Nevertheless, the set of alternatives depends on the shipment size forecast by the first equation, to reflect the technical limitations of the vehicles.

Note that vehicles do not necessarily carry shipments one by one. A measure of the distance between a given shipment size and the average size of shipments carried by a given type of vehicle might have been more relevant here.

The model specified by Abdelwahab and Sargious has been estimated against the Commodity Transport Survey[56] of 1977. The results are interesting and clearly illustrate an interdependency. For example, an increase in the cost of truck transport implies a higher probability of choosing the rail transport mode. It also implies, if the truck mode is chosen, a higher shipment size, certainly to benefit from better rates. The shipment size conditionally to the choice of the rail mode also increases, seemingly accounting for the reasoning of a shipper who considers higher shipment sizes whatsoever given the higher truck transport costs.

This example illustrates both the findings and the limits of the approach. Taking into account the shipment size certainly improves the fit of the model, and allows for a more accurate calculation of elasticities, as is done in Abdelwahab (1998). But the economic causalities are not accounted for, and the model used is weakly linked to the existing theory concerning shipment size choice. For example, the shipment size suffers no technical constraints such as vehicle capacity. Second example: among the variables present in the dataset, the total volume shipped on a given origin-destination was used in the regression, and its influence was estimated positive on the shipment size. According to the authors, this would confirm the existence of economies of scale in road freight transportation. We think that this may be the sum of two effects: first, a larger flow between an origin and destination allows for larger shipments shipped at a larger frequency, which is generally good for the shipper and the receiver. Second, the greater amount of vehicles present on the origin-destination pair allows for a more efficient use of them, better average payloads and a greater availability, which could lead, all other things equal, to the possibility to send smaller shipments: in this case, the second effect identified by Abdelwahab and Sargious pertains to the presence of economies of scale in road freight transport.

Such an analysis would certainly be enriched by a microeconomic analysis.

2.6.3 Microeconomics of the choice of shipment size

A large number of works and methodologies consider explicitly the choice of shipment size, as it is one of the central variables of decision of the logistic function of firms. A subset of these works investigate this question using microeconomic tools and from the perspective of freight transport demand modelling. This section introduces some of them. Note that

[56]Stopped in 1988, renamed Commodity Flow Survey, re-launched in 1993 (Holguin-Veras, 2007) and presented page 124.

some of these models have already been met in Section 2.3; however, what is here discussed is not their validity for a given firm in a well identified environment, but for a large, potentially very heterogeneous population of firms.

The work of Baumol and Vinod (1970) is one of the first models which investigate the issue of shipment size from a freight modelling perspective. Using our notations, a fixed amount shipped per year Q is considered. The objective is to minimise the total logistic cost, denoted g, which consists of four cost components. First, the direct shipping cost c_tQ, where c_t is the transport cost per unit (e.g. ton). This cost does not depend on the shipment size. Second, the in-transit carrying cost atQ, where a is the in-transit cost per unit of time and freight[57], and t the transit time. Third, the ordering cost bQ/s, where b is the cost of ordering a shipment, depending on the transport mode, and s the shipment size. Fourth, the inventory cost. The authors consider only the destination inventory cost $a_ds/2$ (where a_d is the inventory cost per ton and year[58], and $s/2$ is the average inventory level if the shipments are uniformly spread over the year). Then, the total logistic cost is:

$$g = c_tQ + a_ttQ + \frac{bQ}{s} + \frac{a_ds}{2}.$$

The problem of the shipper is then to choose a shipment size s and mode $\{c_t; b; t\}$ that minimises g. These decisions can be considered sequential. Indeed, the optimal shipment size conditionally to the use of a given mode is $\sqrt{2bQ/a_d}$ (this is equivalent to the optimal shipment size in the EOQ model presented in Subsection 2.3.3). The shipper then chooses the mode providing the lowest total logistic cost.

The authors then assume the system presents stochasticity (either in the demand or in the travel time). In that case the a safety-stock is needed, to ensure shortages do not happen too frequently. If the demand is stochastic, for a shipment size s, the maximal delay for filling an order is $s/Q + t$, the average unsatisfied demand that may accumulate during this period therefore is $(s/Q + t)Q$. Given the variance of the demand at each period and for all α between 0 and 1, there is a constant k such

[57] This cost is the willingness of the shipper to pay for a decrease of the travel time of one ton of commodity of one year. It typically encompasses a capital opportunity cost and a depreciation cost.

[58] This cost is the willingness of the shipper to pay for a decrease of the waiting time of one ton of commodity at the destination inventory of one year. This cost typically encompasses a capital opportunity cost, a warehousing cost, and a depreciation cost, etc.

that the following safety-stock:

$$k\sqrt{s+tQ},$$

ensures that the probability of a stock-out is smaller than α. k increases in α. For a given α, the total logistic cost becomes:

$$C = c_t Q + atQ + \frac{bQ}{s} + \frac{a_w s}{2} + k\sqrt{s+tQ}.$$

With this new cost function, there is no close formula for the optimal shipment size. Furthermore, there is no microeconomic reasoning explaining the choice of an optimal α. See Chapter 7 for further discussion about this specific issue.

The analysis led by Hall (1985) is similar in its principle. The framework is identical, except that it takes into account the capacity of the various vehicles available. Denote S the maximal capacity of a given vehicle (with our notations). Therefore, the maximum shipment size is S. The optimal shipment size is:

$$s* = \min\left\{\sqrt{\frac{2bQ}{a_w}}; S\right\}.$$

Hall then examines the joint mode and shipment size decision for a shipper sending a flow of production Q (in pounds per week in the example) on a given origin-destination pair, with three alternatives: either using truckload (TL) contract carriers (who provide direct service, and whose rates are distance dependent), or less-than-truckload (LTL) common carriers (who specialise in grouping small shipments), or through the United Parcel Service (UPS) (specialised in the smallest shipments).

Figure (2.9) illustrates the costs of each mode for a given Q. The best mode is the cheapest one. Figure (2.10) illustrates the optimal shipment size for a flow Q, as well as the mode to use. Both these figures are taken from Hall (1985) It is interesting to note that, given the modes available, some shipment sizes are never optimal.

These works make a step towards accounting for the linkage between the logistic imperatives of firms, and freight transport demand. They have been extended to many distinct frameworks. Tyworth (1991) reviews some of these models, with special emphasis on how the stochastic nature of demand and transit time can be modelled. A more recent review, by Vernimmen and Witlox (2003), is available. From these reviews,

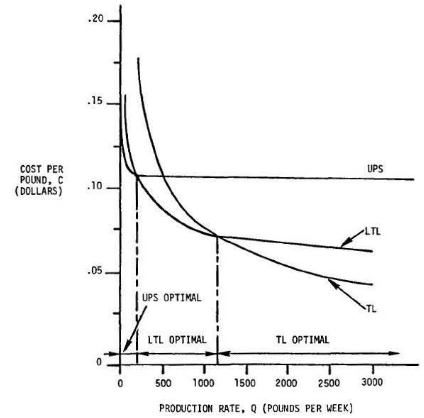

Figure 2.9: Movement cost incurred by each mode (taken from Hall, 1985).

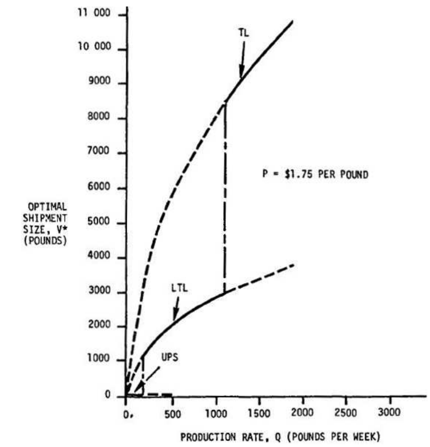

Figure 2.10: Optimal shipment size and mode (taken from Hall, 1985).

it appears the basic elements of joint shipment size and modal choice decisions are all presents in the work by Baumol and Vinod (1970). The subsequent refinements bring little to the economic analysis.

Furthermore, due to data requirements, it has not been possible to proceed to the econometric assessment of these models. Indeed, even the simple model of Baumol and Vinod requires a measure of the commodity flow between the shipper and the receiver. Except for the case of the ECHO survey, this variable is generally not available. Attempts to estimate these models without this variable prove difficult, as confirmed by the example of the work by McFadden et al. (1985). This particular issue is addressed in Chapter 5.

Finally, these models aim at explaining how shippers choose between distinct transport modes. As a consequence, they are by design unable to explain why a shipper would combine two transport modes for a given commodity flow. Modelling this kind of phenomenon requires a more detailed analysis of the logistic imperatives of firms. This is the object of Chapter 7.

2.6.4 Shipment size in freight transport modelling

We presented in Chapter 1 recent advances in freight transport demand methodologies. They pertain to many aspects of freight transport modelling, including considering new decision units, such as the vehicle movement in FRETURB (Routhier et al., 2002) or shipments (but in a rather coarse way) with the logistic chains in EUNET (Jin and Williams, 2005).

But none of these models address explicitly the shipment size decision in their architectures.

The approach of de Jong and Ben-Akiva (2007) aims at addressing this gap[59]. The approach they propose is referred to as "aggregate-disaggregate-aggregate". Production-Consumption matrices, obtained through conventional ways, are disaggregated into a large set of firm-to-firm flows. The way the goods constituting these flows are moved results from a joint shipment-size / transport chain decision, using a model similar to those presented in the previous subsection. This procedure is a major originality in freight transportation modelling.

The second originality is that consolidation is taken into account. Taking into account both the shipment size and transshipments requires indeed the representation of economies of scale. Nevertheless, theoretical questions remain concerning the existence, uniqueness, and convergence towards an equilibrium of such a model[60].

2.7 Conclusion

We began our appraisal of logistics from a freight transportation modelling perspective by defining unambiguously a number of terms, which will be used later in the analysis.

For firms, logistic issues are first and foremost operational. As such, we first studied the tools developed to address them, which are therefore mainly engineering tools. Their accuracy depends on the trade-off made by their users between fine-tuning and robustness. The short presentation of these tools gives an idea of the nature of a logistic issue, and how it may be addressed.

We also showed that when addressing macroscopic issues such as freight transportation at a large scale, interactions between agents play an important role. We presented some elements of analysis with respect to the necessary coordination of firms in a supply chain, as a contradictory force with the competitive drivers resulting from markets structures.

These macroscopic features also exist at an even greater scale, when one does not consider a single market but a regional or national spatial scale. We presented some approaches which focus on the effect of space

[59]This model is still being built in 2007; its principles were presented two years before (de Jong et al., 2005), and a presentation was done in 2007 about its calibration on aggregate data (de Jong et al., 2007). Its estimation on disaggregate data is still to be done in 2007.

[60]These questions are likely to be difficult to address, due to the explicit presence of increasing returns in the model.

and time on the economy. These approaches are symmetric to those addressing logistic issues of the firm.

We finally focused on one of the central questions of our work, which is the identification of the drivers of the shipment size decision, and how they may be accounted for in a large-scale freight model.

We thus tried to fulfill the two objectives of this analysis: identifying the drivers of the logistic decisions of firms, and how they explain the properties of freight transport demand; and presenting a set of approaches, methodologies and models that may provide inspiration in building a more realistic freight transportation demand model.

Modelling microeconomically the choice of shipment size appears to be a fruitful but difficult task. Joint mode choice and shipment size models are able to account at least partially for the linkage between logistic requirements of firms and freight transport demand, in particular mode choice. As such, they are a very interesting direction to improve freight transport models. But their empirical relevancy is not proved. Indeed, assessing these models econometrically requires specific data, not only on shipments, but also on the shipper-receiver relationship, which is seldom observed.

Part II

Systems analysis and metrology

Chapter 3

A systemic representation of the freight transportation system

3.1 Introduction

Substantial progress has been made over recent years in modelling freight transport demand, in several directions. All these lines of progress share one same objective: to improve the realism of models in view of their ability to generate realistic reproductions of known situations and in view of their use as decision support tools, which would notably be possible if they can be used for different extrapolation operations.

Let's consider the category of spatialised models of freight transport demand[1], i.e. models that represent explicitly the spatial dimension of variables of interest (for example: traffic and speed for each infrastructure network arc, vehicle flow for each origin destination pair, etc.) Until recently, these models have been constructed by adapting models of passenger transport demand, with minor methodological improvements aimed at compensating for the inadequacy of passenger transport models to adapt to the context of freight transport. As a result, spatialised freight transport model have long been structured along the classic 'four-phase' representation which forms a common basis of passenger transport demand models.

However passenger transport and freight transport are different in

[1]Here, we will not go into non-spatialised models that relate to aggregate indicators, such as the total number of tkm in a given country per year. These models are built using specific econometric methods and are not of a different type from spatialised models.

many respects and these differences have an impact on the realism of models. Recent advances in freight transport demand modelling progress often consist of better representations of mechanisms totally specific to freight transport, for example: improvements in the representation of freight transport supply, the logistics dimension of choices made by shippers, and even the explicit consideration of shipments in models (Tavasszy, 2006; Combes and Leurent, 2007). Each of these points forms a fundamental difference with passenger transport, thereby limiting the scope of the classic 'four-phase' representation in the frame of freight transport.

Adapting this context to freight transport would be useful in several capacities: it would enable us to draw comparisons between recent works and assess their coherence; it would also enable identification of points that still require further research and as such would potentially constitute a common work basis; it could then be used as a structure for a freight transport model, coherently integrating the latest lines of progress.

Section 3.2 presents more precisely the objectives of this study. Section 3.3 recalls the four stage representation and the reasons for which it is now widely used in passenger transport modelling. The reasons why it is not adequate for freight transport modelling are then detailed. A systemic representation of freight transport is then presented in Section 3.4, as the result of a systemic analysis of the freight transport system. Section 3.5 concludes this study.

3.2 Objective and method

As explained in the previous section, our objective is to propose a framework for modelling freight transport demand, firstly to determine the relative place of recent works, drawing a comparison between them, and secondly to act as a base for the construction of a realistic spatialised model of freight transport demand.

This task shall be conducted in two phases. First, we shall recap the classic four-phase representation of passenger transport demand modelling, the objectives to which it responds and the specificities of freight transport for which it is not suitable.

We will then proceed with a systemic analysis of freight transport, identifying all agents whose decisions affect freight transport operations directly. For each of these agents, we shall identify the decisions they make, the options and resources available to them, the way in which they choose between these alternatives and lastly, the relations between these agents. Throughout this analysis, we shall determine the elements that need to be represented explicitly in a freight transport demand model, those that can be represented in a simplified manner and those that can be overlooked without affecting results. We will use this analysis to construct a representation of the freight transport system.

3.3 The four stages representation

Passenger transport demand modelling resides in a consensual "four-phase" representation (Quinet, 1998). This representation is microeconomically and statistically consistent with the behaviour of the agents of the passenger transport system, and its structure in layers is a convenient base to build a model upon. Given the similarities between the passenger transport system and the freight transport system, it was, as such, a good starting point for modelling freight transport. We shall therefore recap its principles (Subsection 3.3.1) and then explain the limitations of its adaptation for freight transport (Subsection 3.3.2).

3.3.1 The passenger transport system

Initially, the intention of passenger transport demand modelling concerned the sizing of road infrastructures. These models are generally designed to forecast the use of different transport possibilities by passengers. They were therefore constructed using a pragmatic approach: the objective was to forecast traffic, the pertinence of different aspects of passenger transport was assessed in the light of their contribution to traffic formation, and the capacity of modelling methods to factor them in.

These passenger transport models are supply-demand models. Transport supply is described quite simply[2] and the behaviour of transport suppliers (infrastructure managers, public transport operators) is not represented explicitly; we only see the result, considered as exogenous in the models.

The demand - i.e. the need for passengers to travel and the way in which they choose from the different alternatives available to them - is examined in further depth. Initially, passenger transport modelling was solely focused on traffic on the road network. At a very early stage, three phases were identified in the formation of demand, respectively called generation, distribution and assignment, corresponding more or

[2]This simplicity is deceptive: it disregards the multiple tariff regulations present on all transport networks and the specific supplier-demander relations are concealed.

less to the different decisions made at different time scales by passengers. The modal choice was later introduced between the distribution and the assignment stages, when the competition between the different modes of transport (particularly personal cars and public transport) incited more interest.

These decisions show a form of hierarchy since a decision made at a given level determines the options available at the level below. The phase which is the farthest upstream - generation - corresponds more or less to the location decisions of households, activities and companies (and therefore employment). The distribution phase corresponds to the choice of activity by an agent: the choice of a job, the reason for commuting, the choice of shops, schools for children etc. The journeys they have to make stem from this choice of activity. These passengers then need to select their mode of transport and lastly, the itinerary for their journey. This segmentation of decisions impacting traffic formation is well-suited to modelling. These decisions can be represented in the form of superimposed layers, thereby clarifying both hierarchical relations between these decision levels and their common spatial dimension (Figure 3.1).

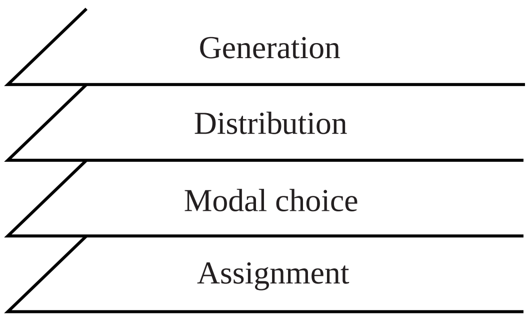

Figure 3.1: The four stages representation

This representation of the passenger transport system does of course have its limits in terms of suitably modelling certain characteristics of passenger transport. Examples of missing elements include: an explicit representation of the use of a private vehicle by several people, chained trips, tariff decisions by suppliers and even the choice of departure time when considering a dynamic context. However it does represent a good foundation for constructing a realistic passenger transport model and a good starting point for attempting to overcome these problems.

The situation is similar for freight transport; the four-phase representation is a good starting point, but it is to a certain extent restrictive, as we shall explain in further detail.

3.3.2 Specificities of the freight transport system

Initially, the freight transport system and the passenger transport system have similar features. For example, if the passenger flow is replaced by the merchandise flow expressed in units of weight, and the transport services in terms of transport time and cost, the itinerary choices can be modelled correctly. We can also note that characteristics are on average more similar for commodities using the same mode of transport, than for commodities using different modes of transport. And finally, the generation stage is more or less functional if we apply economical descriptive variables (GDP, population, employment, possibly categorised into sectors) to forecast the intensity of freight emissions and receptions. The distribution phase is more complex, but the methods used for passenger transport are applicable. In short, the four-phase representation presented above is applicable, to a certain extent, to the freight transport system.

Nevertheless, if we wish to improve the realism of a freight transport model, we need to account for the specific operating characteristics of the freight transport system, which cannot be achieved explicitly in the context of the four-phase representation. We shall highlight three of these limitations.

The decision unit In order to assess the level of aggregation in a freight transport model, it is necessary to identify the decision unit of the freight transport system, *i.e.* the smallest group of freight considered as indivisible in decisions pertaining to its transport. The corresponding notion in passenger transport is straightforward: it is the passenger itself[3]. With freight transport, the decision unit is less clear and many models do not make it explicit; they opt implicitly for an 'atomic' decision unit which is consistent with assignment models with congestion; the 'atoms' are aggregated directly to form origin-destination flows which alone are explicit. Whereas the agents involved in freight transport do not decide the way in which each atom of commodity is to be transported, nor do they make this decision at the scale of an origin-destination flow as a whole. The decision unit for freight transport is intermediate. It must therefore be identified.

[3] Nevertheless, a certain measure of ambiguity resides between the passenger and the vehicle, bypassed by the use of an occupation rate, without necessarily being explicit. Certain commuting cases can also be considered, where the passenger does not choose to make the journey and has limited options in terms of modes of transport.

Decision makers In the context of passenger transport, a large proportion of the decisions involved in traffic formation are made by the passengers themselves. Most of their decisions are relatively well identified, and correctly represented by microeconomic or statistic models (trade off between working time/leisure time, choice of job, etc.). In the case of freight transport, the cargo doesn't make any decisions. The agents constituting the demand for freight transport, *i.e.* the shippers, are very heterogeneous. They may be individuals, different-sized companies, selling products to end-consumers, or to agents supplying other companies. Their decisions are based on many parameters, including transport cost and duration, but also customer satisfaction, punctuality, delivery tracking, stock shortage probability incidence, and so on.

Origins and destinations The destination of a passenger's trip is generally the place where the passenger wants to or needs to be present, since it will be performing an activity at that location. The intermediary stops made by the passenger, to change their mode of transport for example, are short and, in transport models, are not considered as the destination of a trip and the origin of a second trip. Defining the destination of the trip of freight is a more complex task: is the destination the place where the merchandise is consumed, used or processed? Or simply the destination of a transport operation? If we opt for the first definition, we consider that stops at warehouses, potentially very long, are only intermediary stops in the course of a trip. Whereas commodities can be stored for varying reasons, other than synchronisation of different transport operations. The second definition poses a problem in terms of symmetry: short stops at cross-docking platforms would be interpreted as the end of a journey and the start of a second journey, which would not be pertinent.

In the next section, we address these three limitations, concerning respectively the decision unit in freight transport, the identification of decision makers and their criteria, and the different stages in freight transportation.

3.4 The freight transport system: a systemic representation for modelling

A model of freight transport demand has two purposes: it enables the forecast of activity indicators for the freight transport sector (possibly spatialised, as in, for example, the case of heavy goods vehicle traffic per road link), and, ideally, it can be used as a reliable decision-making tool. These two objectives will be more likely to be achieved if the model accounts realistically for the behaviour of the agents involved in freight transport and the way in which they react to changes in their environment. This is why systemic analysis of freight transport is useful, particularly if we wish to incorporate a maximum of microeconomic behaviours into the model.

Fundamentally, freight transport results from the spacial inadequacy between the location of productive resources and the location of the end consumers. Furthermore, the technologies (in the economical sense of the term) enabling the transformation of these resources into products required by end consumers, are often endowed with economies of scale of varying magnitudes, favouring the concentration of production installations. Consumer preferences are such that they prefer to make an effort to obtain goods that are not immediately available, rather than contenting themselves with what is immediately available. This effort is substantial, considering that the transport and logistics sector employed approximately 1.5 million people in France in 2004 (Mariotte, 2007).

Many agents are involved in commodity movements emanating from the motives outlined above. They form the freight transport system. Section 3.4.1 will now describe how this system functions, phase by phase. Then, an integrated representation of the freight transport system is proposed in Section 3.4.2.

3.4.1 Freight transport systemic analysis

The first of the four stage representation limitations mentioned in Section 3.3.2 concerns the decision unit for the freight transport system. It will be addressed in Section 3.4.1.1. Section 3.4.1.2 will then examine the way in which the freight transport offer is constructed, then the determining factors behind the demand are investigated in Section 3.4.1.3. We refer to the freight transport demand as "shippers" and to the freight transport supply as "carriers"[4].

3.4.1.1 The decision unit for the freight transport system

The first fundamental step in the systemic analysis of freight transport involves determining the freight transport decision unit, i.e. the smallest

[4]Therefore, if a company performs itself a transport operation it wants done, it is both a shipper and a carrier.

set of merchandise considered as invisible in decisions pertaining to its transport. This entails defining the freight transport central object, directly concerned by all the decisions made by the different agents involved in the freight transport system.

This decision unit is the shipment (Definition 2.6). To see it, we must first show that this level of detail is necessary to account for the full dimension of decisions concerning freight transport modes. Indeed, characteristics such as the choice of shipment size and its conditioning play a role that is (at least) just as important as the choice of the mode of transport and itinerary, to which they are all, nevertheless, closely linked. The shipment characteristics result from logistic imperatives of shippers and strongly influence the technical options available to the carriers for transporting the consignment, as well as their costs. For example, the shipper might assign a high level of importance to the transport's cost and therefore proceed with large-sized shipments, to use the capacities of large-sized vehicles to best advantage and to therefore offer better tariffs per unit; it might, however, opt for a shorter travel time and therefore dispatch smaller shipments. Symmetrically, the shipment's characteristics are a major determining factor in terms of the way in which it can be transported. For example, if the shipment's dimensions are significantly less than the capacity of the vehicle that the carrier proposes to use, the carrier will have to dedicate a transport operation to this shipment, or else other shipments will need to be found to make more efficient use of the vehicle and driver. If the shipper requires the shipment to be delivered very quickly, such pooling can be harder to execute, particularly in terms of synchronisation.

The second step of the discussion is to show that the shipment level is detailed enough. Indeed, the commodities which form a shipment are considered as a whole by the shippers and the carriers in their decisions pertaining to freight transport operations. For example, in cases where a consignment is divided into several vehicles (due to the size exceeding the vehicle capacity, for example, or to perform a handling operation[5]) , the quantity of freight is still considered as a whole which needs to be transported as a whole, under pre-determined conditions.

In a representation of the freight transport system, it is important that this decision unit be explicitly present. If we respect the spatial

[5] A shipment of twenty pallets of freight, transported by truck, can therefore be transported in two phases, stopping at a platform to transfer the load from one trailer to another. During this handling operation, pallets are handled one by one, for relatively precise technical reasons. Of course, neither shipper nor carrier considers each of the twenty pallets as the subject of a different transport operation.

dimension, we can opt for a representation similar to that presented in Figure 3.2.

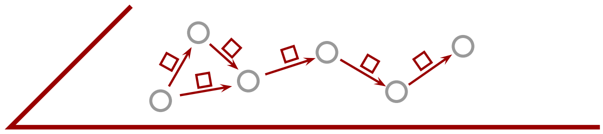

Figure 3.2: The shipment layer

In this diagram, the consignments are represented by red arrows with small squares running above them. These small squares play more than just an aesthetical role, they allow us to highlight the difference between this representation, where all the pertinent characteristics of the shipment (including size and packaging) are explicitly present, and the classic representations in freight transport where the arrows represent freight flow, as a unit of weight, volume or value, per unit of time.

It should be noted that we have little data with regards the shipment, despite the fact that the shipment is the fundamental decision unit for freight transport. To our knowledge, the ECHO database is the only recent database available concerning shipments in France (Guilbault et al., 2006).

3.4.1.2 The freight transport supply

The freight transport supply consists of the supply by carriers, a group of a large number of heterogeneous agents, of transport options with very different characteristics[6]. The carriers can perform the shipping operations themselves (case of own-account transport), or the company's main activity may be transport, etc.

The decisions that have to be made by a carrier can be divided into several different registers. For simplicity, we can divide these into two categories: strategic and operational. The operational decisions concern the way in which the carrier uses the resources available to it to perform the operations within its set objectives. The strategic decisions include the choice of services that the carrier is offering, their tariffs and their characteristics, as well as the financial, material and human resources that the carrier will deploy to supply these services and its relations with

[6]By means of example, we can illustrate the variety inherent in these characteristics: freight transport agents use a conventional segmentation of transport operations according to the weight of shipments, in several categories such as express delivery, parcels, pallets, full truck, etc.

agents offering more or less similar services (partnerships, subcontracting, competition, etc.)

Decisions of a strategic nature will not be subject to an explicit representation. In general, the freight transport market is relatively competitive, at least for certain types of transport operations. Subsequently, prices seen by shippers correspond more or less to the carrier costs, meaning that strategic decisions can be overlooked, in an initial approach. Nevertheless, a certain measure of caution ought to be observed. Let's look at the example of road transport. Clearly, the size and partnerships of road carriers (or those using the road mode) result from their strategic interactions. Whereas it is precisely because these road carriers manage to reach a certain critical size and because they form partnerships that they are able to draw suitable benefits from the economies of scale linked to the fixed capacities of vehicles[7]. Inversely, they are unable to address road infrastructure congestion phenomena by coordinating themselves to use the network of infrastructures effectively, precisely because of the relative dispersal of this market[8]. The strategic dimension therefore holds a significantly important role. Nevertheless, initially we do not intend to include it in our representation of the freight transport system; the cost in terms of complexity would most probably outrun the gain in terms of precision. The strategic dimension shall therefore be accounted for using *ad hoc* hypotheses.

Decisions of an operational order, however, are interesting primarily in terms of freight transport modelling, and must be represented as explicitly as possible. The numerous economies of scale present in freight transport, and the way in which they are exploited, contribute substantially to the characteristics of transport operations which form the freight transport offer. We propose to draw a distinction between the different decisions inherent in the formation of the transport offer, dividing them into three categories, according to whether they concern the location of fixed resources, elementary transport operations, or transport of shipments.

The result of this systemic analysis is represented by Figure 3.3, in which the different decision levels are shown by different layers. In the column on the left, we indicate the decisions to which the layers appearing

[7] A carrier cannot adapt the capacity of its vehicles to the cargo, nor the number of drivers on board. Subsequently, for a given vehicle type, the marginal cost corresponding to the transport of an additional transport unit is low unless the capacity is saturated, while the fixed transport cost is significant: this is the cause for increasing returns to scale.

[8] Which, in theory, is accessible to a monopolistic rail transport operator, for example.

in the middle column correspond. In the column on the right, we indicate observables corresponding to decision layers, *i.e.* borrowing the notion from physics - phenomena that can be measured and thereby enable deductions to be made with regards the way in which the decision layers operate. These observables condition data collection possibilities.

Fixed assets The fixed resources that a carrier may require to perform transport operations include platforms, warehouses, sorting stations and even combined transport sites. The location of these resources strongly conditions the options available to the carrier, since it cannot reorganise its operations, especially when it is the owner. A road carrier, for example, will need to determine the number of break-bulk platforms that it wishes to have, their locations, their sizes, their configurations; it can choose to rent them or to have them built; its decision will depend on the costs of these options as well as the opportunities that they offer him in terms of offering transport operations that meet certain shipper requirements. Interactions of carriers at this level are complex: cooperation on inter-modal nodes, competition due to property rental rates for operating in advantageous zones, etc. The observable corresponding to this decision level is industrial urban planning.

Elementary transport operations The second decision level involves determining elementary transport operations, where we consider that one elementary transport operation consists of the vehicle's journey between two places where at least one loading or unloading operation is performed. The cargo can only be modified, therefore, between two elementary transport operations. A change of driver does not constitute an interruption in an elementary transport operation. These operations are characterised by time tables, itineraries, costs associated to the use of different resources such as vehicles, drivers, fuel, etc. Many options are available to carriers, and the choices that they make are guided by the objective of minimising costs to produce the transport operations requested by the shippers under set terms. The observable corresponding to this decision level is traffic, which can be measured in different ways.

Transport of shipments The third decision level concerns the routing of shipments in elementary transport operations. The shipments can be transported in one single operation (in the context of an elementary transport operation) or in several operations, with stops at break-bulk platforms, or even changes in the mode of transport, etc. Separating these two levels - the transport operations and the shipment itinerary within transport operations - allows us to highlight several important points. Firstly, a shipment's origin and destination does not necessarily coincide with that of the vehicle in which it is transported, which has a distinct impact in modelling. Secondly, understanding the way in which carriers pool their vehicles and their drivers to transport shipments efficiently at the lowest possible cost, is fundamental. The observable corresponding to this decision level is the shipment.

Logistics of carriers The link between elementary transport operations and the routes taken by the shipments is complex, emanating from a set of rules, *i.e.* a protocol. The purpose of this protocol is to manage the flow of goods, it is therefore a logistics protocol. We call it the "logistics of carriers" protocol, because it concerns, exclusively, the execution of consignment transport operations under given conditions and must be distinguished from the logistics-related choices of shippers, which will be presented in the next section.

At the bottom we have added (highlighted in grey) the layer of infrastructure networks as a point of reference, which obviously plays a structuring role in the formation of the freight transport offer. We have also included the layer of the logistics of carriers protocol (dotted lines), forming the link between the elementary transport operations and the modes of transport used for the shipment. The shipment layer is represented in the same way as in Figure 3.2.

3.4.1.3 The freight transport demand

Companies move the goods they have at their disposal, they buy, and they sell, according to logistical imperatives that require clear explanation to provide an understanding of the determining factors of freight transport demand.

Companies generally seek to maximise the profits they generate. This task involves factoring in the needs of the company's customers as well as the choices that will be made by competitor companies. The products proposed by the company are central to these considerations. If we follow the consumer theory of Lancaster (1966), the products have no intrinsic value for consumers, only a certain number of these product's characteristics will determine their value. Amongst these characteristics, three are important from a logistics perspective: the price, the effort that the consumer has to provide to obtain the product (time-frame, shipping costs or eventual delivery costs) and the risk that the product may not be available under the conditions expected by the customer (due, for ex-

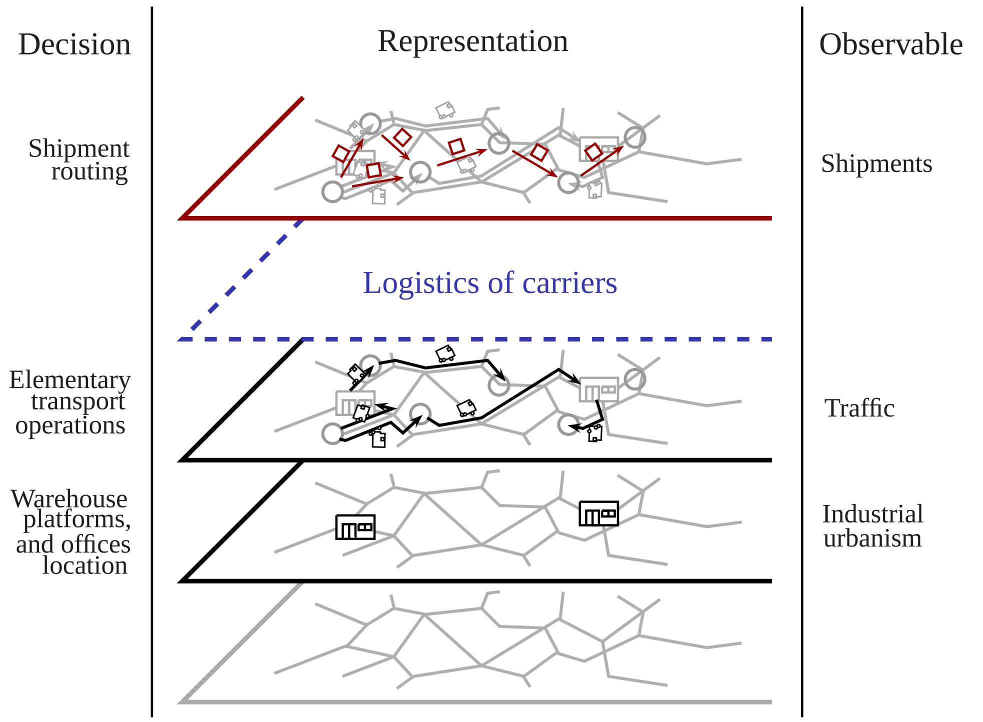

Figure 3.3: Systemic representation of the freight transport supply

ample, to delivery delay, or stock shortage). These three characteristics, which can be referred to as price, generalised distance and reliability, are interconnected for producers. A short generalised distance, as well as a low risk of failure, are only possible at a higher selling price, etc. The trade off that the company will make between the three results from its technological options (operational dimension) and its position vis-à-vis its customers and its competitors (strategic dimension). These two dimensions represent the main components of logistics-related issues, *i.e.* the flow management function[9] of the firm[10].

The options and preferences of companies, as well as the interactions between each other, are too complex at this decision level to allow us

[9]See, for example, Carbone (2004) and Tatineni and Demetsky (2005*b*), who consider that logistics is the management of good flows and their associated information and financial flows.

[10]Logistics-related concerns are of a broader scale. It is easier for a company with good control of its information flows to provide a good level of availability of its products. But the control of information flows can itself have a strategic dimension, since it can endow a company with a certain level of power with relation to other companies in certain contractual negotiations. We will not be examining this dimension here (for further discussion on this subject, see e.g. Dornier and Fender, 2007)

to hope to represent them explicitly. However, decisions made by companies can be divided into hierarchical categories, sorted by their time scales. With this type of representation, decisions are made accounting for their consequences on all levels beneath them, while it would be reasonable to assume that they are made accounting for fixed decisions at the levels above them. We propose distinguishing four decision levels: the location of production installations, the supply decisions, the logistical organisation of supply chains and the demand for transport of shipments. This is illustrated by Figure 3.4. It shows the decisions examined, a graphic illustrative representation and several observable manifestations of these decisions.

The location of production installations The decision level that is the highest upstream in the production organisation resides in the location of production installations. This is represented by small factories drawn in black, on the higher layer. This decision has the greatest impact, for two reasons. Firstly, it conditions the organisation of production and flows, from the suppliers through to the customers. Secondly, delocalising these installations is difficult. For the company, this involves finding a location with a good compromise between installation costs, local production factor costs, costs of transporting the resources, costs of transporting the products to customers. The interactions between companies at this level are complex[11], we will not list them here. A freight transport demand model will, in any case, assume company locations are given. The end demand has a specific role since it is the end of all the production chains; it is shown in grey, as exogenous, in the diagram. The observable corresponding to the decision level for company locations is industrial urban planning.

Production-Consumption flows The following level concerns supply decisions. Here we designate the flow of goods between production installations and places of consumption (which can themselves be the production place of other goods) of goods produced at the place of origin and consumed at the place of destination. These flows are known as production-consumption flows in certain models, see for example Jin and Williams (2005) and ME&P (2002). The organisation of these flows cor-

[11] If we examine, for example, the case where several companies locate to the same area, we can quote examples of the following effects: increase in the job market, increase in the number of customers, increase in property rental prices, etc. This only gives a modest indication of the complexity of the spatial component of choices of companies and their implications (for further discussion, see e.g. Fujita et al., 1999).

responds for companies to the choice of suppliers and customers. Inter-company relations are complex at this decision level: involving, notably, the choice of supplier, a decision-making process in which certain agents may have a high negotiation power, due to their unique access to the market, or because they have a good control of certain information, etc. This decision can be considered as being subordinated to the location decision. We represent it by wide black arrows. The observable corresponding to this decision level consists of consumptions and productions of production installations and places of consumption. It would be measurable if, for example, accounts were available per establishment for companies, or if specific consumption statistics were available per household.

Logistic network design The third decision level involves determining the way in which these supplies are realised by goods flows, *i.e.* the way in which supply chains are organised. Companies have many options: the flows can be just-in-time, or with buffer stocks; the stocks can be pooled, or not; the finish of products can be postponed, etc. The preferences of companies with regards these alternatives depend on the compromise made by the companies between characteristics concerning price, distance and reliability, as presented above. At this level, the possible interactions between companies can, for example, involve the use of a common logistics platform, notably through logistics service providers. The result of this decision is represented on the diagram by small arrows which represent flows and by the warehouses that the companies are liable to use. The observable corresponding to this decision level consists of merchandise flows, measurables, for example, by classic merchandise transport surveys (Sitram in France), or by the calculation of goods imported and exported from different geographical zones.

Logistics of shippers As in the case of the formation of the freight transport supply, we represent explicitly the link between supply decisions and the way in which these supply decisions are concretely realised. This is a complex stage, emanating from a set of rules and also involving a protocol. It is similar in type to the logistics of the carriers, since its aim is to manage flows. But it differs in terms of its imperatives. We therefore call this protocol “logistics of shippers”, both to distinguish it from the logistics protocol of carriers and to clearly demarcate the relationship of subordination: the logistics of carriers protocol aims to fulfil, in the most efficient manner, all shipper requirements, emanating from the logistics of shippers protocol. This layer is represented in the diagram with a dotted line.

Shipment transport demand The final, and most operational, decision level involves the organisation of freight transport. Once a logistics organisation has been selected, flows are materialised by shipments, which will be routed by the shipment transport services available on the freight transport market. The options available to companies are those proposed by the carriers. Here, the aim is to create the goods flow with conditions, costs and reliability corresponding to the requirements of the higher decision levels, which will be a deciding factor in the selection of companies and transport options. The interactions are very much present, since the transport system is generally subject to economies of scale, scope and congestion phenomena. Some carriers manage to exploit these interactions to their advantage, but not all carriers achieve this, as explained above.

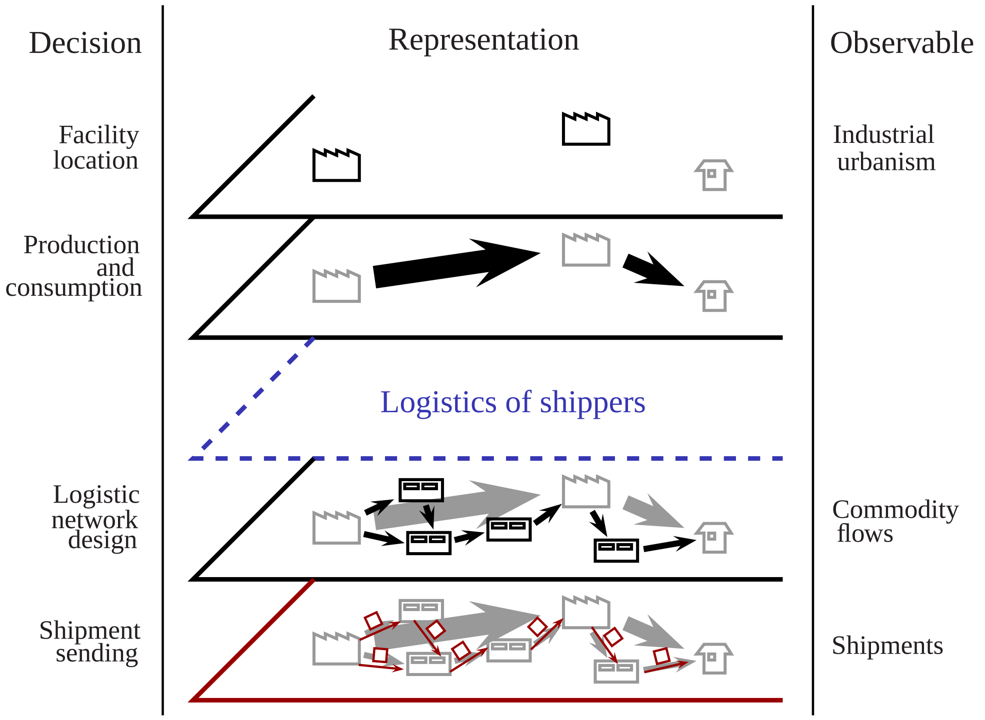

Figure 3.4: Systemic representation of the freight transport demand

We proceeded with a systemic analysis of the freight transport system, with the perspective of modelling freight transport demand. We first identified the decision unit for the freight transport system, then identified, for the formation of the offer as well as for the freight transport demand, the decisions involved, the options available to agents making these decisions, their selection criteria and, to a lesser extent, the eventual interactions between these agents, associated with the decisions examined. We are able to propose a representation of the freight transport system for modelling.

3.4.2 Representation of the freight transport system

The representation that we propose takes the form of a diagram organised into superimposed layers (Figure 3.5) showing (in simplified form) the three diagrams presented on the page. The layers are presented in four different colours: black for layers representing decisions, red for the central interest layer, grey for exogenous layers and blue for layers that do not correspond to decisions but to decision-making protocols.

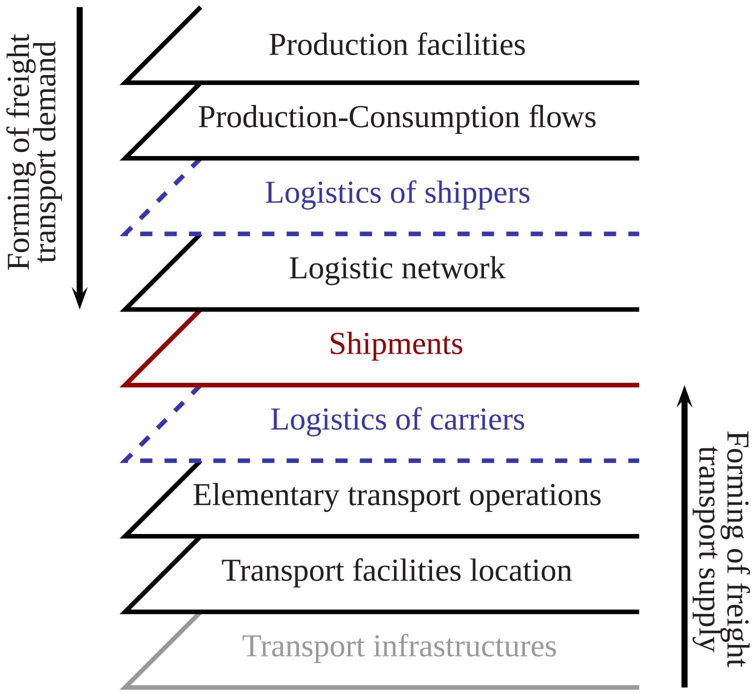

Figure 3.5: Systemic representation of the freight transport system

The consignments layer plays a key pivotal role in this representation. It separates, on one hand, the formation of transport requirements, resulting from the spatial nature of the economy and on the other hand, the way in which these requirements take form, accounting for the technologies available. Since the different decision layers are interlinked bidirectionally, they also account for the symmetrical approach used by shippers to adapt their choices to options offered by carriers, whilst the latter organise themselves to forecast and respond to the best to shippers' needs. The proposed representation is a good illustration of the extent to which this is a complex market game.

3.5 Conclusion

We have sought to identify, by means of a systemic analysis of freight transport, the agents that are directly involved in freight transport, their options, their decision-making modes and their interactions. We have identified a set of decisions that play a fundamental role in the formation of freight transport and are suitable for explicit representation in a spatialised supply-demand model. This representation therefore forms a potential structure for building a realistic freight transport model.

A particular point of interest in this representation lies in the fact that it allows us to distinguish clearly phenomena that do not appear spontaneously in databases collected in the traditional manner. We thus distinguish, upon completion of this analysis, the supply flows (or production-consumption flow), the freight flows, shipments and traffic. In static models, only flows and traffic are systematically informed and explicitly taken into account at present. This involuntary pooling of shipments and this lack of distinction between logistics-related imperatives for shippers on one hand, linked to their chosen method for delivering their products to their customers, and the logistics-related imperatives for carriers on the other hand, linked to the way in which their assigned transport operations are performed using the resources available to them to the best advantage, most probably introduces a degree of bias in these models, or means that these specific problems need to be handled in a rudimentary manner.

This representation does not highlight all the complex phenomena implicated in the operation of the freight transport system. Most particularly, relations between companies are not accounted for explicitly; they are either overlooked or handled, as we explained earlier, by ad hoc hypotheses. Whereas the way in which the different types and different sizes of carriers form their partnerships, the specific role of logistics service providers, the types of agreements set up between shippers and carriers, are all points that do indeed play a significant role in freight transport operations.

Chapter 4

Measuring the motor carriers activity: improvement of a road-side survey protocol

4.1 Introduction

Theoretical models, simulation models, as well as all kind of quantitative and qualitative analyses are based on databases, which describe the activity of the freight transport system in various ways. Among these databases, one type is particularly interesting to study the activity, organisation and costs of road freight transport: they are obtained using roadside surveys.

The purpose of a roadside survey is, among others, to observe the origin and destination of trips made by vehicles passing through a given road. They consist in interrupting a sample of the vehicles passing over this road during their trips, in order to to ask them a series of questions. These questions typically concern their trips' origins and destinations, length, motive, as well as the number of passengers in the case of passenger transport, and, in the case of freight transport, the type and amount of commodity transported.

This kind of survey is used extensively in France to gather data on the origins and destinations of passengers and freight vehicle trips. One of their uses is to yield information on local transport practises. Another one is to be combined with traffic data to build origin destination matrices at the regional and national levels. They usually last 2 to 6 days, and 100 to 150 vehicles can be surveyed each day. Their locations are usually chosen so that all the roads getting in and out of a relatively big city are surveyed, which is necessary to build OD matrices[1].

There are by and large a hundred roadside surveys per year on the French territory. On Figure 4.1 are indicated the locations of the 2100 roadside surveys which have been conducted between 1990 and 2008, both for passenger and freight traffic. Each color corresponds to one of the CETE[2], the regional centers in charge of these surveys.

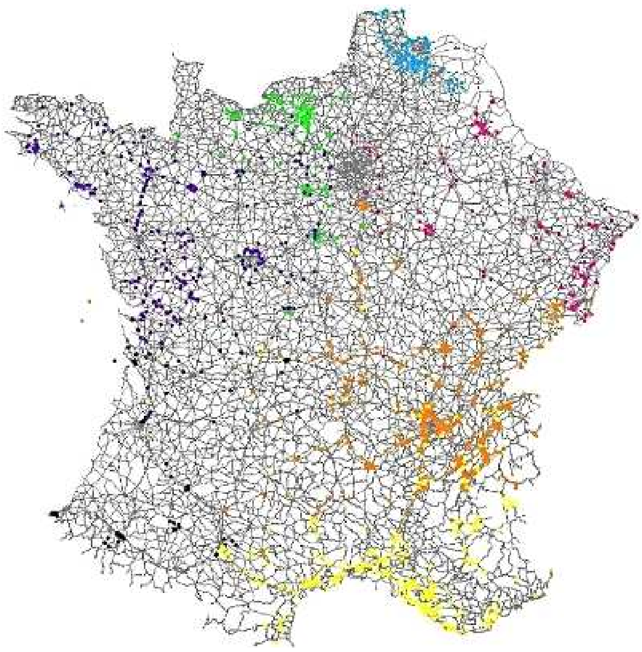

Figure 4.1: Recent roadside surveys in France

In the case of freight transport, these surveys are mainly purported to build OD matrices per commodity type, transport mode and conditioning. However, they can easily be modified in order to yield useful information about other aspects of freight transport, without increasing much their cost. The objective of this chapter is to investigate how this can be done, and to identify which type of information can be obtained.

[1]The measure of an origin-destination flow, *i.e.* the number of passengers or the amount of commodity leaving the origin to reach the destination per period of time, requires that the subset of surveyed links of a given infrastructure network contains at least a graph cut. In that case, the road-side surveys data can be combined with traffic data to provide an unbiased estimate of the OD flow. If no such graph cut exists, other methods are necessary to complete the OD matrix, involving models. See Leurent and Meunier (2009) for details.

[2]*Centre d'Etude Technique de l'Equipement*, technical studies center of the French Transport Ministry.

The outline of this study is the following one: the classic French roadside survey protocol and questionnaire are first presented in Section 4.2. Afdterwards, two surveys are presented. The first one is a pretest survey, in the frame of which an extensive questionnaire has been administered. Section 4.3 presents the construction of this questionnaire and some results of the survey. However, due to specific operational constraints, the classic roadside protocol could not be respected, introducing a possible bias in the results. A second survey has thus been conducted, in accordance to the classic protocol, to determine whether the results of the first one are reliable or not, with a reduced version of the questionnaire built in the previous section. The results are presented in Section 4.4. Although a number of numerical results are provided in this discussion, it is mainly focused on the possibility to gather useful data.

4.2 The classic survey protocol for freight transport

The roadside survey consists concretely in diverting momentarily a number of vehicle from the road so that interviewers can administer them a questionnaire. The operation's presence is indicated by traffic cones and mobile signs. The trucks are diverted by policemen, who alone are authorised to operate inside the traffic (the availability and willingness of the police to participate to the survey is thus absolutely necessary, and sometimes not enthusiastic). They are directed to an available area on the side of the road to be interviewed (typically a rest area for major roads), after which they resume their trips. There can be several interviewers, so that several drivers can be interviewed simultaneously.

The contents of the questionnaire can vary, but the questions asked to truck drivers are essentially the following ones.

- *Number of axles.* Observed by the interviewer, the answer is generally comprised between 2 and 5. The trucks with 4 axles or more are generally semi-trailer trucks, they are constituted of a tractor unit towing a semi-trailer. carrying the freight. The vehicle capacity increases with the number of axles.

- *Type of vehicle.* This variable is also observed by the interviewer. In case of a semi-trailer, the interviewer must fill in the semi-trailer type. The types distinguished are usually: container, box (rigid sider), tanker (for liquid and gas), reefer (equipped with a heating/cooling unit), dry bulk (for dry powder materials), flatbed (a load floor with removable side rails), tautliner (curtain sider), other.

- *Origin of the trip*. The driver is asked his last compulsory stop, whether is was to load or unload freight, or to take the vehicle. Note that the two possibilities are distinguished when asking the question, because the origin of the freight's trip can be different from the origin of the driver's trip. This distinction is not kept in the data.

- *Destination of the trip*. The driver is asked his next compulsory stop, whether is is to load or unload freight, or to take the vehicle.

- *Length of the trip*. The interviewer asks the trip length. The driver's answer can be approximate.

- *Empty or loaded*. The interviewer asks whether the vehicle contains freight or not.

- *Commodity type*. In the case where the vehicle contains freight, the interviewer asks its nature. When the semi-trailer holds a container, the commodity type is unknown to the driver. In the other cases, the driver generally knows what he is carrying, because he has documents describing the freight, the pickup and delivery times, as well as the route.

- *Freight amount*. The driver is asked how many tons of freight he is carrying. This data is also available on the documents the driver has.

- *Hazardous materials*. Some questions can concern specifically hazardous materials. The commodity and risk natures are displayed on specific plates on the vehicles, so that these questions don't have to be asked to the drivers.

This list of questions may vary with the circumstances. For example, some roadside surveys, dedicated to the study of freight traffic through tunnels, are focused on hazardous materials, so that the commodity nature is disregarded if it is not dangerous. An example French roadside survey questionnaire is provided translated in Appendix A.1.

On the whole, this list of questions reaches the objective of providing OD information. There is, anyway, a strong limitation. It is for example impossible to follow the route of the freight if there is a transfer in a warehouse, or to another transport mode. The transport chain, *i.e.*, the number of operations involved by the transport of a given shipment, cannot be observed using roadside surveys.

However, the drivers can answer other questions, which do not concern directly origins and destinations, but which can provide useful information on road freight transport. These questions are examined in the next section.

4.3 Construction of the new questionnaire and pretest survey

Roadside surveys provide opportunities to obtain information from truck drivers. As such, they are excellent occasions to have a closer look at road freight transport. Of course, truck drivers are not managers, and they are not necessarily well informed of the potentially complex processes which determine which trip they have to do and when. But observation shows that they are usually correctly informed about the transport operation they take part in.

In order to identify the kind of information truck drivers can provide, a new questionnaire has been designed. This questionnaire, provided in Appendix A.2, has been tested in the frame of a survey which is presented below. It addresses the following four topics: freight volume, transport organisation, specific schedule imperatives, location and duration of breaks. Before its efficiency is examined on each of these points, the survey is quickly presented, and some of its results are commented.

As explained before, there will be no in-depth econometric analysis in this discussion. Albeit a number of numerical results are presented, their role is to assess the usefulness of the new questionnaire, not to be useable economic indicators.

4.3.1 The survey

The survey took place between the 4th december and the 12th december 2007 at 50 kms of Bordeaux on the RN10 main road, which is part of one of the main routes between Bordeaux and Paris. Both traffic directions were surveyed. Due to the unavailability of police force, the roadside protocol presented in the previous section could not be applied rigorously. As a consequence, the drivers were interviewed on a nearby service area. For this reason, there may be a series of bias in the results presented below. 693 truck drivers were interviewed, 686 of these interviews yielded workable data.

Before getting into the detailed description of the original questions, the sample is briefly described. Some general remarks on interurban freight transport by road are made. 89.8% of the vehicles in the sample have 5 axles. The survey is located on a main road between two big cities, and on a major route of international freight traffic, notably between France and the Iberian peninsula, so that large vehicles, adapted to interurban transport, are expected. For the same reason, the nationalities of the vehicles is not surprising: 45.5% of the vehicles are French[3], 24.6% Spanish and 20.4% Portuguese. With respect to their origins and destinations, 42% of the trips are national, 58% international. 56.9% of the international trips are transit trips.

The vehicle types depend strongly on the number of axles. The trucks with 2 axles are mainly reefers or tautliners, the 3 and 4 axles are mainly flatbeds or dry bulk. About a half of the 5 axle vehicles are tautliners. The other types are by and large equally distributed, except for the containers, which are rare. Tautliner vehicles can be considered as general purpose vehicles, while the other vehicles are used for commodities with specific constraints (handling, temperature, safety, etc.). The use of specific equipments in road freight transport is thus significant. The distribution of vehicle types is relatively similar in both directions.

The commodity type is encoded along the French commodity type nomenclature called NST[4]. The commodity type is obtained for 86% of the loaded vehicles. When examined at the least detail level, the flows appear to be quite symmetric: manufactured products are the prevailing category with 41.7% of the loaded vehicles surveyed. The other most commonly met commodity types are: foodstuff (15.8%), agricultural products (11.4%), minerals and building materials (10.0%), chemicals (7.2%). The hierarchy does not depend on the direction: the commodity flows seem to be of the same nature in both directions.

This seeming symmetry of the commodity flows does not resist to a more detailed examination. Figure 4.2 illustrates the ten most transported commodity types, in each direction, described at the most detailed

[3]Concerning semi-trailers, the tractor unit and the trailer have distinct registration plate. They can be from two distinct countries. However, the proportions are very similar.

[4]At the most detailed level (NST3), the NST (*Nomenclature Statistique Transport*) distinguishes 176 commodity types. These types are regrouped in 52 groupes (NST2) and 10 chapters (NST1). An intermediary category, called sections, takes place between the groups and the chapters.

This nomenclature dates from 1974. It has become progressively obsolete, and has been replaced in 2007 by a new version. The databases examined in this chapter have been encoded with the old version.

level of the NST. It appears that eight of the ten most carried commodity types of each direction do not appear in the other direction. Intrasectorial trade is obvious when the commodity type is described with enough detail, as well as the spatial heterogeneity of production and consumption. If not detailed enough, a commodity typology underestimates this asymmetry.

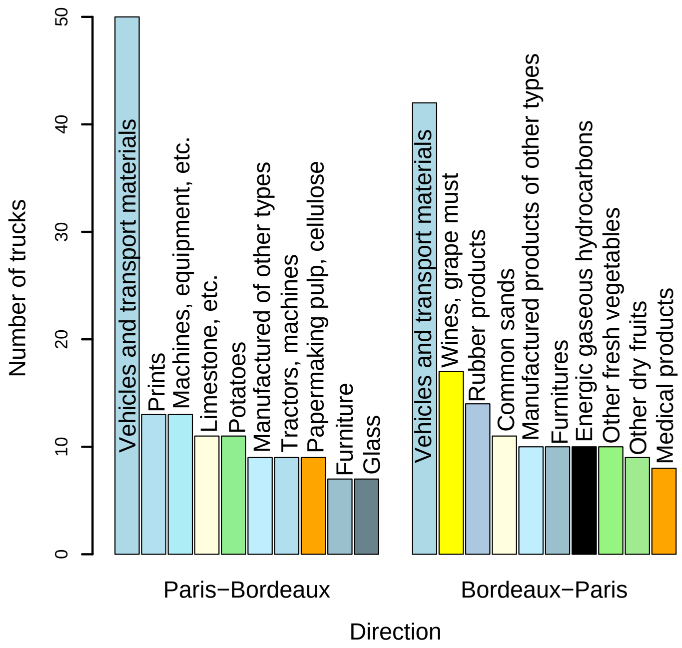

Figure 4.2: Commodity types according to the direction

10.8% of the vehicles interviewed were running empty. However, this percentage depends strongly on the type of vehicle, as illustrated by Table 4.1. Some transport techniques, such as tautliners, are quite versatile, and can address a large set of transport operations, while other are highly specialised, such as tankers and dry bulks. It is difficult to find backhaul freight for a more specialised vehicle. Furthermore, specialised vehicles often carry commodities transported on shorter trips: the flows are more intense, but more concentrated, which also makes finding backhaul freight difficult. These two effects explain jointly the results observed in Table 4.1.

The average trip length is 1176 km, its median 1103 km and its standard deviation 864 km. It depends on the commodity type, the number of axles, and the vehicle type. The average trip length by commodity

Vehicle type	Empty running
Reefer	3.8%
Curtain sider	6.3%
Rigid sider	8.2%
Flatbed	9.1%
Others	12.5%
Container	14.3%
Tanker	22.6%
Dry bulk	34.8%

Table 4.1: Distance run empty according to vehicle type

type ranges from 406 km for the vehicles carrying building materials to 1480 km for the vehicles carrying manufactured goods. Similarly, dry bulk vehicles cover on average 310 km, against 1550 km for reefers. The analysis will not be more detailed: this discussion is just intended to illustrate the close linkages between the trip lengths, the commodity carried, and the type of vehicles used. Road freight transport is an heterogeneous industry, providing heterogeneous services to a heterogeneous demand. This implies heterogeneous costs.

It would be possible to proceed to a myriad of analyses of this kind. However, this is not the objective of this work, so this won't be done here. The main outcomes of this paragraph are reminded: the need to examine in detail the commodity types before concluding on the symmetry of flows, the heterogeneity of flow natures and transport techniques (hence of transport costs, which, in turn, influence the location of activities and finally the commodity flows).

The points specifically addressed by the novel questionnaire will now be examined. To ensure a minimal homogeneity of the transport operations analysed, the following discussion is limited to the 5 axle vehicles, which represent about 90% of the sample.

4.3.2 The volume constraint

The weight capacity is a major technological constraint of freight transport. The ability of carriers to fill their vehicles determines their productivity, the objective being to use the minimal fleet to carry a given amount of commodities. The loading factor, *i.e.* the ratio of the weight of freight carried to the vehicle capacity, is thus a major indicator of the productivity of road freight transport. This ratio is used in almost all spatialised models to convert commodity flows into vehicle flows, which constitute the traffic[5].

As it has been explained in the previous section, the freight weight in a given vehicle is generally measured accurately by roadside surveys, because it is indicated on documents the driver has at its disposal. The vehicle capacity depends mainly on the number of axles, which is observed by the interviewers[6]. As a consequence the loading factor is easily computed; the influence of the weight constraint on the organisation and costs of the carriers can be assessed correctly. For the vehicles with 5 axles, Figure 4.3 illustrates the distribution of the loading factors rounded to one decimal.

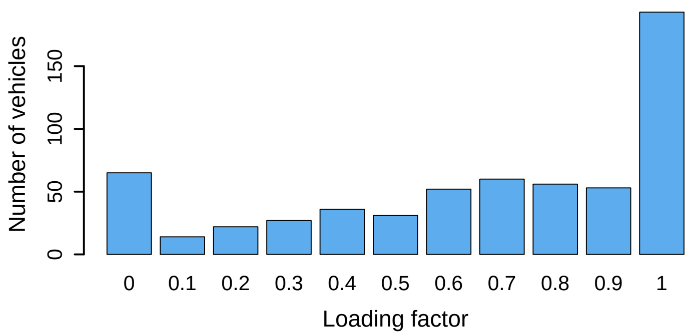

Figure 4.3: Loading factor distribution

Contrarily to the loading factor, which is closely monitored, the occupied volume is usually disregarded in transport statistics. The role of the volume constraint, and particularly its relative importance with respect to the weight constraint in road freight transport costs, is thus not well known. However, this is an important question of road freight transport policy. Indeed, the volume a truck can carry is limited, most often by the regulation limiting the vehicle's dimensions.

[5]For example, the loading factor has been chosen as one of the key factors of freight transport demand in the REDEFINE European project (NEI et al., 1999).

[6]There is a small complication for 5 axles vehicles: the total authorised weight is generally 40t, except when the trip is part of a larger combined transport operation, where the limit is lifted up to 44t. In the former case, the maximal freight amount is about 24t, whereas it is 28t in the latter case. We simplify the matter in this study as follows for 5 axle vehicles: if the freight amount is higher than 24t, we assume the loading factor is 1. If the freight amount is smaller than 24t, we assume the vehicle capacity is 24t.

It is currently impossible to assess the influence of this constraint on the basis of classic roadside surveys, because the freight volume is not observed. This question is absent from the classic questionnaire because the documents accompanying the freight do not detail its volume. However, roadside surveys can provide useful information on this question. Indeed, even if the drivers are not able at all to give the volume of their freight in m^3, they are approximately aware of the place the freight takes in their vehicle, and can tell if the vehicle is, say, half empty, or two thirds full. Therefore, this question has been added to the questionnaire.

If the vehicle is not empty, the interviewer asks the driver if a quarter, a half, three quarters or the totality of the vehicle's volume is used (see Appendix A.2, question Q7). Although some drivers think at first the same question is asked twice (they also have to say how many tons they are carrying), the question is eventually correctly understood. The answers for the 5 axle vehicles are illustrated in Figure 4.4.

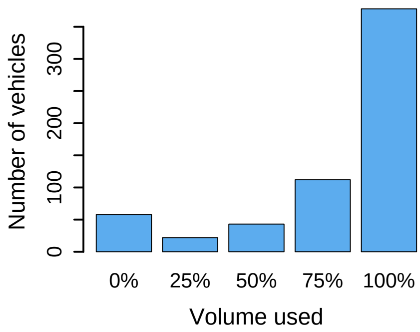

Figure 4.4: Share of volume used

The distributions are fairly similar. However, the volume and the weight constraint do not play the same role. Indeed, a truck full in weight is generally full in volume, but the converse is not true. 42.0% of the loaded 5 axle vehicles are full in volume and not in weight, whereas 26.1% are full both in volume and weight, and only 8.6% in weight and not in volume (Table 4.2). This tends to indicate the volume constraint is more binding than the weight constraint: increasing the weight limit while leaving the volume limit unchanged would only impact 8.6% of the vehicles in the sample, against 42.0% for a change in the volume limit with a constant weight limit.

Furthermore, as illustrated by Figure 4.5, the vehicles full in volume can be far from full in weight. This, combined to the fact that the

		Full volume	
		No	Yes
Full weight	No	23.2%	42.0%
	Yes	8.6%	26.1%

Table 4.2: Weight and volume constraints, RN10 survey.

correlation between the freight weight and the weight volume for loaded vehicles is only 0.16, implies that the weight alone may not be a correct measure of the productivity of road freight transport. Even though the volume measure is far from being accurate, the average loading factors in weight and in volume should be observed jointly. For the sample surveyed, they are respectively 73.8% and 88.1%.

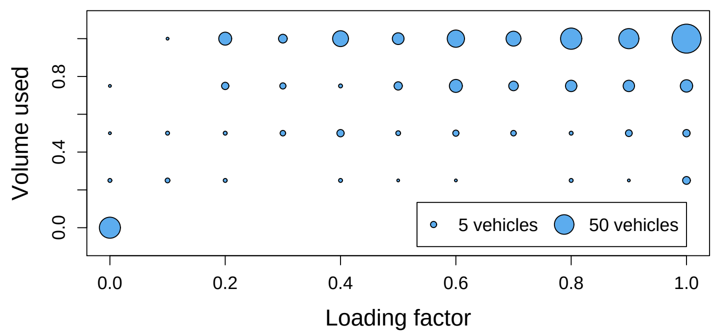

Figure 4.5: Loading factor and volume used in 5 axle vehicles

While the figures given in this section should be considered with care, due to the possible bias resulting from the survey protocol, the feasibility of measuring approximately the volume used in the vehicles is clear. Such data can be a useful basis for TS & W (Truck Size and Weight) studies[7], in which one difficulty is to evaluate the elasticities of road freight transport demand with respect to weight limit and vehicle dimensions. They also explain why increasing the useful volume while keeping the

[7]These studies form a subject of its own in freight transport economics and policy. Its complexity stems from the many issues related to maximum size and weight regulation: road dimensions, wear and tear, safety and passengers perceptions are all to be taken into account. See e.g. McKinnon (2005), which presents the UK and US TS & W discussions and researches.

outer dimensions unchanged is as much a major axis of innovation of the trucking industry as reducing the vehicles weights.

4.3.3 Transport organisation

Microeconomically, road freight transport does not seem to be a complicated technology. Consider the two inputs that are vehicles and drivers[8], they are generally assumed to be perfect complements: they must be used in fixed proportions to produce a given amount of transport, measured in vkm. The cost function of road freight transport is thus proportional to the vkm output, and road freight transport's speed is fully determined by the vehicle limit speed and the breaks the drivers must take.

The complementarity of drivers and vehicles is confirmed econometrically. For example, Delvaux and Duhautois (1998) found by fitting a CES production function which takes the number of vehicles and drivers as its arguments, that the elasticity of substitution between these two factors is 0.2, which is low.

However, motor carriers can, in some cases, organise themselves so as to decrease the travel time on a given origin destination pair. Indeed, the main limitation to the time a driver can drive is the compulsory night break. If two drivers are present in a vehicle (double crew), one of them can drive while the other one sleeps, so that the truck is virtually always moving. A more complex organisation, in relay, is also possible: the idea is to organise the trips of several drivers and vehicles so that each vehicle is always driven and each driver drives when the regulation authorises it[9]. In both cases, the vehicle is always running, so that its average speed is increased. For long distances, the carrier is thus able to reduce the travel time, provided the shipper agrees to pay the extra cost.

Roadside surveys can be useful in observing these practices. This is done with the questionnaire presented here. The interviewer asks "How many drivers are there on board?", and "Are you doing the same trip

[8]In doing so, we disregard the other inputs which are, for example, fuel, tires, insurance, tolls, and so on. The necessary amount of each of these inputs can be assumed proportional to the distance covered.

[9]For example, a driver drives a vehicle eight hours, from the point the freight is loaded, to a service area where another driver takes the truck and continues the trip. The first driver takes his night break in the service area. Ten hours later, he takes a truck which has just arrived from the other direction, driven by another driver. The other driver takes his night break on the service area, while the first driver goes back to his initial departure point with the truck. Such an organisation is quite complex to set up, and requires regular, significant and balanced flows. Furthermore, it is tiresome and very repetitive for drivers.

as the freight?", as well as, if it is a relay, the origin, destination, and length, of the part of the trip done by the current driver (questions Q11 to Q15).

Except for two cases, all the double crews and relays are observed for 5 axle vehicles. Their frequencies[10] are given in Table 4.3, as well as the corresponding average distances and speeds. These frequencies are quite low: it seems double crew and relay organisations are a minority. However, they should be taken with extreme care. Indeed, the objective of relay and double crew organisation is to avoid breaks, so that a survey made on a rest area most probably underestimates their importance. The average trip length and speed corresponding to each organisation are also indicated in Table 4.3. The double crew and relay organisation, which are more expensive than the simple crew organisation, also permit a higher speed. They are used on longer trips.

Organisation	Freq.	Dist. (km)	Speed (km/h)
Simple crew	93.1%	1182	33.9
Double crew	5.4%	2143	40.6
Relay	1.5%	1606	50.8

Table 4.3: Organisation of the transport operations.

From an economic standpoint, double crew and relay are expected to be used on long distances and for high value of time or high depreciation cost goods, such as agricultural products and foodstuff. This is seems to be confirmed by the comparison of commodity types with transport organisations in Table 4.4.

All these statements require confirmation from in-depth econometric analyses and more numerous data. However, the capacity of roadside surveys to yield detailed information on which such studies could be based is demonstrated.

4.3.4 Logistic imperatives

The increasing role of logistic imperatives imposed by shippers to carriers in the recent evolution of the freight transport market has been much discussed. The notion of logistic imperative is however not precisely

[10] The following convention has been taken to consider the case with both a double crew and a relay as a relay, thus reducing the double crew denomination to the case where the two drivers make the same trip as the freight from its origin to its destination.

Commodity type	Simple crew	Double crew	Relay
0 - Agricultural products	11.9%	19.4%	0.0%
1 - Foodstuff	18.0%	22.6%	0.0%
2 - Solid fuels	-	-	-
3 - Petroleum products	5.5%	0.0%	0.0%
4 - Metallurgical minerals	3.2%	0.0%	0.0%
5 - Metallurgical products	2.8%	0.0%	0.0%
6 - Building materials	9.5%	3.2%	0.0%
7 - Fertilizers	0.6%	0.0%	0.0%
8 - Chemicals	7.7%	9.7%	12.5%
9 - Manufactured products	40.8%	45.2%	87.5%

Table 4.4: Transport organisation and commodity type.

defined, and seems to be constituted of several aspects. One may quote, among others, the greater importance of the preferences of customers in the logistic organisations of the shippers, the closer integration of logistic decisions and production or marketing decisions, and the rationalisation of logistic and transport operations.

A first approach of the latter point has been attempted with the present questionnaire. Indeed, the interviewers asked the drivers both their expected arrival times, and imperative arrival times, making a clear distinction between them (questions Q18 and Q19 of the questionnaire). The existence of a specific logistic constraint, distinct from a classical expected arrival time, is thus investigated.

The answers are quite instructive, but not fully expected. For the 5 axle vehicles, the results are provided in Table 4.5. Most drivers answer they have no imperative arrival time. A small part of them answer they have an imperative arrival time, which is the same as the expected arrival time; in other words, the margin is null. It seems that in these two cases, the notion of imperative arrival time is mainly a matter of perception of the driver. By the way, the question has been ill received by many drivers, who considered it questioned their autonomy. Eventually, 16.2% drivers announced an imperative arrival time clearly distinct from the expected arrival time. For these drivers, the time margin increases with the trip's length.

It is difficult to identify any linkage between the presence of an arrival time imperative and the other variables in the survey. In particular, the presence of an imperative does not seem to depend on the commodity type, or on the presence of a double crew or a relay.

Imperative arrival time	Freq.
No	77.0%
Yes, no margin	6.9%
Yes, positive margin	16.2%

Table 4.5: Existence of an arrival time imperative.

On the whole, the existence and definition of logistic imperatives for motor carriers is not really striking. The large amount of drivers for whom there is no imperative, or no clear distinction between the expected arrival time and the imperative arrival time, can be interpreted as the result of a correct reliability of road freight transport. Motor carriers commonly give drivers a margin just large enough for the delivery to be done on time in case of an unexpected delay, thus ensuring the delivery time reliability. Nevertheless, it seems this question is beyond the reach of roadside surveys. Interviewing motor carrier managers may be a much more fruitful approach to this question. As a final note, recall, once again, that these proportions can be biased by the protocol of this survey.

4.3.5 Location and duration of breaks

For drivers to keep watchful and to drive safely, regular breaks are necessary. However, due to the strong competition between motor carriers, drivers tend to skip these breaks, so that they are strongly regulated. This regulation describes precisely when the breaks must be taken, and how long they must last. Subsequently, it leaves little choice to the drivers.

Nevertheless, there are some important questions related to the breaks the drivers must take. One of them is the effective speed of road freight transport, as already discussed in Section 4.3.3: motor carriers can introduce specific organisations to decrease travel times.

Another one is the congestion of service areas on highways and major roads. The breaks being compulsory, the drivers who don't find room on a service area will park somewhere else, which entails road safety as well as a risk of robbery. They can also reconsider their route choice. As such, it is both an important question for infrastructure operators and from a road infrastructure planning perspective.

The usefulness of roadside surveys in investigating these effects has been tested with the present questionnaire, where drivers are asked the number of breaks they have taken since their departure, the location and duration of their longest break, and its motive (questions Q20 to Q23). Unfortunately, their answers yield little useful information, which will subsequently not be presented here. Indeed, apart from choosing the place where they can park, the drivers have little initiative on their breaks. The data only illustrates the rules they follow. A geographical approach, where the route and parking choices of drivers would be analysed together with the occupancy of rest areas. This is not possible using roadside surveys alone.

4.4 Application of the questionnaire: the A10-A20 survey

The survey described in this section took place between May the 14th and May the 20th 2008, on a number of roads between the cities of Limoges and Poitiers, in France. This survey was aimed at studying how the regional passenger and freight traffic is shared between a number of major roads of this region, including the A20 highway passing near Limoges and oriented North-South, and the parallel A10 highway, located between 100km and 200km to the West. The classic roadside protocol presented in Section 4.2 has been applied, so that the possible biases of the previous section should not be feared here.

Initially, the questionnaire for freight vehicle followed the guidelines presented in Section 4.2. We have been offered the opportunity to append some additional questions to this questionnaire. It was impossible to use the questionnaire presented in the previous section, because of its length. As a consequence, given that the freight volume and transport organisation questions proved to be relatively more fruitful, the three following questions (respectively questions Q7, Q11 and Q12 of the previous questionnaire) have been kept:

- How much of your vehicle's volume is used?
- How may drivers are there on board?
- Are you doing the same trip as the freight?

Therefore, the resulting questionnaire corresponds to the first page of the questionnaire in Appendix A.2, *i.e.* questions Q1 to Q12.

Before getting into the analysis of the answers to these specific questions, the sample is briefly described. 630 vehicles have been surveyed. 75% of them are 5 axle vehicles, the average trip length is 514km, and 21.9% vehicles were running empty. This sample is thus not directly comparable to the previous one: the survey has obviously targeted a much more local and diffuse freight transport. This is confirmed by the commodity types: the category 9 “Manufactured products” constitutes only 32.5% of the sample, whereas the category 1 “Foodstuff” constitutes 24.2%.

Despite these dissimilarities, the respective roles of the weight and volume constraint seem to be relatively stable. The figures of Table 4.6, which correspond to the 5 axle loaded vehicles, are by and large consistent with those of Table 4.2. However, it seems the weight constraint is a bit more important in this survey than in the previous one. The correlation between the freight weight and the freight volume of 5 axle loaded vehicles is 0.18. The average loading factor of loaded vehicles is 71.8% in weight, and 84.1% in volume. They are both lower than in the previous survey, but their relative orders of magnitude are similar. On the whole, the discussion of the previous survey is confirmed by these figures.

		Full volume	
		No	Yes
Full weight	No	28.0%	34.4%
	Yes	11.8%	25.7%

Table 4.6: Weight and volume constraints, A10-A20 survey.

By asking the drivers if they do the same trip as the freight and how many they are onboard, double crew and relay organisations are identified. The results are given in Table 4.7. We find again that double crew and relays are relatively rare, and that the average distance is greater with double crew or relay organisations. The speed is not available, as the departure and arrival time have not been asked. Finally, relays are much more numerous here than in the previous case, whereas there are much less double crews. We have no clear interpretation of these differences, which can as well derive from local specificities of transport demand as from the habits of motor carriers. More surveys will certainly prove useful in investigating the drivers of transport organisation decisions.

4.5 Conclusion

The classic French roadside survey questionnaire is generally used to obtain origin destination data. In the case of freight transport, vehicles

Organisation	Freq.	Dist. (km)
Simple crew	92.7%	572
Double crew	1.9%	829
Relay	5.3%	735

Table 4.7: Organisation of transport operations, A10-A20 survey.

are diverted from the traffic, and their drivers are asked their trip's origin, destination, length, the nature and amount of the commodities they carry. Although truck drivers are not motor carrier managers, a possibility remains that this list of question does not make the most of what they know. The objective of this chapter was to investigate this possibility.

In order to determine how roadside surveys can be enhanced, an original questionnaire has been developed and administered in the frame of a roadside survey. Apart from the classic questions, four additional subjects are addressed by this questionnaire: the freight volume, the transport organisation, the presence of specific logistic imperatives, and the breaks. Some of these questions have been used afterwards in another survey, using a reduced questionnaire.

Generally, the drivers can tell precisely how many tons of freight they carry, because they bear documents on which this information is written. This is not the case for the freight's volume. However, the drivers are able to tell approximately how much of their vehicle's volume is occupied by the freight they carry yields useful results. The role of the volume constraint can thus be measured: it even seems to be at least as important in determining road freight transport's productivity as the weight constraint.

If some points of road freight transport operation can be complex, the ratio of one driver per vehicle is generally assumed universal. However, other organisations are sometimes observed, such as double crews or relays. These organisations can be observed by the interviewers of roadside surveys with simple questions. Roadside surveys can thus yield useful data on these practises, on which studies could be based to investigate when and why such organisations are chosen by motor carriers.

Investigating logistic imperatives has proven less easy. The drivers have been asked whether they have an imperative arrival time distinct from the expected arrival time. Most answered that they had no imperative, a fraction answered that their expected arrival time was imperative, and about a sixth had both an expected and an imperative arrival time. However, it has been difficult to discern any relation between this answer and the other variables of the model.

Similarly, the drivers have been asked a series of questions about their breaks. Although the choice of the place the drivers park for their breaks, and the related problems of safety, theft, and service area congestion are important, the answers of the drivers mainly reflect the regulation in force. It seems roadside surveys can provide little information on this question, if they are not combined with other information sources, such as, for example, service area occupancy levels.

Finally, it should be noted that the surveys presented in this chapter have a major shortcoming. Indeed, the distinction has not been made between public and private carriers. It is certainly possible to ask drivers this question, and highly relevant, insofar as public and private road freight transport are not similar. Comparing the organisations of these two activities may certainly yield interesting result. This is a recommended extension to this work.

As explained before, the objective of this chapter was not to proceed to econometric analyses, but to investigate the possibility to enhance roadside surveys, so that they provide useful additional information on freight transport. This possibility appears to be significant. Roadside surveys prove efficient in measuring the role of the volume constraint, even if the approach is approximate, and this role appears to be quite important. They also prove efficient in observing specific road freight transport organisations, such as double crews and relays, two questions on which depends closely the productivity of road freight transport. If these questions were asked systematically, given the number of roadside surveys done in France each year, abundant data would be quickly available to analyse these questions in more depth, to a small extra cost.

Chapter 5

A note on the econometric validity of the EOQ model

5.1 Introduction

Freight transport demand models usually represent how shippers choose transport modes and itineraries. Introducing the shipment size decision in these models is a complex task. Among the difficulties this approach meets, one is critical: the lack of econometric validation of the common microeconomic shipment size choice models on a large, heterogeneous shippers population.

The choice of shipment size is a central question of freight transport modelling insofar as it is closely linked to other characteristics of the demand for freight transport, such as the preferences of shippers concerning the different transport modes. This question is relatively well addressed from a theoretical point of view: a number of microeconomic models explain how shippers decide the shipment size depending on a number of parameters.

These models generally build on inventory theory models, in particular the fundamental EOQ model. The EOQ model represents the choice of shipment size by a shipper as a trade off between some transport and inventory cost components. It is easily generalised to explain – at least partially – the choice of transport mode by shippers, on the basis of few more parameters.

One of the reasons for which such microeconomic models of the choice of shipment size have not been integrated into the larger framework of freight transport demand modelling is that it is difficult to assess econometrically their outreach. This question is crucial: whereas a model such as the EOQ model (presented in Chapter 2) has every reason to be relevant for a firm for which the model's hypotheses hold, it is much less likely to hold simultaneously for all firms of all sectors.

However, the econometric analysis of the EOQ model would require a database describing accurately shipments by their characteristics such as their weight, volume, conditioning, commodity nature, value, etc. Although they are not common in the frame of transport modelling, such databases exist. Nevertheless, even the most basic inventory models require a description of the shipper-receiver relationship. For example, the EOQ model, which is not very demanding in terms of data, requires that the annual rate of the commodity flow between the shipper and the receiver is measured. This particular variable is generally not available in shipment databases. As a consequence, most shipment databases are useless when it comes to estimating inventory models, whatever the number of observations.

The particularity of the French shipment database ECHO is that it contains relatively few observations: only 10,000 shipments are surveyed; as a comparison, the Sweden Commodity Flow Survey contains millions of shipments. This apparent small size is compensated by the large amount of variables describing each shipment. In particular, the relationship between the shipper and the receiver is observed in detail. The total commodity flow rate between the shipper and the receiver, for example, is measured. Furthermore, the way the transport operation has been achieved is also described with great detail. It it thus theoretically possible to assess the EOQ model with the ECHO database[1]: this is the object of this chapter.

The chapter proceeds as follows: the EOQ specification is given in Section 5.2. Section 5.3 then explains concretely how the ECHO database is going to be used to assess econometrically the EOQ model. Section 5.4 presents the model specification and estimation. An extended model is specified and estimated in Section 5.5, before the chapter is concluded in Section 5.6.

5.2 The EOQ model

The EOQ model has already been described in Chapter 2. Its main elements are briefly reminded. Consider a firm sending a regular com-

[1]In fact, the most important econometric result of this chapter, which will be presented in the sequel, is available in Szpiro (1996). It has been indeed obtained incidentally as the authors were studying the structure of freight transport prices. However, the authors did not interpret their results from the perspective of microeconomic modelling of the choice of shipment size.

modity flow of constant rate Q from a given location to another by a given transport mode. The transport operations being discrete by nature, commodities are carried by shipments. We assume that all the shipments are of the same size s, and that each shipment is dispatched as soon as there are enough commodities at the origin. The average origin stock level is thus $s/2$.

The choice of shipment size depends on the structure of the equilibrium prices of freight transport. As a first approximation, the transport cost is assumed to consist of a fixed cost b independent on the shipment size, and a variable cost $K.s$ proportional to the shipment size. The commodity value of time is denoted a. The travel time is denoted t. The cost per period of time is denoted g. We have:

$$g(s) = \frac{a.s}{2} + (a.t + K)Q + \frac{Q.b}{s}. \tag{5.1}$$

The optimal shipment size is obtained by minimising this cost function. It depends only on a, Q, and b:

$$s = \sqrt{\frac{2Qb}{a}}. \tag{5.2}$$

Quite intuitively, the shipment size does not depend on costs which are proportional to the shipment size, such as the pipeline inventory cost $(a.t.Q)$ or the proportional component of the transport cost $(K.t)$.

This model can easily be estimated by linear regression. Indeed, by taking the logarithm of both sides of Equation (5.2), we obtain:

$$\ln s = \frac{1}{2}\ln Q + \frac{1}{2}\ln b - \frac{1}{2}\ln a + \frac{1}{2}\ln 2. \tag{5.3}$$

This equation constitutes the basis of the econometric analysis proposed in the following sections. Note that this equation remains valid whatever the transport mode used by the shipper. From this standpoint, the only difficulty stems from the b parameter: Q and a do not depend on the transport mode chosen by the shipper (at least not directly).

Of course, the EOQ model is not able to account for shippers using two modes simultaneously for a single commodity flow. But such shippers are quite rare, and a model explaining jointly the choice of a shipper to use two transport modes simultaneously and the shipment sizes for each mode is far from simple (see Chapter 7).

5.3 Relevant variables and model specification

The ECHO database describes 10462 shipments. This information can be categorised into four groups: description of the shipping and of the receiving firms, description of the relationship between these two firms, description of the shipment, and description of the transport operation. In each of these categories, questions related to economic, logistic or transport related aspects are available. As a consequence, each shipment is described by hundreds of variables.

Before addressing the identification of the variables which can prove useful in this study, a word should be said about the sampling method. The ECHO survey has been motivated by two important scientific objectives, among others: first, investigating in deep detail the logistic choices of firms. Second, observing in detail how shippers choose transport modes. Quite obviously, given the average sizes of shipments sent by road or plane on one hand and by rail on the other hand, the relative frequency of shipments sent by rail is extremely low. In order to have a workable sample of shipments sent by transport modes other than road, the shipments sent by rail, combined transport or inland waterway have been strongly oversampled. However, we identified no reasons for which this would bias the estimation of an econometric disaggregate model.

We will now determine the relationship which will be estimated using the ECHO database. This relationship will be based on Equation (5.3). This equation strongly incites us to use $\ln s$ as the dependent variable. The shipment size is given in the ECHO database, both in weight and in volume. The weight is chosen as the measure of shipment size[2]. Summary statistics of s and $\ln s$ can be found in Table 5.1.

Choosing the explanatory variables is a little bit more complicated. Our objective is to find the most relevant measures of the three variables of Equation (5.3), *i.e.* Q, a and b, among the several variables available in the ECHO database. Some of the potential explanatory variables available in the ECHO database concern the demand-side of the freight transport market, *i.e.* they describe characteristics of shippers. Other

[2]The effect of the capacities of freight transport vehicles on the choice of shipment size will be disregarded here, so that choosing the weight or the volume variable should be of little consequence. However, the shipment volume is measured in m^3, with zero decimals, whereas the shipment weight is measured in kg, with zero decimals, so that it constitutes a much more accurate measure. As a consequence, the shipment weight is chosen as the measure of shipment size, and is also denoted s, as no ambiguity is possible.

variables concern the supply-side of the freight transport market, such as the transport price or the transport techniques used. They are examined separately.

Demand-side explanatory variables The rate Q of the commodity flow between the shipper and the receiver is not available in the ECHO database. The closest available variable is the total commodity flow denoted Q_{tot}, independent of the commodity type. A significant amount of values are missing. The shipments with missing values are not taken into account. Using Q_{tot} instead of Q without care will lead to underestimating the influence of Q. Unfortunately, there are few alternatives.

There is a chance that this bias is limited. Indeed, in the ECHO survey, the shippers and receivers are establishments, *i.e.* physically well delimited components of firms. At this level of disaggregation, each establishment has one or a few, well identified functions. Therefore, configurations such as production units sending many different commodities to many distinct receivers, or retail centers or plants gathering many commodity types sent by many distinct providers are much more probable than simultaneous flows of different commodity types between a unique pair of establishments. Although this argument lacks numerical support, we will content ourselves with Q_{tot}.

The value of time a, of course, is not available either. Fortunately, in this case, there is a good candidate to replace it. Indeed, for 64.5% of shipments, the market value excluded tax is indicated. Combined with the shipment weight, it is possible to calculate the value density of these shipments. The value density is denoted a_{dens}. Using a_{dens} instead of a is a strong hypothesis. This step can be improved by taking into account the fact that the opportunity cost of capital depends on the economic sector of the shipper and the receiver, and by taking into account the organisation of the supply chain between the shipper and the receiver (make-to-stock, make-to-order, vendor-managed-inventory, etc.) Although these extensions have not been addressed in this simple study, they may prove to be fruitful, and are, to a certain extent, possible with the ECHO database up to a certain extent.

Note that the summary statistics provided in Table 5.1 already speak in favour of a log specification: the distribution of the variables are very skewed when they are taken without transformation. On the contrary, they are much more symmetric after the logarithm transformation.

Supply-side explanatory variables The vast amount of variables necessary to describe the available freight transport alternatives from the

Variable	Min.	Q1	Med.	Mean	Q3	Max	NA's
s (t)	0.00	0.05	0.65	19.58	7.8	10,800	0
Q_{tot} (kt/y)	0	1	18	2,126	350	63,000	1934
a_{dens} (k€/t)	0.00	1.07	4.56	59.37	20.00	10,400	3715
$\ln s$	-6.91	-3.00	-0.43	-0.65	2.05	9.29	0
$\ln Q_{tot}$	-6.90	0.18	2.89	3.02	5.86	13.35	1934
$\ln a_{dens}$	-0.94	6.97	8.43	8.49	9.9	16.16	3715

Table 5.1: Basic EOQ continuous variables summary statistics

demand's standpoint can be categorised into two groups: technical compatibilities, and prices. Technical compatibilities are important: transporting liquid bulk is not the same thing as transporting foodstuff; each operation requires specific tools. However, this dimension will not be analysed in great detail here, apart from the question of modes.

On the contrary, prices are central to the estimation of the EOQ model. Under the assumption made earlier that prices are linear for a given transport mode, the shipment size only depends on the fixed term b. Transport prices are available in the ECHO database, so that b can theoretically be measured. This is a complex task: b is not directly observed. It must be estimated by rebuilding price schedules (*i.e.* prices as functions of shipment sizes), which itself depend on the distance, the transport mode, and many other parameters.

Another strategy is chosen: by definition, b consists of all the costs that do not depend on the shipment size. These costs typically derive from administrative operations, loading, unloading, handling, and control operations, which depend loosely on variables such as distances and vehicle capacities. It is thus more relevant to consider the access cost depends on some technical variables such as the transport mode, the number of transshipments, etc.

These variables are available in the ECHO database. The transport operation of each shipment is described in detail. In particular, when several transport modes have been used with transshipments between them, each stage of the transport operation is detailed. We won't go into that level of detail. The transport modes used will be summarise here by the main mode variable, denoted M. The main mode is determined by a set of priority rules, generally favouring the heavier mode, except for the case of air transport[3]. The modes distinguished are: private carrier,

[3]For example, if road transport and rail transport are used together, the main mode is rail transport. If sea transport, or air transport, are used in a transport

common carrier, rail, combined transport, inland waterway, sea, and air. Given that Equation (5.2) holds whatever the transport mode chosen by the shipper, and that only b changes, b will be considered as a function of the transport modes.

M	Number	Freq.
Private carrier	1727	0.17
Common carrier	6648	0.64
Rail	224	0.02
Combined transport	133	0.01
Inland waterway	44	0.00
Sea	825	0.08
Air	859	0.08
NA's	2	0.00

Table 5.2: Main transport mode summary statistics

On the basis of the previous discussion, the following specification will be estimated:

$$\ln s = \beta_Q \ln Q_{tot} + \beta_a \ln a_{dens} + \sum_{mode} \beta_{mode} X_{mode} + u, \tag{5.4}$$

where u is the error term and $\{X_{mode}\}_{mode}$ dummy variables indicating the transport mode used. Note that b plays the role of a mode-dependent intercept.

5.4 Estimation of the EOQ model

The model is estimated using ordinary least square regression. The results are given in Table 5.3.

The coefficients are highly significant. The R^2 coefficient is close to 0.8. Complementary analyses of the residuals, presented in Appendix B.1, do not invalidate the ordinary least square hypotheses. On the whole, specification (5.4) seems adequate. The estimated model is:

$$\begin{aligned} \ln s \quad = \quad & 0.50 \ln Q_{tot} - 0.44 \ln a_{dens} + 1.05 X_{\text{common carrier}} \\ & +1.46 X_{\text{private carrier}} + 3.42 X_{\text{rail}} + 2.09 X_{\text{combined}} \\ & +4.37 X_{\text{waterway}} + 2.89 X_{\text{sea}} + 1.47 X_{\text{air}} \end{aligned} \tag{5.5}$$

operation, they are its main transport mode.

Coefficients	Estimate	Std. Error	t-value	
β_Q	0.50	0.01	73.27	***
β_a	-0.44	0.01	-37.57	***
$\beta_{\text{common carrier}}$	1.05	0.11	9.47	***
$\beta_{\text{private carrier}}$	1.46	0.11	12.87	***
β_{rail}	3.42	0.18	19.30	***
β_{combined}	2.09	0.20	10.31	***
β_{waterway}	4.37	0.33	13.05	***
β_{sea}	2.89	0.13	21.49	***
β_{air}	1.47	0.14	10.29	***
N	10462			
NA's	4741			
R^2	0.795			
Adjusted R^2	0.795			

Table 5.3: Estimation of the EOQ model

β_Q is close to 0.5, and β_a relatively close to -0.5, as predicted by the theory. It seems that the EOQ model can prove fairly efficient in explaining the shipment size on a large and heterogeneous population of shipments. A simple equation, such as Equation (5.4), explains about 80% of the variance of the shipment sizes in ECHO for which the explicative variables are available (*i.e.* about 55% of the sample). As such, it seems that despite its simplicity, the EOQ shipment size model has a wide econometric outreach. Its simplicity and outreach, combined with its consistency with the mode choice decision, make it a good candidate as a shipment size model in the frame of a large scale freight transport demand model.

Stricto sensu, the EOQ model given by Equation (5.3) is only partially consistent with the estimated parameters. Indeed, even if β_a is close to -0.5, the hypothesis that $\beta_a = -0.5$ is clearly rejected. This can be interpreted as a limit of the EOQ model, or as the inadequacy of the value density a_{dens} as a representation of the commodity value of time. In particular, commodities with a very low value density can still have an important role in a supply chain, or a high warehousing cost, that would explain that $a \gg a_{dens}$. In that case β_a would be underestimated. Further analysis is necessary to investigate this point.

5.5 Exploratory estimation of an extended EOQ model

There are other variables available in the EOQ survey that can influence the shipment size. They will be described and their influence tested in this section. They are examined separately of those identified in Section 5.3 because despite their potential econometric significance, they do not fit in the economic theory introduced in Section 5.2. We thus leave the field of structural econometrics.

The most striking feature of the EOQ model is that the distance between a shipment's origin and destination does not influence the shipment size. However, this variable always play a central role in transport modelling. Its influence is thus tested now. Note that the distance is described by different variables in the ECHO survey: as-the-crow-flies distances, denoted d, are retained here[4]. Given the asymmetry of the distribution of d, $\ln d$ will be used.

Apart from the transport mode, many variables can influence the fixed transport price parameter b. Among them, three will be examined: the number of agents which intervene, physically or administratively, in the management and the achievement of the transport operations, denoted N_{interv}; the number of elementary transport operations, denoted N_{trips}; the fact that the shipment is isolated or part of a round, denoted O. The two latter variables, when available, only concern the part of the transport operation which has been realised in the UE15 territory. Summary statistics of these variables are given by Table 5.4. Variable O is summarised in Table 5.5.

Variable	Min.	Q1	Med.	Mean	Q3	Max	NA's
d (km)	1	74	278	1253	611	18840	8
$\ln d$	0.00	4.30	5.63	5.44	6.42	9.84	8
N_{interv}	0.00	1.00	2.00	2.76	3.00	12.00	720
N_{trips}	1.00	2.00	2.00	2.059	3.00	8.00	720

Table 5.4: Extended EOQ continuous variables summary statistics

We expect N_{interv} and N_{trips} to influence the transport access cost b,

[4]Distances by the shortest path by road are also available in the ECHO survey. But there are many missing values (19.5%), whereas as-the-crow-flies distances are available for almost all the shipments (only eight missing values). Finally, the correlation between these two variables is 0.88, which is quite high.

O	Number	Freq.
Isolated shipment	7647	0.73
Part of a bundle	989	0.09
Part of a round	1767	0.17
NA's	59	0.01

Table 5.5: shipment organisation summary statistics

since more intervenants, and more trips normally mean more transshipment operations. On the contrary, shipments transported in bundles or in routes should meet lower access costs. The distance should have no influence on the choice of shipment size: this will be tested.

Model specification In order to keep the estimation as simple as possible, and since there is no convincing microeconomic model relating b to these variables, the following specification will be estimated:

$$\begin{aligned}\ln s &= \beta_Q \ln Q_{tot} + \beta_a \ln a_{dens} + \sum_{mode} \beta_{mode} X_{mode} \\ &\quad + \beta_d \ln d + \beta_{interv} N_{interv} + \beta_{trips} N_{trips} \\ &\quad + \beta_{bundle} X^O_{bundle} + \beta_{round} X^O_{round} + u,\end{aligned} \tag{5.6}$$

where X_i^O equals 1 if $O = i$, 0 else; and u is the error term. Note that $\beta_{isolated}$, the parameter corresponding to the case of isolated shipments, is set to zero by convention.

Estimation The model is once again estimated using ordinary least square regression. The results of the estimation are given in Table 5.6. A basic analysis of the residuals does not reject the ordinary least square hypotheses (Appendix B.2).

All the coefficients are significant: each of them has a real influence on the shipment size. However, the R^2 is not much improved by introducing the new variables. As it is confirmed by the analysis of covariance in Table 5.7, much of the explanatory power of the model comes from $\ln Q_{tot}$ and $\ln a$, *i.e.* the core of the EOQ model.

The model is not much modified by the introduction of the other variables. The estimated coefficients of $\ln Q_{tot}$ and $\ln a$ remain reasonably (although not exactly) close to the theoretical values. The hierarchy of the main transport modes is not deeply changed. However, one can note than whereas the private carrier and common carrier where significantly

Coefficients	Estimate	Std. Error	t-value	
β_Q	0.44	0.01	61.94	***
β_a	-0.43	0.01	-37.57	***
$\beta_{\text{common carrier}}$	0.95	0.12	7.57	***
$\beta_{\text{private carrier}}$	0.97	0.13	7.26	***
β_{rail}	2.78	0.19	14.56	***
β_{combined}	1.60	0.23	7.04	***
β_{waterway}	3.91	0.38	10.42	***
β_{sea}	1.59	0.19	8.57	***
β_{air}	0.34	0.18	1.85	.
β_d	0.21	0.01	14.60	***
β_{interv}	0.16	0.02	8.92	***
β_{trips}	-0.40	0.02	-21.12	***
β_{bundle}	-0.62	0.07	-9.38	***
β_{round}	-0.69	0.06	-12.18	***
N	10462			
NA's	5134			
R^2	0.827			
Adjusted R^2	0.827			

Table 5.6: Estimation of the extended EOQ model

Variable	Df	Sum Sq.	Mean Sq.	F-value	
$\ln Q_{tot}$	1	13,263.7	13,263.7	7452.8	***
$\ln a_{dens}$	1	28,915.7	28,915.7	16247.5	***
M	7	1,563.8	223.4	125.5	***
$\ln d$	1	345.7	345.7	194.3	***
N_{interv}	1	95.8	95.8	53.8	***
N_{trips}	1	788.6	788.6	443.1	***
O	2	372.3	372.3	104.6	***
Residuals	5,314	9,457.3	1.8		

Table 5.7: Analysis of variance of the extended EOQ model

different in the previous estimation, this is not the case anymore with the extended model. The difference was due to the number of trips: indeed, the average number of trips when the main transport mode is private carrier is 1.02, while it is 2.12 when the main transport mode is common carrier.

N_{interv} has the expected positive effect: a larger number of agents intervening on the chain transport seems to imply a larger b, thus a larger shipment size[5].

What is surprising is the negative sign of N_{trips}, while it was also supposed to be positive. But predicting that β_{trips} would be positive was forgetting the economic rationale of hub-and-spokes transport networks: such networks are especially designed to handle efficiently small shipments. As a consequence, the transshipment cost is smaller in a hub-and-spokes transport network than in more direct transport organisations[6]. Note that this is not contradictory with the assessment that β_{interv} is positive. Indeed, these two variables are hardly related: their correlation is only 0.26.

The influence of the transport organisation variable O on the shipment size is intuitive: shipments sent in bundles and in routes certainly share some fixed transshipment and handling costs, thus their smaller size[7].

The influence of $\ln d$ is most probably the hardest to comment. A

[5]Besides the fact that more agents certainly means more administrative operations, thus a higher cost, a second interpretation can be proposed to this positive effect. Some shippers send large and regular amounts of freight, and are thus a source of constant and significant revenue for carriers. Other shippers are more irregular: they are not able to forecast how many commodities they will send, and are a source of uncertainty of the transport demand the carriers face. Given the operational constraints of carriers, an uncertain demand is a source of costs. A mean for carriers to distinguish regular shippers from irregular shippers is to bind long term contracts with the formers, with reciprocal commitments in terms of prices and volumes, so that the large, regular shippers have an incentive to accept such contracts. Now, if the irregular shippers, which have no long term contract with any carrier, tend to access to the freight transport supply through more go-betweens (such as freight forwarders) than big, regular, shippers, that would explain why they should face higher transport costs. This question is discussed from a more theoretical perspective in Chapter 6.

[6]In fact, taking into account the number of transshipments is just a second step into the technical characteristics of the transport operation chosen by the shipper (the first step being the transport main mode variable). Should the EOQ model be applied in a large scale spatialised freight transport demand model, this should be kept in mind: the transport options should be distinguished by mode but also by number of transshipments: this would form a sound basis to simulate the respective demands for direct and hub-and-spoke freight transport.

N_{interv}, on the contrary, can certainly be disregarded in a first step. It brings little explanatory power to the model (see Table 5.7 in Appendix B.2) and there is little microeconomic basis to model it.

[7]Despite the fact that this variable gives not as much explanatory power to the model as N_{trips}, it may prove much more interesting in a urban framework. For example, in the FRETURB model of freight transport in urban areas, the number of stops in rounds is considered a central variable of the organisation of carriers, closely linked to the distance covered by heavy good vehicles (Chapter 1).

possible explanation is that unused vehicle capacity is more expensive on longer distances. Motor carriers may then adjust their prices in order to induce the shippers to choose shipment size which will optimise their loading factors. The shippers may then send bigger shipments. A modelling approach of this kind of phenomenon can be found in Chapter 6. It can also be an outcome of the structure of the logistic network on the demand (see Chapter 3): if large flows go from plants to regional distribution centers where they are then dispatched, for example towards local retail centers, with a logistic disconnection between both operations (meaning shipments are not targeted towards a given retail center when they leave the plant, but only when they leave the regional distribution center), the macroscopic outcome is that bigger shipments move over longer distances. Coefficient β_d may then result from the structure of the demand.

5.6 Conclusion

In classic transport models, it is generally assumed decisions such as mode choice or itinerary choice fully derive from the characteristics of passengers or commodities. What is relevant in the case of passenger transport modelling, where passengers usually determine themselves the conditions of their trips, is inadequate in the case of freight transport: commodities take no decisions, shippers do (see Chapter 3.) Nevertheless, the characteristics of shippers are usually not explicitly taken into account.

One of the unfortunate consequences of this fact is that in general, freight transport databases pay attention to the commodities transported but not to the shippers sending these commodities. In the particular case of shipment databases, characteristics of shipments (size, commodity type, conditioning, etc.) are available, as well as characteristics of the concerned transport operations (transport mode, etc.)

It appears though that even in the case of the most simple microeconomic shipment size models such as the EOQ model, the relationship between the shipper and the receiver plays a central role. If this relationship is not described (which is the case in most shipment databases), these models cannot be estimated.

The French shipment database ECHO, which describes a rather small set of shipments with many details concerning their characteristics, how they have been transported, their shippers and receivers, and so on, contains one important variable for the estimation of the EOQ model: the commodity flow between the shipper and the receiver. Therefore, it has been been possible to estimate the EOQ model with the ECHO survey.

Given that the estimation concerned a set of shipments of all kinds transported by all kinds of modes, the estimation works surprisingly well. The theoretic EOQ model (which is valid whatever the transport mode used) is rejected by the estimation, but the coefficients obtained are still relatively close to the theoretical values. The fitted model performs efficiently in explaining the choice of shipment size. The consequence is that the main explanatory variables of the choice of shipment size are the value density, the commodity flow rate between the shipper and the receiver, and the fixed transport cost component.

A first approach has been tempted to observe more complex effect. Including the number of trips of the transport operation variable improves significantly the model. We observe that a higher number of trips is associated with smaller shipments. This can be interpreted as the result of the strong rationalisation of transshipment costs in hub and spoke transport networks. Other variables such as whether the shipment is sent in a bundle or not, and the influence of the distance, are also significant. However, they bring little explanatory power. The influence of the distance, in particular, is quite complicated to interpret.

The econometric approach has been kept as simple as possible. This chapter should be considered as a preliminary approach to the econometrics of shipment size. The ECHO survey can be used much more deeply, but this would require refining the EOQ model and econometric methods less simple than the ones used here (in particular to handle vehicle capacity constraints, and to estimate more complicated specifications of the transport fixed cost component). Besides, there are many missing values: variables such as value density are only partially available.

On the whole, the EOQ model seems to be a very good candidate as a microeconomic shipment size model in a spatialised freight transport model with explicit representation of the shipment. One of its advantage is that it is strongly consistent with the choice of mode: the overall cost of transporting a commodity flow with a given transport mode depends on the shipment size. However, measuring this overall cost once the shipment size is available is quite simple, and requires only classical freight transport model variables, such as transport unit cost and travel time.

Part III

Microeconomic Analysis

Chapter 6

On shipment size and freight rates: technical constraints and equilibrium price schedules

6.1 Introduction

Road freight transportation is usually considered as an industry with constant returns to scale. This statement is made from the perspective of carriers: the number of trucks necessary to carry a given amount of freight, hence the transport cost, is assumed proportional to this amount, up to a loading factor which stands for the average amount of freight carried by each truck. It is also made from the perspective of shippers: the cost for a shipper to send a given amount of freight is assumed proportional to this amount (e.g. sending twenty tons is twice as expensive as sending ten tons.) These statements are naturally extended to freight flows: for a carrier, carrying twenty tons per week is twice as expensive as carrying ten tons per week, whereas it is twice as expensive for the shippers to have twenty tons per week carried by road as ten tons per week.

The statement according to which the trucking technology is endowed with constant returns to scale is relatively consensual. It relies both on the observation that the trucking industry is neither constituted of atomic carriers nor oligopolistic[1], and on a number of empirical studies

[1]There were 11,932 long-haul (the corresponding designation being "interurban" in France) freight motor carriers in France in 2006 (SESP, 2008). 10% of these firms realised 73% of the total interurban turnover (Abraham, 2006). The market structure

investigating the structure of costs in the trucking industry[2].

Given that the road freight transport market is quite competitive, the relatively small size of motor carriers and the constant returns to scale of the trucking technology, prices should be equal to the marginal production costs, and transport costs are expected to present constant returns to scale from the perspective of shippers. This assumption is implicitly made in most freight transport models and studies, where rates are expressed per ton, independently of shipment size[3]. However, this is not observed: first, freight rates are highly variable, and do not seem to stick to marginal costs. Second, the price of road transport cost is not the only relevant element from the perspective of shippers: the time spent by commodities during the transport operation is also important. However, this time is not taken into account in an explicit way by carriers, to whom the freight carried does not belong, so that it does not appear in freight rates.

The first of these two arguments has motivated a long discussion concerning whether or not this variability of rates is due to price discrimination. Two of its major points will be presented. The first one was raised by Olson (1972), who demonstrated that even under regulation[4], the American freight transport prices showed a high variability; too high, according to her, to be simply explained by marginal costs. She claimed

---

is thus characterised by the presence of both moderately large firms and very numerous atomic firms. At first glance, this tends to reject both economies and diseconomies of scale.

Nevertheless, the complex relationships the carriers foster with one another make it complicated to measure the concentration of the trucking industry. Subcontracting is indeed quite widespread: a big carrier's activity may be largely based on partnership with smaller carriers; the trucking industry might thus be less fragmented than usually argued (Fernandez et al., 2006).

[2]Xu et al. (1994) review some of these studies. They explain that measuring returns to scale in the trucking industry is closely related to which inputs and outputs are chosen. In particular, they show that, in the case of LTL (Less Than Truckload) transport, although returns to scale with respect to firm size are constant for given output characteristics such as average load and average shipment size, they are slightly increasing if those are allowed to change: bigger firms transport shipments of such nature that their unit costs decrease.

[3]See for example the recent paper by Smith et al. (2007): the paper's title refers to the rates carriers charge, whereas its very subject is the revenues of the carriers per lane. These revenues are econometrically explained by the lane length, the number of shipments and the average size of the shipments. This means that rates are implicitly assumed proportional to shipment size all along.

[4]Among other attributions, the Interstate Commerce Commission regulated the American trucking industry until the Motor Carrier Act of 1980. In particular, Olson claimed that however strict, this regulation left the carriers enough room for seeming price discrimination to appear.

that the carriers practised so-called value-of-service pricing, *i.e.* higher rates for shippers sending higher density of value[5] commodities, in other words, price discrimination. Her straightforward conclusion was that not only was regulation useful, but that it should have been strengthened to prevent these practises.

The second point is the answer of De Vany and Saving (1977), who argued that what may appear to be price discrimination is basically due to an under-estimation of the complexity of the trucking industry's cost structure: the heterogeneity of prices does not prove price discrimination (*i.e.* several prices for a single product) but product heterogeneity. De Vany and Saving illustrated this argument with the example of vehicle queues that may arise at the origins and destinations of unbalanced freight flows on a given location pair, and demonstrate that asymmetric rates can be consistent with perfect competition on a given origin-destination link[6]: fronthaul and backhaul transport cannot be considered as a unique product.

Although quite complex, the approach of De Vany and Saving has the merit of exhibiting the basic source of rate variability on the road freight transport: the large number of distinct products, and the strong joint production nature of the trucking technology. This, combined with the very large number of outputs of the trucking industry[7], explains the asymmetry of rates when fronthaul and backhaul flows are unbalanced[8], the difficulty to assess the economies of scales at the carrier level, the strong relationships between the various spatial markets, the observed rates variability, and incidentally the need to call on complex concepts such as economies of density and economies of network[9] to address the

[5]The density of value of a given commodity is the monetary value of one weight unit of this commodity, e.g. in € per kg.

[6]They consider a given origin-destination link, with queues forming at both places. If the freight transport demands are not symmetric, queuing times may be longer in the fronthaul direction (the direction with the strongest flow) than in the backhaul direction, which provides a rationale for asymmetric pricing even under perfect competition: a shipper in the fronthaul direction pays the risk that the truck used to transport the freight waits a long time at the destination before finding freight back, whereas the shipper sending freight in the backhaul direction benefits from a premium corresponding to the delay the carrier avoids.

[7]Gillen et al. (1990) argued that, in the case of the airline industry, every city pair and every type of service should be considered as separate, although related, markets. This is definitely true, and applies straightforwardly to the trucking industry, although this may not be accurate enough, as we will see later.

[8]Considering the fronthaul and backhaul transport operations as the outputs in fixed proportions of a return trip made by a truck between two places provides a very simple rationale for asymmetric rates (see e.g. Felton, 1981).

[9]Basso and Jara-Diaz (2005) explain clearly the difference between the aggregates

cost function structure. On the whole, there seems to be no clear evidence that the road freight transport rates depart from the carriers' marginal costs. As it will appear more clearly in the sequel, our study supports this point.

The second argument has been first raised by Baumol and Vinod (1970), who explain the role of transit time in how the shippers choose the transport mode. Their approach is based on a simple underlying logistic model, [10], in which firms look for a trade-off between transport cost and inventory cost when determining how a given commodity flow will be transported from its origin to its destination. In this case, transit time and inventory waiting time, *i.e.* the total delay between the moment the commodity is produced and the moment it is consumed implies a cost to the shipper: a user input. Baumol and Vinod's model yields a choice of both mode and shipment size. In this framework, the shippers' full transport cost (including the inventory cost) is endowed with economies of scale. The fact that the shipment size is explicitly considered plays a crucial role: it is the ground on which the logistic model is based.

To sum up this discussion, it appears first and foremost that shipment size should be taken into account explicitly when considering the structure of trucking costs. Second, from the standpoint of shippers, transport costs include user input costs. This has three consequences: first, the assumption that the transport costs are proportional to the amount of freight transported is at best a quite rough approximation; second, considering explicitly the shippers' inputs clarifies the role of the so-called quality attributes used in some empiric studies. Third, the carriers obviously have a strong incentive to bulk shipments of various sizes as much as the vehicles' capacities allow for, which questions the constant returns to scale hypothesis, even for a given type of transport on a given origin-destination pair. This sets further the analysis of the cost structure of the trucking industry into the frame of joint production. Furthermore, if the transport operations of shipments of various sizes are considered as distinct products, then the carrier's cost function should present economies of scope.

---

usually used in empirical studies (total vehicle kilometres, or passenger kilometres or ton kilometres, and so-called output attributes such as average length of haul and so on) and the true product vector, where all the markets and products are distinguished. They explain the concepts of economies of density and network economies, and present their respective limitations. They introduce the concept of economies of network scope to overcome some of these limits. This concept addresses the effects on costs of an incremental increase in the network served, with reasonable hypotheses on the flows resulting from this increase.

[10] The EOQ model, which will be presented later.

The general purpose of this paper is to re-examine the cost structure of motor carriers and shippers by taking explicitly into account the shipment size, the logistic preferences of shippers, and the technical constraints of carriers. This is a systemic approach: the analysis is not limited to shippers, or carriers, but addresses their interaction. As a consequence, the attention is focused on the freight transport market equilibrium.

More precisely, the first objective is to represent the perfect competition equilibrium on the road freight transport market with distinct shipment sizes. Firstly, some qualitative elements are provided in Section 6.2. Then, the formal framework of this study is introduced in Section 6.3. The second objective is to investigate the specific influence of the vehicle capacity constraint on the equilibrium freight rates. To do so, a simple price schedule model is derived in Section 6.4, where the capacity constraint is naively taken into account; this approach is proved to inadequate. An advanced model is therefore developed in Section 6.5. Despite strong assumptions due to otherwise insuperable technical difficulties, this framework captures in an explicit and consistent way the qualitative elements previously introduced. The third objective is to analyse the implications of this analysis with respect to the cost structure of carriers and of shippers, which is done in Section 6.6, before the study is summed up and concluded in Section 6.7.

6.2 Qualitative elements of description of the road freight transport market

The complexity of the trucking industry sometimes happens to be underestimated. Although this study is not econometric, some qualitative elements about the trucking industry are reminded, so that the models presented later are as realistic as possible, despite their theoretical nature.

The trucking industry encompasses a large number of distinct but interdependent markets, characterised by their origins and destinations, the amounts transported and the transport techniques used. For simplicity, this study is focused on the transport of palletised goods by semi-trailer; a minimal homogeneity of the operations involved is thus insured, since the conditioning and the equipments involved are highly standardised. This also offers a natural, one-dimensional measure of the shipment size and of the vehicle capacity[11].

[11] Even for palletised commodity transport, matters are unfortunately not that sim-

Two markets are usually distinguished in the field of palletised freight transport: first, the classical or industrial transport market: unspecific with respect to the nature of the freight transported, it is characterised by complex price schedules, with a distinct unit price for each shipment size (*i.e.* number of pallets). Shipments are loaded, carried and delivered using the same vehicle. Second, the consolidation market: mainly devoted to the transport of expensive or perishable freight, it is characterised by systematic break-bulk operations[12], and presents paradoxically simpler price schedules. There are indeed usually two, at most three different rate classes. For example, shipments of which the size is comprised between 1 and 15 pallets are charged a unique unit rate plus a fixed term, shipments of more than 15 pallets are charged another unit rate and fixed term, whereas a third price is charged for full truckload shipments.

Each of these markets corresponds to a given transport technique. This study is focused on the industrial transport market. The operations involved in the transport operations shall now be detailed, then some empirical observations on price schedules will be presented.

In the case of the industrial transport market, each shipment's transport can be broken down into an access movement and a haulage movement: shipments are gathered in the origin zone, then carried to the destination zone where they are delivered. The access movement is the movement made by the vehicle to reach the very place the shipment is to be to taken from or delivered. The access movement is assumed to be relatively short with respect to the haulage movement. As a consequence, if the shipment is cancelled, the modification of the vehicle route's length can be assumed marginal. We can thus consider that the access movement is specific to the shipment, whereas the haulage movement is common to all the shipments transported (thus the joint production nature of the trucking technology).

Concerning the rates that the carriers charge, data are hard to obtain. However, some qualitative features of the carriers' price schedules seem to be quite general[13]. Consider a given price schedule $p(.)$ function of s

---

ple. For example, two types of palettes are in circulation in France: the so-called "France" palette, 100×120 cm, and the "Europe" palette, 80×120 cm. A semi-trailer can carry 26 France palettes and 33 Europe palettes. Furthermore, palettes can be stacked if their height is low enough. However, a given commodity type is usually conditioned using a single palette type; trucks thus hardly carry the two palette types. Such refinements are disregarded here.

[12] The shipments, usually quite small, are gathered and brought to a cross-dock platform where they are unloaded, then loaded into other trucks which carry them to their destinations. This is a hub-and-spokes organisational scheme.

[13] This information was made available to the author through informal contacts

the shipment size on a given origin-destination pair. This price schedule has the following properties: first, the price schedules do not tend toward zero when the shipment size tends toward zero:

Property 6.1 $\lim_{s \to 0} p(s) > 0$.

Second, the unit prices are decreasing:

Property 6.2 $s_1 < s_2 \Rightarrow \dfrac{p(s_1)}{s_1} \geq \dfrac{p(s_2)}{s_2}$.

This property ensures, not surprisingly, that $p(.)$ is under-additive[14] (if that property did not hold, it would be profitable for the shippers to split some shipments, which is contradictory). It also implies that considering per ton transport costs and prices is necessarily restrictive.

Third, the price schedule is such that shipments nearly as large as the capacity, denoted S, of the truck are charged so as to bear virtually the whole transport cost:

Property 6.3 $\lim_{s \to S} p'(s) = 0$.

For example, the transport of a shipment using 90% of the vehicle's capacity can be charged 98% of the full truck load rate.

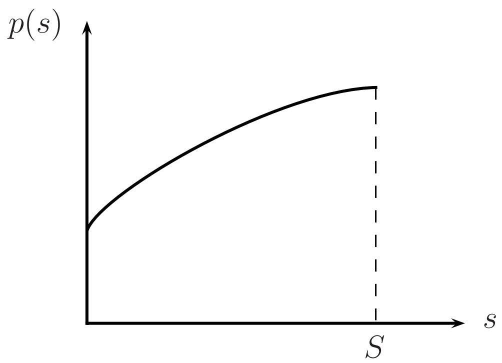

Figure 6.1: Qualitative shape of a carrier's price schedule

The general shape of a price schedule abiding to those three properties is indicatively illustrated by Figure 6.1. The remaining part of this study

with carriers. According to them, carriers do not publish their price schedules. An econometric study would thus certainly require a specific survey at this level to fill in this gap. Besides, some of these elements are more or less corroborated by Abad (2007), among others.

[14]A function f is under-additive if $\forall\ (x; y)$, $f(x + y) \leq f(x) + f(y)$. Note that if $x \mapsto f(x)/x$ is positive and decreasing, then f is under-additive. However, the contrary is not true (Baumol, 1977).

is notably aimed at finding out the minimal set of hypotheses necessary to account for those three properties, in the frame of a perfect competition equilibrium. The framework upon which the following models are built is presented in the next section, on the basis of these qualitative elements.

6.3 General framework

This section provides a simple framework, which will constitute the basis for further analysis. The spatial configuration of the freight transport market considered is presented first in Subsection 6.3.1, then a set of assumptions describing the freight transport supply (motor carriers) is provided in Subsection 6.3.2. Finally, Subsection 6.3.3 provides the representation of the freight transport demand for transport (shippers).

6.3.1 Spatial configuration

It is thereafter assumed that carriers gather a set of shipments in a given origin area, then move towards a destination area where these shipments are delivered. The loading and delivering of each shipment is assumed to require an access movement and loading and unloading operations, whereas the movement of the vehicle from the origin to the destination area with all the shipments on board is referred to as the haulage movement. This representation is illustrated by Figure 6.2:

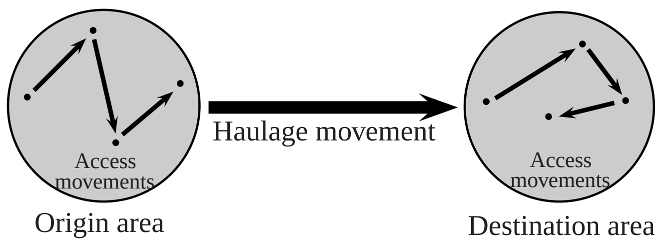

Figure 6.2: Model framework.

Note that the routing problem is set aside: the average cost for carriers of access movements is thereafter considered as a parameter, as well as the cost of haulage movements, as explained in the next subsection.

6.3.2 The freight transport supply

The detailed assumptions describing the technology of carriers and the freight transport market structure are the following ones:

(A1) Carriers are in perfect competition. They are price-takers and they maximise their expected profits.

(A2) Transporting a shipment implies a constant, shipment-specific access cost b, and a joint hauling cost c.

(A3) There is no routing problem. In particular, the backhaul movements, as well as the possibility for the carriers to route their vehicles through complex itineraries, are ignored.

(A4) Trucks are of capacity S.

(A5) The fleets of carriers are fully flexible.

(A6) The market is a three stage game: 1. carriers display their price schedules; 2. shippers choose the carriers who will carry their shipments; 3. shippers decide the sizes of the shipments they want to send, on the basis of the price schedules displayed.

(A7) From the perspective of each carrier, the number of shippers who choose them is a random number. However, the carrier decides this number's expected value.

Assumptions (A1) and (A2) derive straightforwardly from the above discussion. (A3) is not realistic, but it allows for the spatial relationships to be set aside. Backhaul movements and more complex spatial relationships should certainly be taken into account in a comprehensive model of road freight transport, but they can be ignored when addressing the phenomenons this study focuses on. Assuming a unique size of trucks S is tenable, as long as the freight movements considered are on a significantly long distance (e.g. more than 200 km).

(A5) requires some explanation. Sub-contracting and partnerships (either formal or informal) are widespread in the trucking industry[15]. In many cases, relatively large carriers are bounded to shippers by more

[15]For example, the analysis of the French survey on firms (EAE – Enquête Annuelle d'Entreprise) of 1998 reveals that subcontracting made up to 13% of the long-distance road freight transport's turnover at this time in France. This may seem low, but as motor carriers subcontract to motor carriers, the overall proportion of subcontracted turnover is mechanically low (to see that, consider there is only two motor carriers, the first one subcontracting all its activity to the second one: the overall subcontracting proportion would be only 50%).

86% of the subcontracted activity stemmed from 10% of the French motor carriers, mainly big carriers. This tends to point out that big motor carriers subcontract to small carriers. Unfortunately, there is no information on which proportion of the small carriers' activity comes from subcontracting (SESP, 2001).

or less long-term contracts, in which the shippers commit themselves to sending flows of given amounts (the commitment being more or less flexible), while the carriers provide in return guarantees on their rates. These carriers may then resort to other carriers (typically of the same size or smaller) by subcontracting, partnership or spot markets[16] in order to achieve their commitments. Small shippers, or shippers who cannot or do not want to commit themselves on some amount of freight, can also resort to freight spot markets, e.g. through freight exchanges. Economically, all happens as if the carriers had flexible fleets. This is the reason why no fixed cost is assumed in the carriers' cost functions.

The other consequence of the way the trucking industry works is that carriers cannot know for sure the amount of freight they will have to carry. Nevertheless, they have some control on this variable: they can bind up more or less contracts, decide to be present or not on spot markets, etc. But there is a level of uncertainty on the amount of freight which they eventually have to carry. This is what Assumption (A6) stands for.

Assumptions (A6) and (A7) also have a technical role; they may not be fully realistic but they describe unambiguously the relationships of the agents considered in this framework. In particular, Assumption (A6) sets this study in a static framework: only one period is considered.

6.3.3 The freight transport demand

Taking into account the shipment size makes it necessary to explicit how the shippers determine it. This approach, already chosen by Baumol and Vinod (1970) to explain the modal choice made by shippers from the standpoint of their logistic imperatives, relies on the century old Economic Order Quantity model originally designed by Harris (1913), then rigorously derived and popularised by Wilson (1934) (for details, see Erlenkotter, 1989). Hall (1985) applied it with price schedules instead of costs to model the choice of shippers given the exogenous freight transport market prices (the price schedules considered were linear). It

[16] It should be noted that in France, truck drivers who work alone do not depend on the same social regulation as truck drivers employed by motor carriers: in particular, they are not considered as employees, but as directors. As a consequence, they have no minimal wages. This provides a strong rationale for the existence of a number of very small carriers, who, despite the fact that they belong to no organisation, have a clear competitive advantage. Other carriers resort to these small carriers through freight exchanges. These small carriers can thus be considered as a freely available fleet, that the other carriers can use when they need to.

is recalled here, and generalised to the case of any price schedule[17].

When sending commodities somewhere, a shipper has to decide how they will be transported. This implies determining a number of parameters, to be chosen among a large set of options. These options concern for example the speed and cost of the transport operation, the conditioning, the delivery time reliability, etc.

The shipment size is a variable of particular interest: smaller shipments mean lower inventory costs (*i.e.* the time commodities wait to be transported or consumed at the origins and destinations) but larger transport costs (due to the extra vehicle movements needed), whereas larger shipments mean the contrary. As usual, a trade-off exists.

Consider a shippers sending from origin O to destination D a commodity flow at rate ϕ. All other things equal, the shipper wants these goods to be moved at a minimal total cost. This cost encompasses a transport cost, resulting from the road freight market rates, and an inventory cost. Assume the shipper sends shipments of size s. For each shipment, the transport cost is $p(s)$. Meanwhile, the time spent by a unit of good between the moment it is produced and the moment it arrives in B implies an opportunity cost a €/t.day to the shipper. We call a the unit value of waiting time savings.

If the shipments are all of the same size s and if each shipment is sent as soon as possible[18], then the shipper sends ϕ/s shipments per unit of time. The per unit of time generalised transport cost g consists thus of the cost $p(s)$ of sending a shipment of size s times the shipment frequency $1/s$, plus the average origin stock level $s/2$ times the unit value of time

[17]The generalization of the EOQ model to non-linear price schedules is a well-known subject of operational research, as is explained for example in Abad (2007), where a series of papers on this subject are presented. All these works aim at describing accurately the optimal behaviour of a given firm facing a complex environment (non-linear rates, random demand, discounts for large orders, etc.) The objective of our study is to provide the simplest possible shipment size model with non-linear freight rates.

[18]We do not demonstrate that this is, in this particular framework, the optimal shipment schedule, but is should be noted that this depends closely on the hypotheses of the model. In particular, if the inventory cost is not the same at the origin and at the destination, this regularity does not hold: irregular shipment schedules prove to be more efficient (see e.g. Goyal, 1977).

a [19]:

$$g(s) = \frac{a}{2}s + \phi\frac{p(s)}{s}.$$

How does g depend on s? If p is differentiable, then so is g, and:

$$g'(s) = \frac{a}{2} + \phi\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{p(s)}{s}\right).$$

Under adequate hypotheses, this equation has a single solution on $\mathbb{R}_+^*$. In other words:

Proposition 6.4 *If p is a continuous and differentiable function of s, if $p(0) > 0$ and if $p(s)/s$ is decreasing and strictly convex on $\mathbb{R}_+^*$, then there is a unique shipment size s^* minimising the unit transport cost. s^* is uniquely determined by:*

$$-\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{p(s)}{s}\right) = \frac{a}{2\phi}. \tag{6.1}$$

s^* is a decreasing function of a/ϕ.

Proof: $p(s)/s$ is decreasing and strictly convex implies that that the first derivative of $s \mapsto p(s)/s$ is non-positive and increasing towards zero. As a consequence, since a and ϕ are positive, Equation (6.1) has indeed a unique solution[20]. Q.E.D.

The classical EOQ model is a particular case of Proposition 6.4, in which the price schedule is constant (the transport price does not depend on the shipment size). In fact, the EOQ result remains valid as long as the price schedule is linear in s:

[19] Notwithstanding a travel time related waiting cost that is not taken into account: the travel time being constant, the per ton travel time inventory cost is indeed independent from the shipment size, it is thus legitimate to ignore it in the generalised transport cost function. In a model that would address the choice between two or more transport modes, this would not be correct.

[20] Note that if $p(s)/s$ is not strictly convex, Equation (6.1) is still a necessary condition. Furthermore, if $p(s)/s$ strictly convex, then it is continuous and its left and right derivatives exist everywhere. As a consequence, Proposition 6.4 still holds, provided that Equation (6.1) is replaced by:

$$-\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{p(s)}{s}\right)\Big|_{s=s^{*+}} \leq \frac{b}{2\phi} \leq -\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{p(s)}{s}\right)\Big|_{s=s^{*-}}$$

Corollary 6.5 *When $p(s) = b + cs$ with $b, c > 0$, the optimal shipment size is:*

$$s^* = \sqrt{\frac{2\phi b}{a}} \tag{6.2}$$

Proof: $\mathrm{d}/\mathrm{d}s(p(s)/s) = -b/s^2$, thus the result. Q.E.D.

This is the form of the price schedules used in Hall (1985). Observe that c has no influence on the shipment size choice. Observe also that this corollary does not hold when $b = 0$. In that case, the transport price is proportional to the size of the shipment: as a consequence it is optimal for the shippers to choose an arbitrary small shipment size, so as to suppress the inventory costs.

Finally, when the hypotheses of Proposition 6.4 do not hold, there is no general solution; each case calls for an adequate method.

It follows from this discussion that any shipper's behaviour derives from the rate ϕ of the emitted commodity flow and from the unit value of time a. As a consequence, in the following of the paper, the shippers' population is described by the joint distribution of ϕ and a, denoted $F_{\phi,a}$.

6.4 Basic model: equilibrium freight rates with a constant loading factor

Due to market competition, each carrier needs to minimise its costs. This means that to transport a given set of shipments, each carrier needs to use as few vehicles as possible, or, in other words, to maximise the loading factor of its fleet.

This maximisation is a twofold problem. First, the carrier must be able to determine the minimal number of vehicles necessary to carry any set of shipments. This is a classic operations research problem known as the bin-packing problem[21]. It is known as a NP-hard problem. In the following, it is assumed the carriers are always able to find the optimal solution.

Second, the carrier may influence the sizes of the shipments it has to transport by means of its freight rate. Some shipment sizes are more or less desirable to the carriers. For example, a full truck load means

[21] In this case, a 1D bin-packing problem, as the shipment sizes are expressed in number of pallets, as well as the vehicle capacity; the 2D and 3D bin-packing problems, concerning cases where the disposition in a 2D or 3D space of shipments has to be taken explicitly into account, are much more complicated.

a full vehicle, which is a good thing. On the contrary, a big shipment can be problematic: either the carrier is able to find a small shipment to use the remaining payload, or the vehicle will run partially empty. As a consequence, a small shipment may be the opportunity for a carrier to optimise its fleet's loading factor. Carriers might thus want to favour some shipment sizes and to penalize others, so that shippers modify their choices in a way that decreases the costs of carriers.

The objective of this section is to determine whether it is necessary to take this phenomenon into account to model the equilibrium freight rates. To do so, it is first assumed this is not the case:

Hypothesis 6.6 *All the carriers consider that whatever the rates they charge, the average loading factor of their fleet is constant.*

In other words, carriers assume that the number of vehicles they need to carry a given set of shipments is proportional to the number of pallets to be carried. Furthermore, they assume that this number of vehicles does not depend on the sizes of shipments. Everything happens here as if the average loading factor were a kind of constant of the trucking industry, resulting from long observation and practise; denote λ this loading factor.

In this section, the freight transport market equilibrium price schedule is derived under the assumptions introduced in Subsection 6.3.2 and under Hypothesis 6.6.

To derive the equilibrium price schedule, let us assume that all the carriers charge the same following price schedule:

$$p(s) = b + K.s,$$

where $p(s)$ is the rate charged for the transport of a shipment of size s, b the access cost and K a positive constant. Then it results from Corollary 6.5 that the sizes of the shipments sent by the shippers do not depend on K.

Let us now calculate the profits of carriers for this price schedule and loading factor. Consider one carrier. Denote $\{s_i\}_{i=1...n}$ the sizes of the shipments it has to carry, with n the number of shipments. Denote n_t the minimal number of vehicles necessary to carry them. The fleet's observed average loading factor Λ is by definition:

$$\Lambda = \frac{\sum_{i=1}^{n} s_i}{S.n_t}. \tag{6.3}$$

As a consequence, the profit π_t of the carrier is:

$$\pi_t = \sum_{i=1}^{n} (b + K.s_i) - n.b - n_t.c,$$

or, using Equation (6.3):

$$\pi_t = \left(K - \frac{c}{S.\Lambda}\right) \cdot \left(\sum_{i=1}^{n} s_i\right).$$

Due to Hypothesis 6.6, the carrier expects the average load factor to be equal to:

$$\Lambda = \lambda,$$

so that its profit is:

$$\pi_t = \left(K - \frac{c}{S.\lambda}\right) \cdot \left(\sum_{i=1}^{n} s_i\right).$$

This implies that the road freight transport market is at the perfect competition equilibrium when $K = c/(S.\lambda)$, that is to say when the price schedule is:

$$p(s) = b + \frac{c.s}{S.\lambda}. \tag{6.4}$$

When all the carriers charge this price schedule, their expected profit is zero. They can stay in the market, but none can undercut or surpass the price schedule.

Furthermore, this equilibrium is unique. Assume indeed another equilibrium price schedule $p_2(.)$. There is at least one shipment size s' for which $p_2(s') < p(s')$. Any carrier will expect to make a positive profit by charging $p_2(s)$ for all $s \neq s'$ and $p(s')$ for $s = s'$ (or any other value larger than $p_2(s')$). As a consequence:

Proposition 6.7 *Under Hypothesis 6.6, the perfect competition equilbrium price schedule on the road freight transport market is:*

$$p(s) = b + \frac{c.s}{S.\lambda}.$$

Figure 6.3 illustrates this equilibrium price schedule. Two of its features are worth noting. First, the equilibrium price schedule is linear, but does not tend toward zero when the shipment size gets small, due to the access costs: however small a shipment may be, the necessity for the carrier to reach the origin and destination places to load and deliver the shipment incurs costs that do not depend on the shipment size, and in particular which do not tend towards zero for very small shipments.

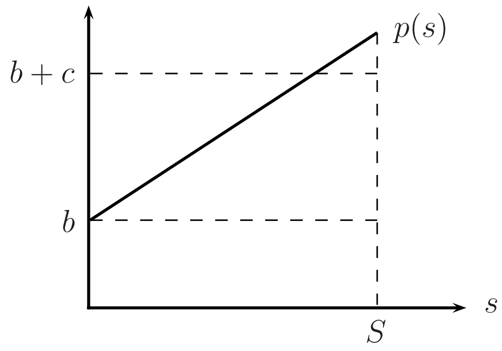

Figure 6.3: equilibrium price schedule under Hypothesis 6.6.

This is the basic cause of the famous "last-kilometer" problematic. Besides, this implies that assuming the transport costs are proportional to the amount of freight shipped is incorrect. This also proves that this simple model is able to account for Properties 6.1 and 6.2 (the unit price is clearly decreasing), but not for Property 6.3.

Second, it stems from the assumptions previously introduced that the cost for a carrier to transport an FTL shipment is $b + c$. However, the equilibrium rate for an FTL shipment is strictly larger than $b + c$ (the price schedule crosses the upper dotted line in Figure 6.3). This is a direct consequence of Hypothesis 6.6: all the carriers expect an average loading factor λ, even a carrier which would have to carry only FTL shipments. This is clearly contradictory. Hypothesis 6.6 is not tenable. This hypothesis is relaxed in the next section.

6.5 Advanced model: freight rates with an endogenous loading factor

The equilibrium freight rate derived in the previous model is not realistic: any carrier would make a positive profit by specialising in FTL shipments and charging any price comprised between $b + c$ and $p^e(S)$. With an average loading factor of 1, such a carrier would indeed at least cover its costs. It is realistic to assume the carriers are aware of this fact; however, this statement is inconsistent with Hypothesis 6.6.

As a consequence, Hypothesis 6.6 is relaxed thereafter. It it assumed in this section that carriers consider the average loading factors of their fleets depend on the sizes of the shipments they have to carry. The objective is to derive the equilibrium price schedules.

Nevertheless, there is a significant technical difficulty in such an approach: to investigate how the average loading factor of a carrier's fleet depends on the sizes of the shipments carried, the bin-packing problem must be considered explicitly. The model thus becomes very complicated, lest the framework is kept extremely simple. As a consequence, the scope of this study is limited to a stylised model of which the derivation will hopefully yield useful insights, despite its sketchy character.

The simplifications made so that the model is warkable are detailed in Subsection 6.5.1. The equations characterising the supply-side perfect competition equilibrium prices are then derived in Subsection 6.5.2, and for the demand in Subsection 6.5.3. We finally derive the market equilibrium in Subsection 6.5.4.

6.5.1 Framework simplification

Relaxing Hypothesis 6.6 means that the loading factor is not considered as a constant by the carriers. On the contrary, each carrier may influence the number and size of the shipments to be carried by modifying the rates it charges. To take this phenomenon into account, it is necessary to determine how the loading factor depends on the sizes of the shipments to be carried.

Due to the structure of the bin-packing problem, this is impossible in the general case. To overcome this obstacle, the following simplification is set forth: from now on, it is assumed that the shippers may choose a shipment size of 1, 2 or 3 units, and that the vehicles' capacity is 3. In this framework, the bin-packing problem has a simple, explicit solution: the vehicles can be filled with three shipments of size 1, or one shipment of size 1 and one shipment of size 2, or one shipment of size 3.

There are n shippers. For simplicity, each shipper sends an amount $\phi = 1$ of commodities per unit of time. The shipper population is therefore fully described by n and F_a, the unit value of time distribution. The density of F_a is denoted f_a.

Denote n_i the number of shippers sending shipments of size i. Necessarily, $\sum_i n_i = n$.

Denote n_i^s the number of shipments of size i sent per period of time. Due to the flow $\phi = 1$ sent by each shipper per period of time, these numbers verify $n_i^s = n_i/i$. The following analysis could be extended to any joint distribution $F_{\phi,a}$ without major difficulties, but the calculations would be tedious, without bringing anything crucial to the study.

The approach is focused on one period. From Subsection 6.3.2, the events sequence is: first, the carriers determine the rates they charge so as to maximise their expected profit; then each shipper chooses a carrier on the basis of these rates, finally they choose the size of the shipments they send, on the basis of the logistic behaviour introduced in Subsection 6.3.3.

Now should be determined how the carriers anticipate their costs and profits. Consider a given carrier. Denote Q_i the number of shipments of size i this carrier has to transport. Assumption (A7) states that this number is random. Q_i is thus a random variable, of which the expected value is denoted q_i. From Assumption (A7), provided it is competitive, the carrier decides q_i, but not Q_i. More precisely, the following assumption[22] is made with respect to the distribution of Q_i:

Hypothesis 6.8 *Q_i is uniformly distributed, of mean q_i and standard deviation proportional to $(q_i)^2$:*

$$Q_i \hookrightarrow \mathcal{U}\left(\left(1-\frac{K_s}{2}\right)q_i, \left(1+\frac{K_s}{2}\right)q_i\right). \tag{6.5}$$

with $K_s \leq 2$ a positive constant.

In other words, the carriers assume the relative variability of the demand they face does not depend on q_i: whatever q_i, the probability that Q_i is comprised between e.g. $q_i \pm 10\%$ is constant – there is no risk pooling effect.

The market equilibrium is addressed in three steps. First, the carriers' profit constraint is stated, thus deriving conditions on the equilibrium price schedule for given $\{n_i^s\}_i$. Second, the analysis jumps back to the behaviour of the demand, before the market equilibrium is eventually derived.

[22] This assumption is clearly *ad hoc*. As it will appear later, its main merits are first to ensure constant returns of the carriers' cost function, allowing the analysis to focus on the scope effects and to disregard the scale effects (using the vocabulary of Baumol, 1977), second to allow an analytical approach, impossible if Q_i's distribution is not explicited.

Is it a realistic assumption? This is quite hard to say from a theoretical point of view. The following reasoning, based on simple assumptions, yields different properties.

Assume the shippers choose randomly the carriers for each shipment, with the probability for one carrier to be chosen being proportional to its activity q_i. Q_i's distribution is then a multinomial $\mathcal{B}(n,p)$ with $n = n_i^s$ and $p = q_i/n_i^s$, so that its mean and variance are respectively $n.p = q_i$ and $n.p.(1-p) = q_i.(1-q_i/n_i^s)$. The standard deviation of Q_i thus increases more slowly than q_i.

However, shippers sending a lot of shipments do not choose a carrier for each shipment independently. Some shipment lots are thus allocated simultaneously, which would tend to increase the variance of Q_i.

6.5.2 Profit constraints

The carriers' profits should be non-negative for any transport demands. They must determine the $\{q_i\}_i$ and the price schedule $p(\cdot)$ so as to reach this objective.

The amounts of shipments of each size the carrier has to transport are $\{Q_1; Q_2; Q_3\}$. Denote C the carrier's cost function. Then its profit is:

$$\pi(Q_1, Q_2, Q_3) = \sum_{i=1}^{3} p(i).Q_i - C(Q_1, Q_2, Q_3).$$

Note that C, and as a consequence π are separable in $\{Q_1; Q_2\}$ on one hand and Q_3 on the other hand. π can re rewritten as as:

$$\pi(Q_1, Q_2, Q_3) = \left[\sum_{i=1}^{2} p(i).Q_i - C(Q_1, Q_2)\right] + \Big[p(3).Q_3 - C(Q_3)\Big],$$

or, equivalently:

$$\pi(Q_1, Q_2, Q_3) = \pi(Q_1, Q_2) + \pi(Q_3)$$

where we use the same symbols π and C to represent the three functions $\pi(\cdot)$, $\pi(\cdot, \cdot)$ and $\pi(\cdot, \cdot, \cdot)$ and $C(\cdot)$, $C(\cdot, \cdot)$ and $C(\cdot, \cdot, \cdot)$ respectively.

This has two consequences. First, the perfect competition equilibrium price $p(3)$ can be derived independently of $p(1)$ and $p(2)$. To determine it, it is sufficient to know $C(Q_3)$. But a shipment of size 3 is a FTL shipment, which means that one vehicle and one access movement are needed per vehicle:

$$C(Q_3) = (b + c)Q_3.$$

As a consequence the marginal cost of carrying a shipment of size 3 is $b + c$, and so is $p(3)$ at the perfect competition market equilibrium:

$$p(3) = b + c. \tag{6.6}$$

Second, the equilibrium prices $p(1)$ and $p(2)$ are interdependent. Indeed, if the carrier has to transport Q_1 shipments of size 1 and Q_2 shipments of size 2, its revenue is $p(1)Q_1 + p(2)Q_2$. Its cost depends on the sign of $Q_1 - Q_2$. If $Q_1 \geq Q_2$, Q_2 vehicles will carry Q_2 shipments of size 1 and Q_2 shipments of size 2; and $(Q_1 - Q_2)/3$ vehicles will carry $3Q_1$ shipments of size 1. On the contrary, if $Q_1 < Q_2$, Q_2 vehicles will be necessary to move all the shipments.

As a consequence, the transport cost is:

$$C(Q_1, Q_2) = b(Q_1 + Q_2) + \frac{c}{3}\Delta + cQ_2,$$

where Δ is defined as follows:

$$\Delta = (Q_1 - Q_2)^+.$$

Subsequently, Δ is also a random variable. The first term of C stands for the access costs, and the last two for the long haul costs.

Δ is a variable of considerable importance in this analysis: it means that, due to Assumption (A7), even if there are, on the whole, more shipments of size 1 than shipments of size 2 (*i.e.* $n_1^s > n_2^s$), a carrier may still have to transport more shipments of size 2 than shipments of size 1, and as a consequence, some of its vehicles would be partially empty. The carriers take this eventuality into account.

The profit of the carrier for the LTL shipments is:

$$\pi(Q_1; Q_2) = (p(1) - b)Q_1 + (p(2) - b - c)Q_2 - \frac{c}{3}\Delta.$$

It is a random variable as well. Its expected value is a function of q_1 and q_2:

$$\mathbb{E}(\pi)(q_1; q_2) = (p(1) - b)q_1 + (p(2) - b - c)q_2 - \frac{c}{3}\delta(q_1, q_2),$$

where δ, defined as:

$$\delta(q_1, q_2) = \mathbb{E}(\Delta),$$

is also function of q_1 and q_2. Note that δ is positively homogeneous of degree 1 (for all $\nu > 0$, $\delta(\nu q_1, \nu q_2) = \nu\delta(q_1, q_2)$; this result depends crucially from Hypothesis 6.8), which implies that C presents constant returns to scale.

At the perfect competition equilibrium, the carriers are price-taker. They choose, for given prices $p(1)$ and $p(2)$, the quantities q_1 and q_2 they produce so as to optimise their expected profits. In other words, the following first-order conditions hold:

$$\frac{\partial\mathbb{E}(\pi)}{\partial q_1} = \frac{\partial\mathbb{E}(\pi)}{\partial q_2} = 0.$$

Or, equivalently:

$$p(1) = b + \frac{c}{3}\frac{\partial\delta}{\partial q_1}(q_1, q_2), \tag{6.7}$$

and:

$$p(2) = b + c + \frac{c}{3}\frac{\partial\delta}{\partial q_2}(q_1, q_2). \tag{6.8}$$

Equations (6.7) and (6.8) determine, for each carrier, the optimal q_1 and q_2. But, since δ is homogeneous of degree one, its partial derivatives are homogeneous of degree zero, that is to say: $\partial\delta/\partial q_i(q_1, q_2) = \partial\delta/\partial q_i(q_1/q_2, 1) = \partial\delta/\partial q_i(q_1/q_2)$.

Therefore, for given prices, q_1 and q_2 are not determined uniquely: Equations (6.7) and (6.8) only yield the ratio q_1/q_2 a carrier should opt for in order to maximise its own profit. This result is consistent with the fact that C presents constant returns to scale. Denote r this ratio.

Quite straightforwardly, if the quantities (q_1, q_2) chosen by each carrier share the same ratio r, and if the supplies and demands are balanced, necessarily $r = n_1^s/n_2^s$. This determines the perfect competition equilibrium prices for a given transport demand:

Proposition 6.9 *For the freight transport market to be at the perfect competition equilibrium with given freight transport demands n_1^s and n_2^s, the equilibrium price schedule necessarily verifies:*

$$\left\{\begin{array}{lcl} p(1) & = & b + \frac{\partial\delta}{\partial q_1}(n_1^s, n_2^s)c/3 \\ p(2) & = & b + c + \frac{\partial\delta}{\partial q_2}(n_1^s, n_2^s)c/3 \\ p(3) & = & b + c, \end{array}\right. \tag{6.9}$$

The behaviour of these prices is clarified by the following result:

Lemma 6.10 *When n_1^s/n_2^s increases from 0 to $+\infty$, $\partial\delta/\partial q_1$ increases from 0 to 1 and $\partial\delta/\partial q_2$ decreases from 0 to -1.*

Proof: see Appendix C.1.

The evolution of the perfect competition prices for given transport demand is illustrated by Figure 6.4, for three values of K_s, with b and c fixed at the values of 0.5 and 2 respectively, and where the access cost b and full transport cost $b + c$ have been represented by two dashed black horizontal lines.

Figure 6.5 illustrates the price schedules[23] for various ratios n_1^s/n_2^s, with the following parameters $b = 0.5$, $c = 2$ and $K_s = 0.4$. As previously two horizontal dashed lines have been drawn at levels b and $b + c$.

[23] Note that the under-additivity property does not derive automatically from Proposition 6.9, although obviously, if $2p(1) < p(2)$, no shipper will decide to send shipments of size 2. However if $n_i^s = 0$ for any i, the corresponding price is irrelevant; the proposition remains thus valid.

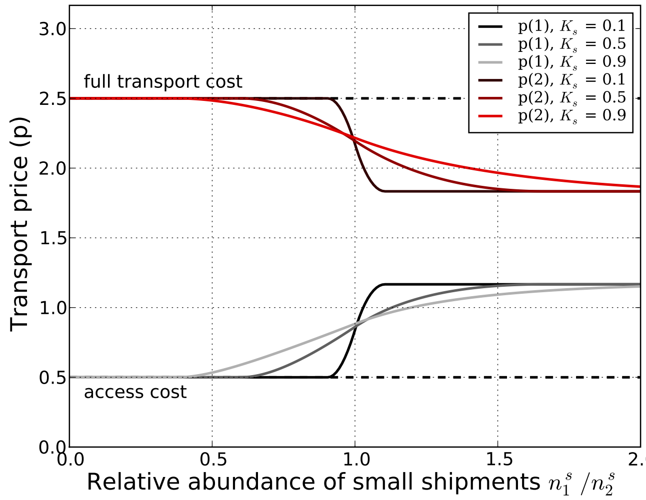

Figure 6.4: Equibrium prices for a given transport demand

From Proposition 6.9, as well as from Lemma 6.10, the equilibrium prices have the following properties. First, each shipment bears the cost b of its access movement. Second, the assignment of the haulage cost depends on the relative frequencies of shipments of size 1 and 2. If there are many more shipments of size 1 than shipments of size 2, the vehicles run full. For each carrier, an additional shipment of size 1 means an extra third of vehicle, and an additional shipment of size 2 means two thirds of a vehicle which will be easily filled by one of the several shipments of size 1 to be carried. Each shipment thus bears its share of the haulage cost, *i.e.* $c/3$ for a shipment of size 1 and $2c/3$ for a shipment of size 2. This situation corresponds to the right side of Figure 6.4, and to the light grey curves in Figure 6.5.

On the contrary, if n_1^s is low compared to n_2^s, things are radically different. The carriers expect to have a significant number of vehicles running partially empty. An additional shipment of size 1 means an opportunity to fill one of these vehicles; in other words, there is no need to use another vehicle to transport them, so that the marginal cost of carrying this shipment is the access cost. An additional shipment of size 2 means the opposite thing: another vehicle will be needed only for this shipment. In that case, the marginal cost is $b + c$. This corresponds to the left side of Figure 6.4, and to the darker curves in Figure 6.5.

Although the framework of this study is simplistic, the results derived are endowed with some interesting features. First, the joint production

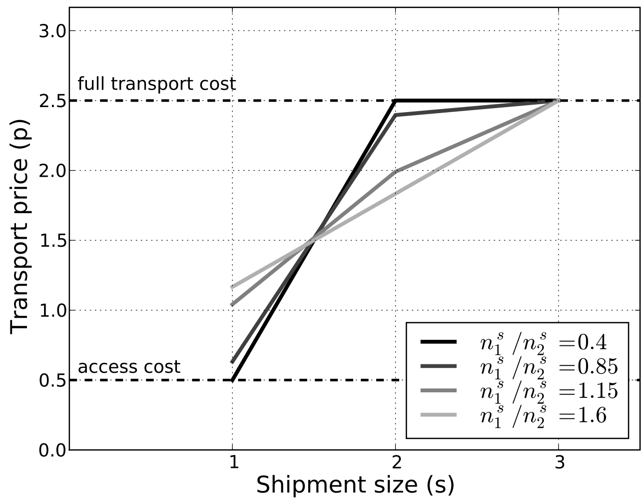

Figure 6.5: Price schedule for a given, fixed demand

nature of the trucking technology is clearly visible: the price of carrying shipments of different sizes are, in some cases, closely related. Second, the results of the model are consistent with the qualitative elements introduced in Section 6.2, in particular with Property 6.3. This proves the pivotal role of the constraint capacity in the formation of freight rates, and the consistency between perfect competition and complex price schedules, as discussed in the introduction.

The next step of this analysis is to derive the demand behaviour and the market equilibrium. However, before doing so, the average loading factor of the carrier's fleets will be calculated.

For each carrier, the fleet's loading factor is a random variable, depending on the number of shipments of each size it has to carry. Since any truck holding a shipment of size 3 is full (and consequently its loading factor is 1), the focus is here limited to the loading factor of trucks moving shipments of size 1 and 2, *i.e.* less-than-truckload (LTL) shipments. Denote Λ_{LTL} this loading factor. The loading factor of a carrier which has to transport Q_1 (resp. Q_2) shipments of size 1 (resp. 2) is:

$$\Lambda_{LTL} = \begin{cases} 1 & \text{if} \quad Q_1 \geq Q_2 \\ \frac{Q_1+2Q_2}{3Q_2} & \text{else} \end{cases} \tag{6.10}$$

Denote λ_{LTL} the expected value of Λ_{LTL}. It is possible but tedious to derive an exact formula of λ_{LTL}. The calculations have thus been relegated to Appendix C.2. Figure 6.6 depicts the evolution of the expected loading factor with n_1^s/n_2^s and K_s.

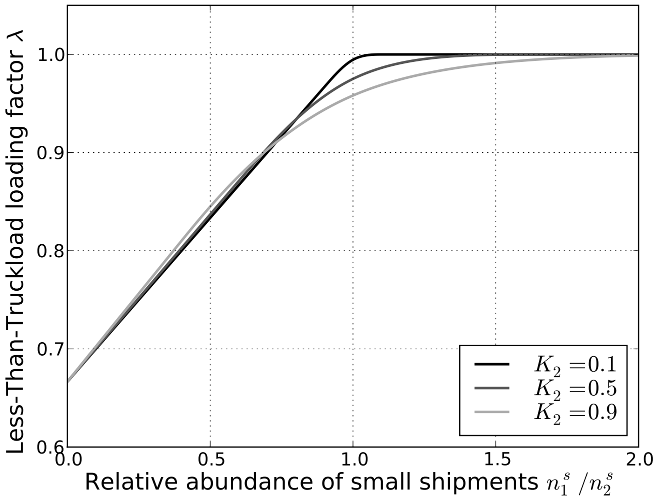

Figure 6.6: LTL trucks expected loading factor

This illustration calls for two observations. First, the average loading factor is theoretically lower than $2/3 + n_1^s/3n_2^s$. This is approximately what is observed in Figure 6.6, but not exactly. Indeed, for relatively small values of n_1^s/n_2^s (say, $n_1^s/n_2^s \leq 0.75$), the expected loading factor is actually higher than this upper bound, and increases slightly with K_s. These two rather paradoxical properties actually show a quite unexpected limitation of Hypothesis 6.8: it is indeed impossible to have simultaneously a fixed total transport demand at the industry level and carriers facing independent, stochastic demands of which the variances increase proportionally to the expected values. As a matter of fact, Hypothesis 6.8 would be consistent with random industry transport demands of variance proportional to n_1^s and n_2^s respectively; in that case, the $2/3 + n_1^s/3n_2^s$ upper born would not hold anymore. Anyway, the inconsistency is not dramatic; the framework is thus kept as is thereafter.

The second observation is that the LTL loading factor generally decreases when K_s increases. This point will be further discussed in Subsection 6.6.4.

6.5.3 Transport demand

Now that the behaviour of the carriers has been explicited, and that the equilibrium prices have been derived for given transport demands, the focus is brought on the demand side of the freight transport market. In other words, n_1^s, n_2^s and n_3^s are derived as functions of a given price schedule $p(\cdot)$.

Consider a shipper sending an amount $\phi = 1$ of commodity per unit of time, with an opportunity cost of travel time a. This shipper can send shipments of size 1, 2 or 3. The generalised transport cost of each of these alternatives are respectively:

$$\begin{cases} g(1) &= a/2 + p(1) \\ g(2) &= a + p(2)/2 \\ g(3) &= 3a/2 + p(3)/3 \end{cases}$$

Each firm chooses the most economic shipment size, which depends on their unit value of time a. A shipment size of 1 is advantageous when $g(1) \leq g(2)$ and $g(1) \leq g(3)$, that is to say:

$$\begin{cases} a/2 + p(1) &\leq a + p(2)/2 \\ a/2 + p(1) &\leq 3a/2 + p(3)/3 \end{cases}$$

A shipment of size 2 is better than a shipment of size 3 if $g(2) \leq g(3)$, *i.e.*:

$$a + p(2)/2 \leq 3a/2 + p(3)/3$$

The optimal shipment size is then straightforwardly determined.

Proposition 6.11 *For a shipper of unit value of time a shipping $\phi = 1$ units of a given commodity per year, facing a price schedule $p(.)$, and if $p(2) \leq p(1) + p(3)/3$, the optimal shipment size s^* is:*

$$\begin{cases} s^* = 3 & \text{for} \quad a \in [\,0\,;\, p(2) - 2p(3)/3\,[\\ s^* = 2 & \text{for} \quad a \in [\,p(2) - 2p(3)/3\,;\, 2p(1) - p(2)\,[\\ s^* = 1 & \text{for} \quad a \in [\,2p(1) - p(2)\,;\, +\infty\,[; \end{cases}$$

in the case where $p(2) > p(1) + p(3)/3$, the optimal shipment size is:

$$\begin{cases} s^* = 3 & \text{for} \quad a \in [\,0\,;\, p(1) - p(3)/3\,[\\ s^* = 1 & \text{for} \quad a \in [\,p(1) - p(3)/3\,;\, +\infty\,[. \end{cases}$$

Figure 6.7 shows how the cost of each alternative depends on a, in the case where a shipment size of 2 can be optimal (*i.e.* when $p(2) \leq p(1) + p(3)/3$, the first case in Proposition 6.11). The parameters for this figure are $p(1) = 0.8$, $p(2) = 1.2$ and $p(3) = 1.5$. $g(1), g(2)$ and $g(3)$ have been drawn on the whole range covered by a, their ordinates are carried over to the left vertical axis. The minimum envelope is represented by a wider black piecewise linear curve. On the right vertical axis is carried over the optimal shipment size, drawn in red on the figure.

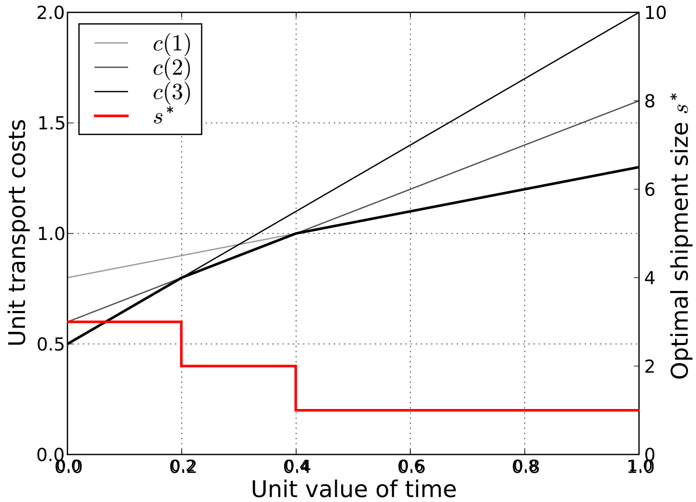

Figure 6.7: Optimal shipment size with respect to a

One can indeed observe that if $p(2)/2$ is too high, all firms choose either a shipment size of 1 or 3. This is the case illustrated by Figure 6.8, drawn with the same parameters as the previous one, apart from $p(2) = 1.4$.

Two results are confirmed by Proposition 6.11. First, higher values of time imply lower shipment sizes: the higher a, the more expensive the inventory cost relatively to the transport cost, thus the result. Second, a small increase in the freight rate for the intermediate shipment size leads to a fast decrease in the number of shipments of this size to be carried.

Proposition 6.11 makes it possible to calculate the demands for transport n_i^s for a given price schedule $p(.)$. Note that for a flow $\phi = 1$, a firm for which $s^* = j$ with $j \in \{1; 2; 3\}$ sends $1/j$ shipments per unit of time. As a consequence, with N the number of firms, f_a the density of distribution of the firm's unit values of time, and F_a the corresponding c.d.f.:

Proposition 6.12 *For a given price schedule $p(.)$, if $p(2) \leq p(1) + p(3)/3$, the demands for transport of each shipment size are:*

$$\begin{cases} n_1^s(p) = N.\Big(1 - F_a\big(2p(1) - p(2)\big)\Big) \\ n_2^s(p) = \frac{N}{2}.\Big(F_a\big(2p(1) - p(2)\big) - F_a\big(p(2) - \frac{2}{3}p(3)\big)\Big) \\ n_3^s(p) = \frac{N}{3}.F_a\left(p(2) - \frac{2}{3}p(3)\right) \end{cases} \tag{6.11}$$

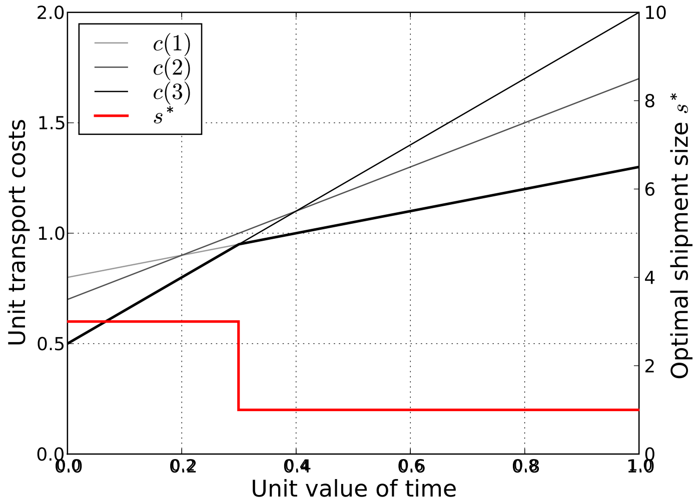

Figure 6.8: Optimal shipment size, where $s = 2$ is never optimal.

In the case where $p(2) > p(1) + p(3)/3$*:*

$$\begin{cases} n_1^s(p) = N.\Big(1 - F_a\left(p(1) - \frac{1}{3}p(3)\right)\Big) \\ n_2^s(p) = 0 \\ n_3^s(p) = \frac{N}{3}.F_a\left(p(1) - \frac{1}{3}p(3)\right) \end{cases} \tag{6.12}$$

Straightforwardly, the demands for transport of shipments of size i depend therefore as follows on the transport price schedule:

Corollary 6.13 *For all i in $\{1;2;3\}$, $n_i^s(p)$ is decreasing in p_i and increasing in p_j with $j \neq i$.*

The transport of shipments of different sizes are distinct services; it follows from Corollary 6.13 that they are substitutes. Once determined the behaviour of the demand for a given price schedule, we may analyse the potential existence and unicity of a perfect competition market equilibrium.

6.5.4 Market equilibrium

The transport market is at the perfect competition equilibrium if the demands and supplies of all the products considered are balanced, and if Propositions 6.9 and 6.11 hold. The existence and unicity of such an equilibrium relies mainly on Lemma 6.10 and Corollary 6.13. It is indeed sufficient to show that equation systems (6.9), (6.11) and (6.12) have a unique solution in $p(.)$ and $\{n_i^s\}_i$.

The idea behind this result is quite simple: due to the technical constraints of the carriers, the price of sending shipments of size 1 increases relatively to the price of sending shipments of size 2 when shipments of size 1 are relatively more frequent. This tends to decrease the demand for transport of shipments of size 1 with respect to the demand for transport of shipment of size 2, these two services being substitutes for one another.

More precisely, note first that from Proposition 6.9, we have $p(3) = b + c$. The problem is to find out whether equation systems (6.9), (6.11) and (6.12) determine uniquely $p(1)$, $p(2)$, n_1^s, n_2^s and n_3^s, or not.

To see that this is the case, let $r = n_1^s/n_2^s$. Given that $\partial\delta/\partial n_i^s$ is function of r for $i = 1, 2$, Proposition 6.9 and Lemma 6.10 imply that $p(1)$ (resp. $p(2)$) is an increasing (resp. decreasing) function of r. Let $f_{p(1)}$ and $f_{p(2)}$ express $p(1)$ and $p(2)$ as functions of r:

$$\begin{cases} p(1) &= f_{p(1)}(r) \\ p(2) &= f_{p(2)}(r) \end{cases}$$

These two functions are monotonic: $f_{p(1)}$ is increasing from b to $b + c/3$ on $\mathbb{R}_+$ whereas $f_{p(2)}$ is decreasing from $b + c$ to $b + 2c/3$ on the same range.

At the same time, from Proposition 6.12, n_1^s and n_2^s depend univoquely on $p(.)$. Equation systems (6.11) and (6.12) imply that n_1^s and n_2^s can be written as functions of $p(1)$ and $p(2)$. Denote respectively $f_{n_1^s}$ and $f_{n_2^s}$ these two functions:

$$\begin{cases} n_1^s &= f_{n_1^s}(p(1), p(2)), \\ n_2^s &= f_{n_2^s}(p(1), p(2)). \end{cases}$$

From Corollary 6.13, $f_{n_1^s}$ is a decreasing function of $p(1)$, and an increasing function of $p(2)$, and $f_{n_2^s}$ is an increasing function of $p(1)$ and a decreasing function of $p(2)$. n_1^s and n_2^s can be written as functions of r, by replacing $p(1)$ and $p(2)$:

$$\begin{cases} n_1^s &= f_{n_1^s}(f_{p(1)}(r), f_{p(2)}(r)), \\ n_2^s &= f_{n_2^s}(f_{p(1)}(r), f_{p(2)}(r)). \end{cases}$$

The case where $n_2^s = 0$ should be examined with special care. It derives from the equation above that n_2^s is an decreasing function of r, from $f_{n_2^s}(b + c/3, b + 2c/3)$ to $f_{n_2^s}(b, b + c)$ on $\mathbb{R}_+$. If $f_{n_2^s}(b, b + c)$ is zero, then n_2^s is zero for all r in $\mathbb{R}_+$. The competitive equilibrium is then unique and described by $n_2^s = 0$, $p(1) = b + c/3$, $p(2) = b + 2c/3$.

If $f_{n_2^s}(b, b+c) > 0$, then there is a constant $K \in \mathbb{R}_+$ such that $\forall r > K$, $f_{n_2^s}(f_{p(1)}(r), f_{p(2)}(r)) > 0$.

As a consequence, r can be written on $]K; +\infty[$ as a function of $p(1)$ and $p(2)$:

$$r = \frac{f_{n_1^s}(p(1), p(2))}{f_{n_2^s}(p(1), p(2))},$$

where the right-hand side is decreasing in $p(1)$ and increasing in $p(2)$. We obtain by replacing $p(1)$ and $p(2)$ that $r = f_r(r)$ with:

$$f_r : r \mapsto \frac{f_{n_1^s}(f_{p(1)}(r), f_{p(2)}(r))}{f_{n_2^s}(f_{p(1)}(r), f_{p(2)}(r))} \tag{6.13}$$

f_r is decreasing from $f_{n_1^s}(b, b+c)/f_{n_2^s}(b, b+c)$ to $f_{n_1^s}(b+c/3, b+2c/3)/f_{n_2^s}(b+c/3, b+2c/3)$ on $\mathbb{R}_+$ if $f_{n_2^s}(b, b+c) > 0$ and from $+\infty$ to $f_{n_1^s}(b+c/3, b+2c/3)/f_{n_2^s}(b+c/3, b+2c/3)$ on $]K; +\infty[$ else. In both cases, $r = f_r(r)$ has a unique positive solution.

It results from this discussion that if the freight transport market is at the perfect competition equilibrium, either $n_2^s(r) = 0$ for all r, or $r = f_r(r)$. Conversely, each of these two cases, which are mutually exclusive, describe such an equilibrium. $p(1)$ and $p(2)$ are straightforwardly and uniquely determined by equation system (6.9); n_1^s, n_2^s and n_3^s derive from $p(1)$ and $p(2)$ using equation systems (6.11) and (6.12).

As a consequence:

Proposition 6.14 *For any unit value of time distribution F_a, there is a unique perfect competition equilibrium on the freight transport market. This equilibrium is described by the unique solution r^* of the following equation:*

$$r^* = f_r(r^*) \tag{6.14}$$

This result ensures that the framework proposed in this paper behaves correctly. This model is thus able, despite the limited realism of the assumptions it is based on, to propose an explanation for the empiric facts presented in Section 6.2. This model indeed illustrates how the technical constraints of carriers constitute a strong rationale for non trivial price schedules, depending non-linearly on the shipment size. Before discussing how this framework might be extended, some of its properties will be examined.

6.6 Road freight transport costs analysis

Representing the shipment size and the vehicle capacity constraint in a perfect competition equilibrium model is possible, as it has been shown in the previous section. The aim of this section is to re-examine some classical questions of cost structure analysis on the basis of this model.

The starting point, addressed in Subsection 6.6.1, is to examine the structure of the transport cost function of carriers and shippers. Subsection 6.6.2 then focuses on how these costs depend on such basic parameters as b, the access cost, or c, the hauling cost. It may be interesting to take a step back and examine how the transport industry's productivity depends on the population of shippers which constitute the transport demand. An experiment of this kind is presented in Subsection 6.6.3. Subsection 6.6.4 eventually addresses a somewhat annex question: how the productivity of carriers depends on the information they have on the numbers and sizes of the shipments they have to carry.

6.6.1 Transport costs structure

From the structure of costs in a given industry result the market structure at the equilibrium and the potential need for regulation. The objective of this section is to analyse the structure of the costs of carriers, as well as the structure of transport costs from the perspective of shippers.

Cost structure of carriers

The cost function of the carriers is given in Subsection 6.5.2. Due to Hypothesis 6.8, there are no scale effect: for all $\nu > 0$:

$$C(\nu q_1, \nu q_2, \nu q_3) = \nu C(q_1, q_2, q_3)$$

The road transport technology is thus endowed with constant returns to scale: the sizes of the carriers are not determined at the perfect competition equilibrium.

Concerning scope economies, as stated in Subsection 6.5.2, C is separable in $\{q_1; q_2\}$ and q_3, which means that the cost of producing a given bundle $\{q_1; q_2\}$ does not depend on the amount of q_3 produced, and conversely. However, q_1 clearly does influence the expected cost of carrying q_2 shipments of size 2, which implies that there are scope effects.

To be more precise, let us focus on the cost of carrying Q_1 shipments of size 1 and Q_2 shipments of size 2, with $q_1 = \mathbb{E}(Q_1)$ and $q_2 = \mathbb{E}(Q_2)$. We have:

$$\mathbb{E}(C(q_1, q_2)) = b(q_1 + q_2) + \frac{c}{3}\delta(q_1, q_2) + c.q_2$$

From Lemma 6.10, $\partial\delta/\partial q_1$ is a decreasing function of q_2:

$$\frac{\partial^2 C}{\partial q_1 \partial q_2} \leq 0$$

This equation holds on the whole domain of the cost function. Therefore:

Proposition 6.15 *The transport technology is characterised by constant returns to scale, and by economies of scope with respect to q_1 and q_2.*

It is more profitable for the carriers to transport shipments of size 1 and 2 together than to specialise in any of these two types of operation. This is consistent with the demonstration in Subsection 6.5.2 that for given prices, each carrier produces a bundle of transport operations characterised by a specific ratio $r = q_1/q_2$.

Transport costs from the perspective of shippers

The inputs carriers use to produce transport operations are the access and hauling movements. As for shippers, the operations produced by carriers are not the only inputs necessary to carry a flow of commodities from one place to another: the time spent by commodities during transport operations or waiting in inventories cannot be by-passed. As such, this time must be considered as an input; a user input, insofar as it is not provided by carriers, which provide the transport operations, but by shippers, which consume them.

This user input must be considered when addressing the structure of transport costs from the perspective of shippers. Furthermore, this structure is based on the equilibrium freight rates: it depends on the population of shippers. With respect to the shipper, the transport cost g is function of the flow rate[24] ϕ, the unit value of time a and the equilibrium price schedule $p(.)$.

Before g is derived as a function of ϕ, it is calculated for each shipment size s. Denote $g(\phi, s)/\phi$ the per ton generalised transport cost for a given

[24] It has been assumed hitherto that $\phi = 1$ for all shippers, for technical simplicity. As already explained, the results of the previous sections generalise easily to any ϕ for each shipper: the optimal shipment choice depends only on a/ϕ, so that Proposition 6.11 remains valid if a is replaced by a/ϕ. Proposition 6.12 is a bit trickier to generalise: the optimal shipment size depends only on a/ϕ, but the number of shipments is proportional to ϕ. Anyway, the main properties of the model remain, particularly the equilibrium existence and unicity.

shipment size s:

$$\begin{cases} g(\phi,1)/\phi &= a/2\phi + p(1) \\ g(\phi,2)/\phi &= a/\phi + p(2)/2 \\ g(\phi,3)/\phi &= 3a/2\phi + p(3)/3 \end{cases}$$

The optimal shipment size is therefore given by Proposition 6.11, provided a is replaced by a/ϕ. As a consequence, if $p(2) \leq p(1) + p(3)/3$, the generalised transport cost is:

$$\begin{cases} g(\phi) = a/2 + \phi.p(1) & \text{on} \quad [0;\ a/(2p(1) - p(2))[\\ g(\phi) = a + \phi.p(2)/2 & \text{on} \quad [a/(2p(1) - p(2));\ a/(p(2) - 2p(3)/3)[\\ g(\phi) = 3a/2 + \phi.p(3)/3 & \text{on} \quad [a/(p(2) - 2p(3)/3);\ +\infty[; \end{cases} \tag{6.15}$$

or, in the contrary case:

$$\begin{cases} g(\phi) = a/2 + \phi.p(1) & \text{on} \quad [0;\ a/(p(1) - p(3)/3)[\\ g(\phi) = 3a/2 + \phi.p(3)/3 & \text{on} \quad [a/(p(1) - p(3)/3);\ +\infty[; \end{cases} \tag{6.16}$$

As a consequence, from the subadditivity of $p(.)$:

Proposition 6.16 *From the perspective of shippers, transport costs exhibit economies of scale.*

Increasing returns are a feature of the EOQ model: they arise from the possibility of shippers to modify the trade-off they make between inventory costs and transport costs when the rate of the flow they need transported changes.

The capacity constraint increases the returns to scale in the zone where shippers send shipments of intermediate size. These shipments are cumbersome for carriers, who thereof increase $p(2)$, so that shippers tend to send either shipments of size 1 or 3. The concavity of the cost function is intensified by the capacity constraint.

6.6.2 Comparative statics: spatial parameters

The parameters from which the market equilibrium basically derives are the access cost b in the origin or destination area, which decrease with the density of activities in these areas (although the high levels of congestion in urban centers might mitigate the relevancy of this statement), and the cost c of the long haul movement. This section examines how the equilibrium price schedule depends on those two parameters, for a given population of firms F_a.

The influence of access costs

When the access cost b increases, a series of effects compete. Since the access cost constitutes an basic component of the equilibrium freight rates, those are expected to increase at the same speed as b, as can be seen in equation system (6.9). The share of the freight rates in the transport costs of shippers thus increases, so that shippers are expected to modify their logistic trade-off towards more inventory waiting time and less frequent shipments. The general increase in the sizes of shipments is thus unquestionable. This section is aimed at investigated the evolution of the freight transport market equilibrium at an additional level of detail.

When b increases, $p(1)$ and $p(2)$ increase (supposedly) at a similar pace. n_1^s should decrease with b, because the advantage of a small size gets mitigated by the per shipment component of the transport cost. The change in n_2^s is ambiguous: while some shippers sending shipments of size 2 opt for shipments of size 3, some shippers sending shipments of size 1 opt for shipments of size 2. The resultant evolution of n_2^s derives from two opposite trends, of which the outcome is indeterminate. The formal analysis, detailed in Appendix C.3, yields the following characterisation of the evolution of r^*:

Proposition 6.17 *The equilibrium ratio r^* decreases with b if and only if the following condition holds:*

$$\left(1+\frac{r^*}{2}\right).f_a(2p(1)-p(2))-\frac{r^*}{6}.f_a(p(2)-2p(3)/3)\geq 0 \tag{6.17}$$

This is a sufficient condition for $p(1)$ and $p(2)$ to be strictly increasing with b.

Such a condition cannot be true for all r^* and f_a. Consider indeed an equilibrium with r^* neither null nor indeterminate. Then it is always possible to modify slightly f_a so that the equilibrium prices and demands remain unchanged and $f_a(p(2)-2p(3)/3)$ is indefinitely large. For example, one can add some mass to f_a in the neighborhood of $p(2)-2p(3)/3$ so that the slope of F_a increases while F_a is unchanged at $p(2)-2p(3)/3$. Anyway, Equation (6.17) generally holds, as long as f_a does not vary too quickly.

As a conclusion, r^* generally decreases when b increases: the higher access costs met, for example, in less dense areas, prompt shippers to send bigger shipment. Therefore, the relative frequency of small shipments with respect to intermediate shipments generally decreases. This phenomenon is illustrated by Figure 6.6, which has been drawn with the following parameters: $c = 1.0$, $F_a : x \mapsto 1 - \exp(-x)$, and for various values of K_s. The limit case where K is almost zero is represented by the dashed curve.

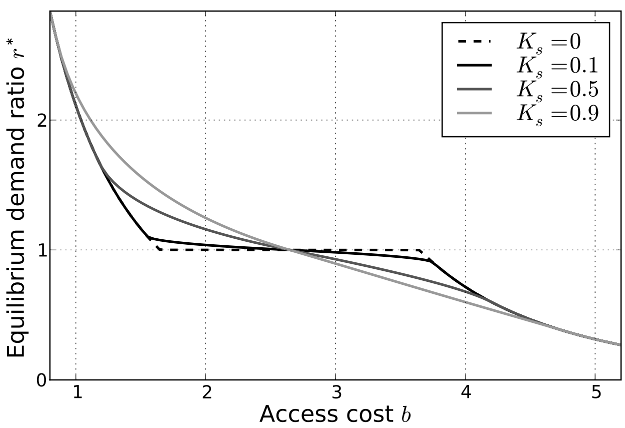

Figure 6.9: Influence of b on r^*, F_a smooth

However, this trend can be reversed in some situations due to the capacity constraint of carriers. This is the case depicted by Figure 6.10, where the unit value of time distribution F_a has been designed expressly so that Equation 6.17 does not hold at some point:

$$F_a : x \mapsto \begin{cases} 0.9(1.0 - \exp(-a)) & \forall a \in [0; 1.104[\\ 0.9(1.0 - \exp(-a) + 5.0 * (a - 1.104)) & \forall a \in [1.104; 1.124[\\ 0.9(1.0 - \exp(-a)) + 0.1 & \forall a \in [1.124[\end{cases} \tag{6.18}$$

Other parameters are unchanged.

The hitches in the middle of the curves show that the behaviour of r^* depends crucially on Equation 6.17. At some point, a large set of carriers shift from shipment size 2 to shipment size 3, so that r^* locally increases. However, the general direction taken by the r^* curves is toward zero, which is relatively consistent with our previous discussion.

This analysis illustrates the fairly complex behaviour of average load factors, transport demand and price schedule: even in a trivial model such as the one proposed here, it is quite a difficult task to deliver general conclusions on the influence of structural parameters on the market equilibrium. It may even happen, although it is quite improbable, that

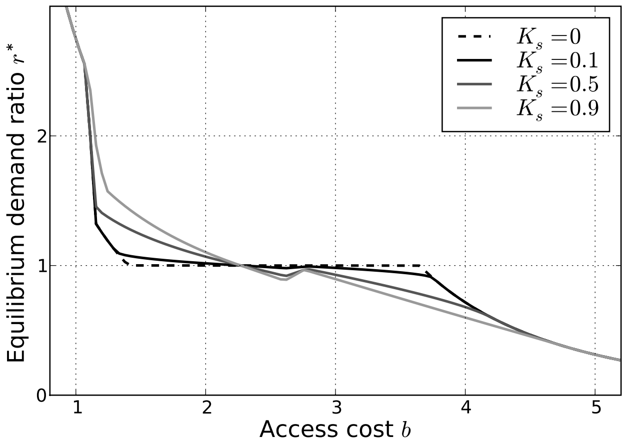

Figure 6.10: Influence of b on r^*, F_a irregular

a general increase in the access costs b implies a decrease of the freight rates from some shipment sizes.

The influence of haulage costs

The effect of the haulage cost c on the market equilibrium is not trivial. At first sight, the haulage cost is assigned to each shipment more or less on the basis of the amount of the vehicles' capacity they use. As such, it is basically a per ton cost, so that it should play no role in the shipment size choice. However, due to the capacity constraint, the freight rates depart from this scheme (this is the difference between the models introduced respectively in Sections 6.4 and 6.5).

Indeed, the main reason for which the capacity constraint has to be taken into account is that some shipment sizes imply an expected capacity loss to carriers. This loss is all the more expensive as c increases[25]. Therefore, the higher c, the more small shipments will be favored and intermediate shipments penalised by the equilibrium freight rates, so as to improve the average loading factor of the carriers' fleets. This intuition

[25] An interesting case to consider to understand the influence of the haulage costs on the freight transport market equilibrium is $c = 0$. In such a situation, the capacity constraint plays no role: each shipment should only pay its access movement. The equilibrium price schedules of carriers are constant and equal to b. The carriers do not care about the loading factor of their fleet. This proves that the influence of the loading factor on the equilibrium freight rates depends fully on c.

is consistent with the following result:

Proposition 6.18 *The equilibrium ratio r^* increases with c.*

Proof: See Appendix C.4.

The evolution between c and r^* is illustrated by Figure 6.11, drawn for $b = 2.2$ and F_a an exponential distribution of expected value 1.

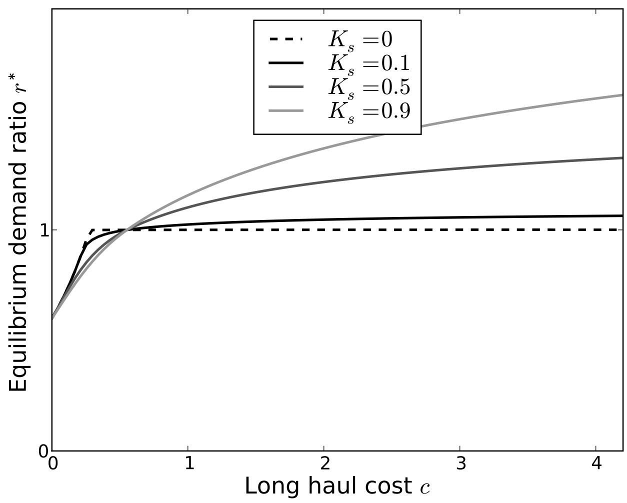

Figure 6.11: Influence of c on r^*

Note that a higher K_s implies a higher r^* for large values of c. It was indeed stated in Subsection 6.5.2 that the average loading factor of LTL trucks decreased when K_s increased. The cost of the unused capacity is proportional to c. As a consequence, for a given, large value of c, a higher K_s implies a modification of the equilibrium freight rates so as to foster smaller shipments, *i.e.* a smaller r^*, so as to minimise the unused vehicle capacity.

The haulage cost c has indeed an influence on the choice of shipment size. This result is intuitive: unused capacity is much more expensive on 500 km than on 150 km. The equilibrium freight rates are certainly influenced by this circumstance. By representing explicitly the technical constraints of the carriers and the logistic choices of the shippers, this approach is able to address this phenomenon.

6.6.3 Shippers characteristics and carriers productivity

This subsection investigates the linkage between F_a and the freight transport market equilibrium, all other things being equal.

This is a complex relationship. Basically, if one shipper, or a subset of the shippers' population find their unit value of time increased, they will send smaller shipments. The average shipment size should then decrease when F_a shifts towards higher values. However, the behaviour of r^* — and thus of prices — is ambiguous: the prevailing trend is toward a higher r^*, but if a number of shippers shift more or less simultaneously from shipment size 3 to shipment size 2, r^* might decrease.

The first case is illustrated by Figure 6.12, where $b = 2.5$, $c = 1$, and F_a is an exponential distribution of expected value 1, of which the expected value is multiplied by a parameter ranging from 0.1 to 2.5. As previously, the r^* curve is drawn for four different values of the variability parameter K_s. As expected, r^* increases when the value of time of the shippers increase.

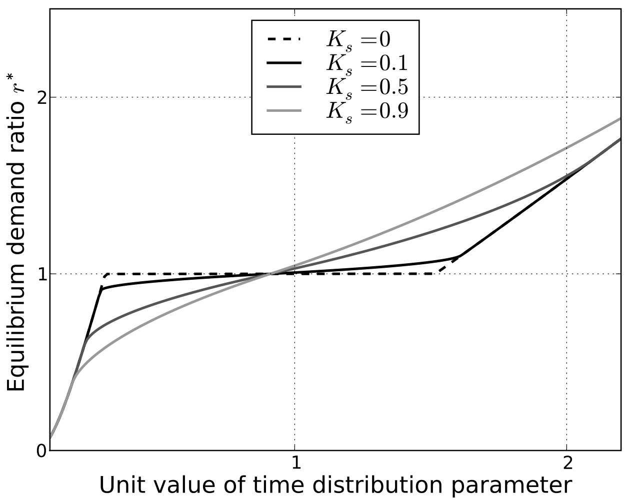

Figure 6.12: Influence of F_a on r^*, F_a smooth

However, this behaviour is not systematic. Figure 6.13 illustrates the movement of r^* in the same conditions as above, except that F_a is not smooth, but has the irregular behaviour described by Equation (6.18). r^* is indeed decreasing on a given range of values when the expected value of F_a increases.

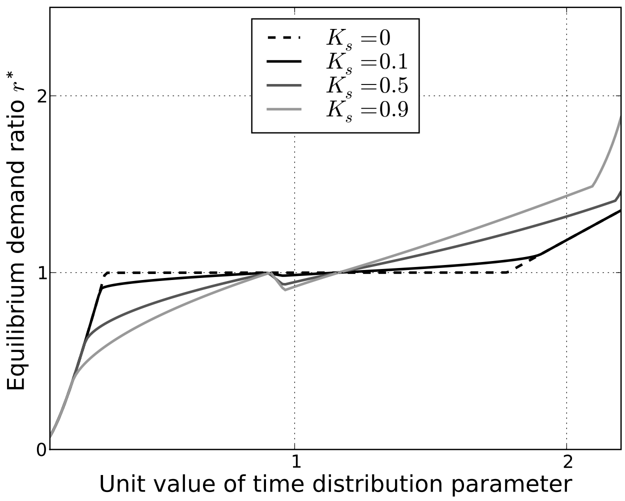

Figure 6.13: Influence of F_a on r^*, F_a irregular

r^* is directly linked to the average loading factor of the vehicles carrying LTL shipments. But the average loading factor of a carrier concerns as much the vehicles carrying FTL shipments as the vehicles carrying LTL shipments.

This average loading factor may increase or decrease when shippers send smaller shipments. Indeed, assume for example that all the shippers have the same unit value of time. If this value is very small, they all send shipments of size 3, so that all the vehicles are full. For some intermediate unit value of time, they all send shipments of size 2: the average loading factor is then 2/3. Finally, for a high unit value of time, they all send shipments of size 1, in which case vehicles are full again. This simple example illustrates the difficulty to identify a simple relationship between the characteristics of the demand for transport and the productivity of carriers, even in the frame of this particularly simple model.

Figure 6.14 illustrates this phenomenon in the frame of another set of hypotheses. The following parameters are assumed $b = 2.5$, $c = 1.0$, F_a exponential. The overall expected loading factor has been calculated along the following steps. The expected number of LTL trucks n^t_{LTL} derives as follows from the number of LTL shipments n^s_1 and n^s_2, and of the expected loading factor of a LTL truck λ^t_{LTL} defined in Section 6.5.2 and calculated in Appendix C.2:

$$n^t_{LTL} = \frac{1}{\lambda^t_{LTL}} \left(\frac{n^s_1}{3} + \frac{2n^s_2}{3} \right)$$

The number of FTL trucks is derived in a similar fashion, although more simply:

$$n^t_{FTL} = n^s_3$$

The average loading factor λ^t is then:

$$\lambda^t = \frac{n^t_{LTL}}{n^t_{FTL} + n^t_{LTL}} \lambda^t_{LTL} + \frac{n^t_{FTL}}{n^t_{FTL} + n^t_{LTL}}$$

All these variables are calculated at the equilibrium.

As one can observe on this figure, the average loading factor of the carriers first decreases, then increases when the unit values of time of the shippers increase. The reason is basically the same as in the previous examples: first, the shippers have small unit values of time, so that they send FTL shipments. The vehicles run full, thus the average loading factor of 1. As the values of time increase, some shippers start to send smaller shipments. Therefore, a large amount of shipments of intermediate size is to be expected first, which is problematic from the standpoint of the road freight transportation industry (the vehicles carrying shipments of size 2 run partially empty if there is no shipment of size 1 to use the remaining capacity). The average loading factor decreases. However, whereas the unit values of time of shippers continue to increase, more and more shipments of size 1 are send, so that λ_{LTL} progressively increases from 2/3 to 1. In the end, the average loading factor of carriers tends towards 1, so that the productivity of carriers increases[26].

The picture drawn here is, of course, quite schematic. In particular, the very high equilibrium loading factors are due to the idealized representation of the freight transport market: carriers face many more constraints than the ones taken into account in this analysis; these additional constraints explain the much lower loading factors actually observed.

Despite this limitation, one conclusion at least can be drawn: the linkage between the transport demand characteristics and the productivity of road freight transport is complex. Apart from the simple recommendation to always distinguish FTL and LTL transport when possible, little can be said about road freight transport's productivity, with respect to the capacity constraint of the vehicles and the choice of shipment size of shippers.

[26] The evolution of the average loading factor is due to the fact that small shipments are perfectly consolidated in this model. However, this is not realistic. Indeed, for the shippers with high unit values of times, the shippers can decide not to fill their vehicles to reduce the gathering time, so that the travel time decreases. Furthermore, the carriers can opt for carriers who use smaller and/or faster vehicles. On the whole, this model is limited by the fact that the travel time is considered as a parameter.

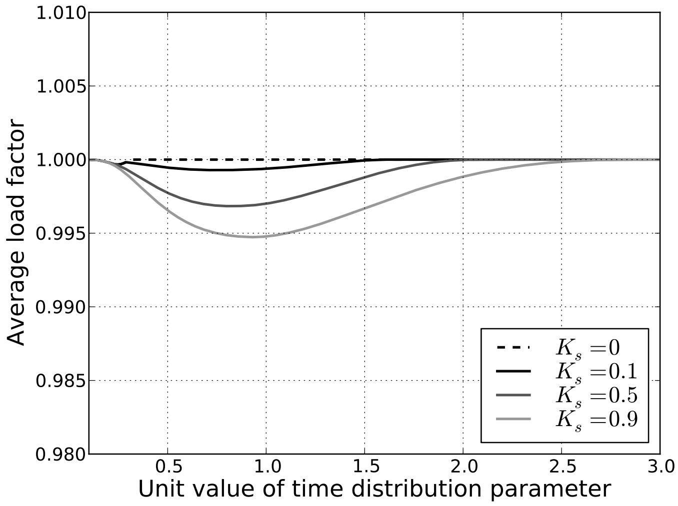

Figure 6.14: Influence of F_a on the average loading factor

6.6.4 Information and road freight transport productivity

One parameter plays a special role in the equilibrium of the freight transport market: the variability of the transport demand parameter K_s. In this framework, K_s results from the way shippers choose carriers. As shippers do not coordinate their decisions, it may happen, for example, that one carrier has to transport two shipments of size 2, and another one two shipments of size 1. Their vehicles would thus run partially empty, whereas they could improve significantly their productivities by swapping two of these shipments. Due to the fact that the carriers have to choose their price schedules before knowing exactly how many shipments of each size they will have to transport, a higher variability of the demand means higher rates.

It also means lower loading factors, both if the loading factor is calculated for the LTL trucks, as illustrated in Figure 6.6, or if it encompasses all the vehicles, as depicted by Figure 6.14. From the standpoint of carriers, the variability of the demand is a fundamental driver of productivity.

This observation has an important consequence: any carrier would be ready to pay to reduce the variability of its demand, or, in other words, to have more information on it. This value of information provides a strong rationale for contracts between carriers and shippers where the carrier would offer attractive prices whereas the shipper guarantees either regular flows or accurate information on the shipments it will send. A shipper with regular flows would thus be offered better rates than a shipper with irregular flows. The carriers would be ready to offer better rates in compensation of guarantees on their freight flows.

Up to this point, K_s has been considered as a uniform parameter. However, it is certainly possible to build a model with a distributed K_s; distinct rates and contract types should emerge. The difference of K_s between shippers would thus appear to be one explanation of the heterogeneity in the types of contracts carriers and shippers bind with one another; these contracts range indeed from spot transactions to several year long contracts with invitation to tender.

6.7 Conclusion

The general purpose of this study is to re-examine the structure of the costs of carriers as well as the structure of transport costs from the perspective of shippers. The focus has been brought on representing explicitly the size of shipments, as well as the technical constraints of carriers.

After some qualitative elements concerning the road freight transport market are described, a simple formal framework is introduced. A set of hypotheses describes the road freight transport market structure. In particular, carriers are assumed to be in perfect competition. The behaviour of shippers is detailed: the shipment size is modeled as the result of a trade-off made by shippers between transport costs and inventory costs.

A simple equilibrium model is derived, in which the capacity constraint of carriers plays a trivial role. It is proved that this model is not consistent with some qualitative properties of the freight transport market. A more sophisticated model is thus introduced, where the joint production nature of the road freight transport technology is explicitly represented. It demonstrates that perfect competition and marginal cost pricing are consistent with complex freight rates, due to the interactions between freight rates for different shipment sizes.

The properties of the equilibrium transport demands and freight rates are then analysed. However simple the representation of the trucking technology is, the results are non trivial. The cost structure analysis is proceeded to first. Consistently with the hypotheses of the model, the carriers technology is endowed with constant returns to scale. It is also endowed with economies of scope: quite intuitively, carriers have a strong incentive to bulk LTL shipments when possible. Due to the user input they bring, the transport cost is endowed with increasing returns to scale from the shippers' perspective. These increasing returns to scale are consistent with perfect competition between carriers, but provide an incentive for shippers to increase their sizes. This demonstrates the importance of distinguishing carriers and shippers when addressing the structure of freight transport costs.

The spatial properties of the model are then investigated. Considering the shipment as the unit of the model allows the distinction between access costs and hauling costs. Larger access costs tend to increase the sizes of shipments, as shippers pay them on the basis of the number of shipments sent, and not on their sizes. On the contrary, larger hauling costs influence the freight rates so that the distribution of shipment sizes optimise the loading factors of the carriers, since unused capacity gets more expensive.

The characteristics of the demand play an important role in the equilibrium freight rates and road freight transport productivity, but no simple relationship links them. In particular, an increase in the unit values of time of the commodities shipped does not necessarily imply a decrease in the loading factors.

A particular attention is brought to the analysis of the demand variability parameter. This parameter has been considered as fixed and uniform along all the analysis. However, it proves to have an interesting role. Indeed, if considered as an amount or quality of information carriers have on their demands, it is shown to be positively related to their productivity: a carrier may reduce its costs if it has more information on the shipments to be transported. A value of information naturally emerges from this approach: it could be a further explanation of the heterogeneity in the natures of contracts the shippers bind with the carriers, as well as in freight rates.

As a conclusion, it should first be noted that the framework of this study lacks a considerable amount of features, which cannot be ignored in the frame of a comprehensive model of the freight transport market. Furthermore, a series of strong hypotheses were made so that the analysis keeps tractable. There are no backhaul movements, no possibility for complex routes, the loading factor is only calculated during the hauling movement, not the access movements. There is no possibility for the carriers to use smaller or larger vehicles, to set break-bulk platforms in order to optimise the bulking of small shipments, or to limit the loading factors of their vehicles in order to improve the travel time.

Nevertheless, this framework introduces with the shipment size an additional level of detail in the microeconomic analysis of the road freight transport market. It is a first step towards making the linkage between the freight transport market and the logistic imperatives of shippers. It allows for the representation of a series of qualitative properties which have no signification in the classical representation by flows. It highlights the complexity of the apparently simple trucking technology, and argues that the heterogeneity in freight rates is not necessarily inconsistent with perfect competition and constant returns to scale for the carriers.

Despite some considerable technical difficulties, particularly with respect to the combinatorial nature of the bin-packing problem, this approach raises a large number of questions, notably concerning such phenomenons as unbalanced flows, location of logistic platforms, or the linkage between travel time and the type of vehicle and loading factor choices by carriers, etc.

Chapter 7

Logistic imperatives and modal choice

7.1 Introduction

Many transport externalities depend strongly on the transport mode used, so that modal shift is perceived as a priority to diminish the social, economic and environmental impacts of freight transport. The relative lack of success of modal shift transport policies meet in practise indicates that mode choice may be partially misunderstood. Why do shippers use a given transport mode? Why is it difficult to induce them to modify their choices? These are central questions of freight transport economics and modeling.

The classic modeling approach to the question of modal choice is based on a representation of the transport modes by their characteristics, in the tradition of the consumer theory of Lancaster (1966). These characteristics are typically the transport price and the travel duration. This approach yields the well-known value of time, or value of travel time savings, the marginal substitution rate between time and price (Quinet and Vickermann, 2004), which is a key parameter of spatialised transport models in general (Small and Verhoef, 2006). This approach generally proves econometrically insufficient, so that other variables are introduced: mode specific constants in many cases, attempts to introduce some other characteristics, such as reliability (see e.g. Fowkes et al., 2004). These approaches have two main shortcomings. First, there is no notion of shipment size, yet a crucial decision variable in freight transport. Second, they cannot explain why a firm would use two transport modes simultaneously, although some firms do it in practise.

Several strategies are conceivable to take into account the shipment size in freight transport models, either microeconomically or econometrically. One of the most interesting ones, because it has the virtue to set the shipper's transport decisions in a more general logistic framework, is the work of Baumol and Vinod (1970), based on the very well known Economic Order Quantity (EOQ) model of Harris (1913). In this model, the shipment size stemf from a trade off between the fixed shipment costs and the inventory holding costs. Baumol and Vinod derived the preferences of shippers with respect to transport modes on the basis of the EOQ model: this is the first step in integrating logistic elements in freight transport economics. It has been done again in analytical, theoretical models (see e.g. Hall, 1985), but only very recently in spatialised, applied models. The first attempt to build a spatialised freight transport demand model with explicit shipment sizes is that of de Jong and Ben-Akiva (2007), based on a generalised EOQ formula. In all these studies, the models still yield a unique transport mode for a given shipper.

The choice of shipment size is also a central variable in inventory theory. In this field, the focus is not put on travel demand but on limiting simultaneously the amount of commodities stocked at a given instant and the number of customers waiting for their orders. The problem for the firm is then to decide at each period of time (for a periodic review model) the amount of commodities that should be produced or carried to the inventory. Some extensions of these models address specifically the problem of firms using simultaneously two transport modes[1]. Their objectives is to provide algorithms which determine when to send which shipment by which mode, when the demand is uncertain. This is a complex problem. As a consequence, the solutions available in the literature often rely on limiting hypotheses, so that their outreach is questionable. Furthermore, as usually in complex problems of operations research, their results are hard to analyse from a microeconomic perspective. The demand's stochasticity is a fundamental hypothesis of these models.

Explaining with a microeconomic language why a shipper would use two distinct transport modes simultaneously for a given commodity flow is an important theoretical issue, for three reasons. First, such firms do exist. It is typical of some firms to use a heavy, slow and potentially not very reliable transport mode (such as rail or sea transport) to save on transport costs, in parallel to a fast, expensive transport mode which

[1] The references addressing this problem quoted in this chapter are Karlin and Scarf (1958), Barankin (1961), Morton (1969), Moinzadeh and Nahmias (1988), Huh et al. (2009), and Su and Zhang (2009). Their assumptions and results will be detailed later on. They are but a small subset of the studies on optimal simultaneous use of two modes.

ensures the firm stays reactive and is able to solve quickly an unexpected issue (such as an unexpectedly large demand or travel time). Second, these firms could make the main target of a modal shift transport policy. It is thus important to identify which parameter has the highest influence on their choices. Third, a microeconomic model explaining the simultaneous use of two transport modes necessarily encompasses a description of the shipper's logistic imperatives. Modeling these imperatives and explaining their influence on the choices of the shippers is currently a major issue of freight transport economics. As it will appear later, the impossibility to forecast precisely the demand is one of the most important drivers of the logistic imperatives of the shippers.

The objective of this paper is provide a microeconomic explanation for the simultaneous use of two transport modes by a unique shipper, for a unique commodity flow. The methodology adopted consists in designing a model inspired from the inventory theory, simple enough for a microeconomic analysis to be possible. This is based on a simple single commodity periodic review model. The economic analysis is focused on the preferences of shippers with respect to modal choice, and on the role of their logistic imperatives in their preferences.

The classic case with one transport mode is first presented in Section 7.2, including an economic analysis of the results. A simple heuristic logistic policy for the simultaneous use of two modes is then presented and analysed in Section 7.3.

7.2 A model with one transport mode

The first step of this study consists in the presentation of a simple model with only one transport mode. Section 7.2.1 presents the framework of the model. The model itself is developed in Section 7.2.2, and analysed in Section 7.2.3. Numerical applications are presented in Section 7.2.4, and further comments in Section 7.2.5.

7.2.1 Framework

Consider a firm, thereafter called the shipper, owning a production unit in A, and a retail center in B. This shipper produces commodities of many distinct types in A, and sells them to its customers in B. There is only one transport mode available to carry all these commodities. This mode is used with a given frequency, say, without loss of generality, once per day, to carry all these commodities. The transport lead-time[2] is denoted l.

As a profit maximiser, whatever the market structure, the shipper minimises its costs (including, if any, the user costs, see Mohring, 1985)[3]. The analysis is focused here on the supply chain associated to a unique commodity type. By supply chain, we mean the commodity flow and stocks, from production to delivery.

Customers order each day a given amount of this commodity type. If they are not served straightaway, they wait. However, they regard waiting as a discomfort; in other words, they are willing to pay to reduce or avoid waiting time. As a consequence, pending orders incur a cost the shipper takes into account. The shipper can decrease the number of customers waiting by increasing the inventory level.

However, a high inventory implies two types of costs. First, warehousing requires space, buildings, monitoring, and handling, which are costly. Second, commodities owned by the shipper imply a capital opportunity cost as well as, depending on the commodity type, a depreciation cost.

The problem for the shipper is to decide the amount of commodities to ship each day, so as to minimise the related transport and inventory costs.

Demand distribution During each time period $t \geq 0$, the customers order a amount D_t of the considered commodity at the retail center. The demands $\{D_t\}_{t\geq 0}$ are assumed stochastic, independent and identically distributed. For analytical convenience, their common distribution is assumed discrete[4]:

$$\left\{ \begin{array}{lcl} \mathbb{P}\{D_t = d - \sigma\} & = & 1/2, \\ \mathbb{P}\{D_t = d + \sigma\} & = & 1/2, \end{array} \right. \tag{7.1}$$

[2]The model presented in this chapter is built from the standpoint of the shipper. As a consequence, apart from some aspects such as the vehicle type (the shipper needs berths to load and unload trucks quickly, railway sidings to have access to rail transport) the way the transport operation is practically realised is not relevant. In particular, if the carrier combines several vehicle movements in addition to break-bulk operations (e.g. for a motor carrier carrying LTL shipments) or moves railcars through marshalling yards (for a rail carrier), the actual time the commodities spend moving can be much smaller than the whole transport operation duration. This is why we prefer to speak about transport lead time rather than about travel time.

[3]Mohring's statement is based on the following intuition: if it costs the shipper less than 1 to reduce by 1 the cost customers perceive, the shipper may proceed to this reduction, increase its price by 1, have the same number of customers and increase its own profit, even though the shipper's cost function does not encompass all the elements of the customers' cost function, such as waiting time.

[4]Note that this assumption is only needed in Section 7.3.

with d and σ positive constants, and $\sigma \leq d$. As a consequence, D_t's expected value is:

$$\mathbb{E}(D_t) = d,$$

and its variance is:

$$\mathbb{V}(D_t) = \sigma^2.$$

D_t's cumulative distribution demand is denoted F_D.

System dynamics The framework is modeled as a single-product single-location inventory system under periodic review, where excess demand is kept and the replenishment lead time is positive.

The destination inventory at time t is denoted I_t. Note that this variable can be negative. In that case, it stands for the amount of orders pending or, equivalently, the number of customers waiting. The shipment sent at time t is denoted s_t. The pipeline inventory (*i.e.* the amount of commodities being carried at time t) is denoted I_t^p.

The destination inventory dynamics are simple: at the beginning of period $t+1$, the destination inventory is the destination at time t plus the shipment sent at time $t-l$ minus the demand ordered during period t:

$$I_{t+1} = I_t + s_{t-l} - D_t \tag{7.2}$$

The pipeline inventory dynamics are even simpler. Quite naturally:

$$I_{t+1}^p = I_t^p + s_t - s_{t-l} \tag{7.3}$$

For analytical convenience, the following conventions are taken: $\forall t < 0, D_t = d,$ and $s_t = d$.

Costs In this framework, four types of costs are distinguished. Each of them is thereafter defined and calculated over period t.

The first cost considered is the direct transport cost C_t, *i.e.* the amount paid by the shipper to the carrier for the commodities to be actually moved. In the framework considered, the transport system is designed for a large number of commodity types. With respect to the structure of costs, this means that a higher shipment size s_t means more vehicle capacity, but no change in the shipment frequency, for example, and a negligible change in the loading and unloading times. As a consequence, the daily transport cost is assumed proportional to the shipment size:

$$C_t = c_t s_t. \tag{7.4}$$

Second, the pipeline inventory cost C_p, *i.e.* the inventory cost due to the time the commodities wait during the transport operations. This cost is assumed proportional to the time spent by the commodities during transport, up to a coefficient a_c standing for the amount the shipper would be ready to pay to decrease by 1 day the time waited by one ton of commodity in the shipper's inventory (a_c encompasses the capital opportunity cost and the depreciation cost, but not the warehousing cost).

During period t, this cost is equal to the amount of commodities currently being carried (it thus depends on the transport lead-time l) times the unit commodity value of time.

$$C_p = a_c \sum_{i=t-l}^{t-1} s_i. \tag{7.5}$$

Third, the destination inventory cost: proportional to the time the commodities wait in the retail center, before they are sold, this cost consists of the capital and depreciation cost a_c, and of the warehousing cost, assumed proportional to the amount of commodities being stocked, up to a coefficient denoted a_w:

$$C_d = (a_c + a_w)(I_t)^+, \tag{7.6}$$

where $(.)^+$ denotes the positive value.

Fourth, the customer cost[5]: it is assumed proportional to the number of customers waiting for their orders at time t, equal to $|I_t|$ if $I_t < 0$, times the unit customer value of time (the amount the customers would be ready to pay to reduce their waiting time of one day), denoted by a.

$$C_c = a(I_t)^-. \tag{7.7}$$

The shipper's objective is to minimise the sum of these four costs[6]. The sum of these costs is thereafter called the full logistic cost C:

$$C = C_t + C_p + C_d + C_c. \tag{7.8}$$

[5]This cost type can be referred to as penalty or shortage cost in the inventory theory literature.

[6]Strictly speaking, there is a cost component for each period of time. Typically, the objective for the shipper would then be to minimise an actualised sum of these costs. However, this simple analysis will be limited to the minimisation of the expected cost per time period.

From Equations (7.4) to (7.7), Equation (7.8) becomes:

$$C = c_t s_t + a_c \sum_{i=t-l}^{t-1} s_i + (a_c + a_w)(I_t)^+ + a(I_t)^-. \tag{7.9}$$

The objective of the shipper is to determine at each period t the shipment size s_t which minimises the logistic cost C under the system dynamics given by Equations (7.2) and (7.3).

The detailed approach will not be reported here. We will take for granted that in this framework, the optimal logistic policy is an order-up-to policy. In other words, the amount sent in each period from the production unit to the retail center exactly compensates the quantity ordered at the retail center. As a consequence, the destination inventory plus the pipeline inventory, *i.e.* the overall inventory, remains constant. The level of the overall inventory is chosen so as to minimise the expected cost for any time t. See Karlin and Scarf (1958) for a formal analysis.

7.2.2 The optimal logistic policy

In this framework, the difficulty for the shipper stems from the demand uncertainty and the transport lead-time: the system's state at time t derives from decisions the shipper has taken at least l days earlier. The objective of the shipper is to address this uncertainty optimally with respect to the costs introduced earlier.

The best way to do so is to follow an optimal order-up-to policy. In this section, we show how this result stems intuitively from the problem's structure. Then, the optimal inventory level is derived.

Description of the optimal logistic policy Given the regularity of the problem, one can expect the optimal logistic policy to be such that the destination inventory's expected value is equal to a given target. Let I_s denote this target, called the safety stock. In other words, the objective is:

$$\mathbb{E}(I_t) = I_s.$$

Now consider the information available to the shipper at time t: the shipper knows the inventory I_t and the shipments sent over the l previous periods. The shipper decides at time t the size of the shipment s_t which will arrive at $t + l$.

By combining Equation (7.2) $l+1$ times we obtain:

$$I_{t+l+1} = I_t - \sum_{i=t}^{t+l} D_i + \sum_{i=t-l}^{t-1} s_i + s_t, \tag{7.10}$$

where I_t and $\{s_i\}_{i=t-l,\ldots,t-1}$ are given, s_t to be decided, $\{D_i\}_{i=t,\ldots,t+l}$ yet unobserved.

By taking the expected value of Equation (7.10), necessarily:

$$\mathbb{E}(I_{t+l+1}) = I_t - (l+1)d + \sum_{i=t-l}^{t-1} s_i + s_t,$$

so that to obtain $\mathbb{E}(I_{t+l+1}) = I_s$, s_t must be equal to:

$$s_t = I_s + (l+1)d - I_t - \sum_{i=t-l}^{t-1} s_i. \tag{7.11}$$

If this policy is applied, then, by replacing s_t from the equation above in (7.10) and translating $l+1$ periods backwards, the destination inventory at time t is:

$$I_t = I_s + \sum_{i=t-l-1}^{t-1} (d - D_i). \tag{7.12}$$

(Note that this is true only for $t \geq l$. For simplicity, we will only consider $t \geq l$ in the following.)

This equation is of central importance. First, it tells us that under the optimal logistic policy, the destination inventory is a random variable centered on I_s, and of which the variance is clearly related to the variance of D_t and to the transport lead time. Second, the distribution of I_t is identical for all $t \geq l$. Third, as it will be explained in more detail later, the expected full logistic cost depends almost entirely on this equation.

Let us introduce right away the following notations:

$$\left\{ \begin{array}{lcl} \mu_I & = & \mathbb{E}(I_t), \\ \sigma_I^2 & = & \mathbb{V}(I_t). \end{array} \right. \tag{7.13}$$

From Equation (7.12), the destination inventory's expected value is:

$$\mu_I = I_s, \tag{7.14}$$

and its standard deviation is:

$$\sigma_I = \sqrt{l+1}\sigma. \tag{7.15}$$

Besides, Equation (7.12) yields:

$$I_{t+1} - I_t = D_{t-l-1} - D_t,$$

whereas from Equation (7.2):

$$I_{t+1} - I_t = s_{t-l} - D_t.$$

As a consequence, the optimal order-up-to logistic policy simply reduces to:

$$s_t = D_{t-1}, \tag{7.16}$$

thus its name.

At each period, as soon as the demand is observed, a shipment is sent to compensate exactly the amount of commodity which has just left the destination inventory. Along this policy, the inventory at time t is given by Equation (7.12). One thing remains to be decided: the value of I_s.

Incidentally, Equation (7.16) implies that:

$$\mathbb{E}(s_t) = d. \tag{7.17}$$

The optimal safety stock The choice of a target value for the safety stock relies solely on Equations (7.9), (7.12) and (7.16). The objective of this section is to derive the optimal I_s, *i.e.* the safety stock which minimises the expected value of the full logistic cost, $\mathbb{E}(C)$.

Let us first consider the transport and pipeline inventory costs. Both of them only depend on the shipment sizes $\{s_t\}$. From Equation (7.16):

$$\mathbb{E}(s_t) = d,$$

on average these two costs do not depend on I_s.

As a consequence, denote:

$$\mathbb{E}(C)(I_s) = \mathbb{E}(C_d)(I_s) + \mathbb{E}(C_c)(I_s) + K,$$

where K is a constant. The I_s variable is omitted from now on, unless necessary.

From Equations (7.6) and (7.12), the expected destination inventory cost is:

$$\mathbb{E}(C_d) = (a_c + a_w)\mathbb{E}\left[\left(I_s + \sum_{i=t-l-1}^{t-1}(d - D_i)\right)^+\right],$$

and from Equation (7.7), the expected customer cost is:

$$\mathbb{E}(C_c) = a\mathbb{E}\left[\left(I_s + \sum_{i=t-l-1}^{t-1} (d - D_i)\right)^-\right].$$

These two equations reveal the role of I_s: a large safety stock means fewer customers waiting, but a higher inventory cost, and conversely. Note that I_s can be negative: the shipper decides that the consumers usually wait, its only objective is that they do not wait too long.

Let D_t^r denote the sum of the orders placed over $\{t-l-1, \ldots, t-1\}$:

$$D_t^r = \sum_{i=t-l-1}^{t-1} D_i,$$

so that Equation (7.12) becomes:

$$I_t = I_s + (l+1)d - D_t^r.$$

The D_t^r are identically distributed, but not independent. Strictly speaking, D_t^r is a discrete random variable. However, due to the central limit theorem, the distribution of D_t^r is almost gaussian, provided the travel time l is not too short. As a consequence, in order to avoid tedious considerations, which would bring little to the analysis, D_t^r is considered continuous and normally distributed.

If D_t^r is normally distributed, then I_t is also normally distributed, of mean μ_I and variance σ_I^2. The c.d.f. of I_t, denoted F_I, is thus approximately:

$$F_I = \Phi\left(\frac{x - \mu_I}{\sigma_I}\right),$$

and its distribution density, denoted f_I, is:

$$f_I = \frac{1}{\sigma_I}\varphi\left(\frac{x - \mu_I}{\sigma_I}\right).$$

It is now possible to determine how $\mathbb{E}(C_d)$ varies with I_s. Indeed:

$$\mathbb{E}\left[(I_t)^+\right] = \int_0^{+\infty} x f_I(x) dx.$$

By replacing f_I:

$$\mathbb{E}\left[(I_t)^+\right] = \int_0^{+\infty} \frac{x}{\sigma_I}\varphi\left(\frac{x - \mu_I}{\sigma_I}\right) dx;$$

then by variable substitution:

$$\mathbb{E}\left[(I_t)^+\right] = \int_{-\frac{\mu_I}{\sigma_I}}^{+\infty} (\sigma_I v + \mu_I)\varphi(v)\,dv. \tag{7.18}$$

From Equations (7.14) and (7.15), $\partial\mu_I/\partial I_s = 1$, and $\partial\sigma_I/\partial I_s = 0$. As a consequence:

$$\frac{\partial\mathbb{E}\left[(I_t)^+\right]}{\partial I_s} = \int_{-\frac{\mu_I}{\sigma_I}}^{+\infty} \varphi(v)\,dv,$$

which, given that $\Phi(x) = 1 - \Phi(-x)$, is equivalent to:

$$\frac{\partial\mathbb{E}\left[(I_t)^+\right]}{\partial I_s} = \Phi\left(\frac{\mu_I}{\sigma_I}\right).$$

The same kind of calculations applied to $\mathbb{E}\left[(I_t)^-\right]$ yield:

$$\frac{\partial\mathbb{E}\left[(I_t)^-\right]}{\partial I_s} = \Phi\left(\frac{\mu_I}{\sigma_I}\right) - 1,$$

so that the behaviour of $\mathbb{E}(C)$ with respect to I_s is given by:

$$\frac{\partial\mathbb{E}(C)}{\partial I_s} = (a_c + a_w + a)\Phi\left(\frac{\mu_I}{\sigma_I}\right) - a, \tag{7.19}$$

which, given Equations (7.14) and (7.15), is equivalent to:

$$\frac{\partial\mathbb{E}(C)}{\partial I_s} = (a_c + a_w + a)\Phi\left(\frac{I_s}{\sqrt{l+1}\sigma}\right) - a. \tag{7.20}$$

Therefore, the expected full logistic cost is a convex function of I_s, minimised when the derivatives of the expected destination inventory cost and the expected customer cost are equal except for the sign. The optimal safety stock is derived straightforwardly from Equation (7.20):

Proposition 7.1 *The safety stock minimising the expected full logistic cost is approximately:*

$$I_s^* = \sigma\sqrt{l+1}\,\Phi^{-1}\left(\frac{a}{a_c + a_w + a}\right). \tag{7.21}$$

The role of the safety stock is clearly illustrated by this result: I_s^* does not depend on d, and is proportional to σ. Apart from the pure transport cost, all the logistic costs considered in this framework stem from the demand's uncertainty, which plays a crucial role. Indeed, if the demand were certain, there would be no need for a safety stock; in fact, there would be neither inventory nor customer costs: the shipper would send each day the exact amount which would fulfill the needs of his customers.

However, the demand uncertainty is not the only source of the need for a safety stock: the safety stock is made necessary by the positive transport lead time. A greater transport lead time means a less reactive system, and the need for a greater safety stock. As a consequence, the transport lead time is a central parameter for the shipper.

Up to this point, we have just provided very classical elements of inventory theory. The remaining paragraphs of this section aim at analysing these results from a microeconomic, freight transport modelling perspective.

7.2.3 Transport demand analysis

Now that the shipper's logistic policy has been introduced, it is analysed microeconomically. This section is focused on how the safety stock and the full logistic cost depend on the model parameters. These parameters are grouped into three categories: the inventory and customer costs, the transport mode characteristics, and the demand characteristics.

Inventory and customer costs The microeconomic interpretation is quite straightforward: the optimal safety stock is a trade off between destination inventory costs and the customer cost. Since the first ones tend to increase with a_c or a_w, while the latter one increases with a, it is not surprising to observe from Equation (7.21) that I_s^* decreases with a_c and a_w, and increases with a. Quite intuitively, simple calculations would easily show that the expected full logistic cost increases with each of these three parameters.

Transport characteristics, value of time The transport mode used by the shipper is described by two of its characteristics: its unit cost c_t and its lead time l. The roles these two parameters play are not symmetric.

Indeed, from Equation (7.21), the cost parameter c_τ has no influence on the safety stock level. Since there is only one mode available, an increase in c_t does not impact the shipper's transport demand either.

The only consequence is a change in the full logistic cost:

$$\frac{d\mathbb{E}(C)}{dc_t} = d, \tag{7.22}$$

where $\mathbb{E}(C)$ is the expected full logistic cost under the optimal logistic policy, a function of the model's parameters.

On the contrary, the transport lead-time has a non-trivial impact both on the shipper's logistic policy and full cost. First, it appears straightforwardly from Equation (7.21) that the safety stock increases with l. Consistently with what was explained at the end of Section 7.2.2, the safety stock addresses the demand uncertainty and the system's lack of reactivity. An increase in the transport lead time means a loss of reactivity, which comes at a cost which is not limited to an increase in the pipeline inventory cost.

To calculate the derivative of the full logistic cost with respect to l, first note that I_s^* is a function of l. However, from the envelop theorem:

$$\frac{d\mathbb{E}(C)}{dl} = \left.\frac{\partial\mathbb{E}(C)}{\partial l}\right|_{I_s=I_s^*}$$

As a consequence, μ_I can be considered fixed in Equation (7.18), and we have:

$$\frac{\partial\mathbb{E}\left[(I_t)^+\right]}{\partial l} = \frac{\partial\sigma_I}{\partial l}\int_{-\frac{\mu_I}{\sigma_I}}^{+\infty} v\varphi(v)dv.$$

Given that $\varphi'(v) = -v\varphi(v)$:

$$\frac{\partial\mathbb{E}\left[(I_t)^+\right]}{\partial l} = \frac{\partial\sigma_I}{\partial l}\left[-\varphi(v)\right]_{-\frac{\mu_I}{\sigma_I}}^{+\infty},$$

and, from the symmetry of φ:

$$\frac{\partial\mathbb{E}\left[(I_t)^+\right]}{\partial l} = \frac{\partial\sigma_I}{\partial l}\varphi\left(\frac{\mu_I}{\sigma_I}\right).$$

The same result is obtained for $(I_t)^-$.

Besides, Equation (7.15) implies that:

$$\frac{\partial\sigma_I}{\partial l} = \frac{\sigma}{2\sqrt{l+1}}.$$

From this result, as well as from Equation (7.21):

$$\frac{d\mathbb{E}(C)}{dl} = a_c d + \frac{\sigma}{2\sqrt{l+1}}(a+a_c+a_w)\varphi\circ\Phi^{-1}\left(\frac{a}{a+a_c+a_w}\right).$$

In order to clarify notations, the following function is introduced:

$$\zeta : (x, y) \mapsto (x + y)\phi \circ \Phi^{-1}\left(\frac{x}{x + y}\right), \tag{7.23}$$

so that:

$$(a + a_c + a_w)\varphi \circ \Phi^{-1}\left(\frac{a}{a + a_c + a_w}\right) = \zeta(a, a_c + a_w).$$

ζ has the following properties:

Lemma 7.2 *ζ is positive and increasing in x and y. Besides, $\zeta(x, y) = \zeta(y, x)$.*

Proof: See Appendix D.1.

In the following, $\zeta(a, a_c + a_w)$ is simply written ζ.

Using these notations, the derivative of the full logistic cost with respect to the transport lead time can be written:

$$\frac{d\mathbb{E}(C)}{dl} = a_c d + \frac{\zeta\sigma}{2\sqrt{l + 1}}. \tag{7.24}$$

When the transport lead time increases, the full logistic cost increases because the time the commodities spend in the transport operation increases, but not only. The second reason why the full logistic cost increases if the transport lead time is one day longer is that the sizes of shipments must be decided one day earlier, which means the variability of the destination inventory is greater. This greater variability implies a change in the safety stock, which is not a problem in itself since, due to the envelop theorem, such a shift leaves the full logistic cost unchanged, but it also means that $\mathbb{E}[(I_t)^+]$ and $\mathbb{E}[(I_t)^-]$ and, as a direct consequence, the inventory and customer costs, increase. Note that $\zeta(a, a_c + a_w)$ can be interpreted the marginal cost of variability. From Lemma 7.2, and as expected, it is an increasing function of each of its parameters.

The shipper's value of time can be defined as the cost increase the shipper would be ready to accept for the travel time to be decreased by one unit:

$$\alpha = \frac{d\mathbb{E}(C)}{dl} \bigg/ \frac{d\mathbb{E}(C)}{dc_t}.$$

Given the previous discussion:

Proposition 7.3 *the shipper's value of time is equal to:*

$$\alpha = a_c + \frac{\zeta}{2\sqrt{l+1}}\frac{\sigma}{d}. \tag{7.25}$$

As expected, α encompasses the commodity value of time, but it also encompasses another term, which can be interpreted as the value of speed in the frame of a supply chain.

Note that this value of time is not necessarily closely related to the carrier's value of time as it is observed in classic roadside surveys for example. Indeed, the transport lead time offered by the carrier can rely on a complex organisation, so that the linkage between the transport lead time and the actual transport operation durations is not trivial. A deeper analysis of this difference is found in Chapter 3.

Demand characteristics To determine the influence of d on the safety stock and the full logistic cost, first observe that as stated above, I_s^* does not depend on d. From Equations (7.14) and (7.15), this is also the case of μ_I and σ_I, so that the inventory and customer cost do not depend on d. As a consequence, a change in d only impacts the transport and pipeline inventory costs. From Equation (7.9), which gives the cost function, and Equation (7.17), according to which the expected shipment size is d:

$$\frac{d\mathbb{E}(C)}{dd} = c_t + (l+1)a_c. \tag{7.26}$$

Whereas d influences the transport and pipeline inventory costs, the demand variability σ impacts the inventory and customer costs. From Equation (7.18):

$$\frac{\partial\mathbb{E}\left[(I_t)^+\right]}{\partial\sigma} = \frac{\partial\sigma_I}{\partial\sigma}\varphi\left(\frac{\mu_I}{\sigma_I}\right),$$

and from Equation (7.15) $\partial\sigma_I/\partial\sigma = \sqrt{l+1}$ so that:

$$\frac{\partial\mathbb{E}\left[(I_t)^+\right]}{\partial\sigma} = \sqrt{l+1}\varphi\left(\frac{\sigma_I}{\mu_I}\right).$$

$\partial\mathbb{E}[(I_t)^-]/\partial\sigma$ is derived similarly. As a consequence:

$$\frac{d\mathbb{E}(C)}{d\sigma} = \zeta\sqrt{l+1}. \tag{7.27}$$

The full logistic cost is thus linear in d and σ. When d increases, the transport cost and pipeline inventory cost increase. In this case, the important cost parameters are the transport unit cost c_t, the commodity value of time a_c, and the transport lead time l, on which depends the pipeline inventory cost.

When σ increases, the inventory cost and the customer cost increase. The cost increase is equal to ζ, which can be interpreted, as above, as the marginal cost of standard deviation, times $\sqrt{l+1}$, which means that the full logistic cost is all the more responsive to a change in σ as the travel lead time is great. This is consistent with Equation (7.25), according to which the amount the shipper is ready to pay for a reduction of the transport lead time is proportional to σ.

7.2.4 Numerical applications

The model presented here is not intended to be used in an operational context, and has not been the object of econometric investigations. However, some numerical applications are presented here, to illustrate the model's behaviour and to provide orders of magnitude. The values given to the various parameters are purely indicative. Two parts of supply chains are considered: laptop computers, and cars.

Laptop computers Consider a shipper producing laptop computers in China and selling them in Europe[7].

Since a laptop weights approximatively 2 kg and is sold by and large 1000€, its value density is about 500 k€ per ton of commodity. If we assume an opportunity cost of capital of about 20% and a depreciation cost also equal to 20%, a_c is about 200 k€per ton per year, or 550 €per ton and per day.

The warehousing cost can be roughly approximated on the basis of average warehouse rents, about 50€ per year and m^2 in France. On the hypothesis that it is possible to stock about 3 tons of commodities per m^2, a_w is estimated at 0.05€ per ton and per day.

The customer cost is much trickier to estimate. According to many industry studies, 8% of retail items are out of stock at any one time (Su and Zhang, 2009). In our model, if the probability that I_t is negative is 8%, then $a/(a+a_c+a_w)$ is 0.92, so that a is approximately $11(a_c+a_w)$.

[7] In the framework presented above, the shipper produces the commodities and sells them in the retail center. However, the producer and the retailer of a given commodity are often distinct firms. In a perfectly competitive environment, and provided the firms communicate each other all the information they have, the results previously presented still hold. Of course, this is generally not the case, thus the strategic dimension of supply chain management.

In the absence of more accurate data, the following value is retained: $a = 6000€$ per day and per ton, which stands, in this case, for 12€ per laptop per day.

Estimates of the daily demand expected value and variance are unavailable. The following assumption is made: $d = 5$ units, or 0.01 ton, and $\sigma = 3$ units, or 0.006 ton. Given that the example is focused on a single computer brand, small numbers are reasonable.

Now assume the computers are transported by plane and by truck for the pickup and delivery movements. The transport lead time is assumed to be 5 days. The transport rate is assumed to be 2000€ per ton.

Given these parameters, the safety stock is (in units) $I_s^* = 10.1$ units. As a is much larger than $a_c + a_w$, we find as expected that the average amount of computers actually stocked at a given time, $\mathbb{E}[(I_t)^+] = 10.4$, is much larger than the average number of customers waiting at a given time $\mathbb{E}[(I_t)^-] = 0.3$. The per day cost components are:

$$\begin{aligned} C_t &= 20.0€ \text{ per day}, \\ C_p &= 27.5€ \text{ per day}, \\ C_d &= 11.4€ \text{ per day}, \\ C_c &= 3.4€ \text{ per day}, \end{aligned}$$

so that the direct transport costs constitute less than half of the full logistic cost, but the generalised transport cost (including the pipeline inventory cost) more than 75% of the full logistic cost. Finally, the shipper's value of time is $\alpha = 673.7€$ per day or 28.1€ per hour.

Now assume the computers are transported by sea. The transport rate is assumed to be approximately 100€ per ton, and the transport lead time approximately 45 days. The safety stock is then of 28.1 units. There are on average 28.8 computers stocked, and 0.8 customers waiting. The cost elements are:

$$\begin{aligned} C_t &= 1.0€ \text{ per day}, \\ C_p &= 247.5€ \text{ per day}, \\ C_d &= 31.7€ \text{ per day}, \\ C_c &= 9.4€ \text{ per day}, \end{aligned}$$

so that the full logistic cost is 289.6€ per day, for 62.3€ per day in the previous case, so that air transport is more competitive in this example, due to the pipeline inventory cost. Note that in this case, the value of time is 594.7€ per day, not very different from the previous one. The model variables are recapitulated in Table 7.1.

	Air	Sea
I_s^*	10.1	28.1
$\mathbb{E}[(I_t)^+]$	10.4	28.8
$\mathbb{E}[(I_t)^-]$	0.3	0.8
α	673.7	594.7
C_t	20.0	1.0
C_p	27.5	247.5
C_d	11.4	31.7
C_c	3.4	9.4
C	62.3	289.6

Table 7.1: Laptop supply chain example

Cars Consider now a shipper sending cars from a producing unit located in Europe to a central distribution platform also located in Europe, at about, from which the cars are then sent to the car dealers. Car dealers will be considered as the customers of the model, which is justified by the fact that they have no stock: if they wait for their deliveries, so do the final customers. We focus on a particular car model.

The following assumptions are taken. The cars are worth 25000€. Their weight is one ton (for simplicity). The capital cost is 15%, and the depreciation cost 15%, so that $a_c = 20$€ per day. The warehousing cost is $a_w = 0.05$€ per day. As above: $a = 8(a_c + a_w) = 160$ (the decision to buy a car is taken in a wider time scale, and other parameters such as the attachment to a particular brand, or the need of a particular car type, can mitigate the importance of the waiting time to the customer; anyway, all these assumptions are of course very coarse). The daily demand of the considered car model is $d = 10$, and the variance $\sigma = 6$.

Two transport modes will be compared. First, motor carriers: the lead time is 2 days, and the rate about 200€ per ton. Second, railroad transport: the lead time is 5 days, and the rate about 50€ per ton.

The model variables in the two cases are compared in Table 7.2. It appears again that the higher lead time of rail transport implies higher pipeline inventory costs, and higher costs due to the destination inventory variability. These effects do not compensate the price decrease, so that rail transport is competitive.

	Road	Rail
I_s^*	12.7	20.7
$\mathbb{E}[(I_t)^+]$	13.2	21.6
$\mathbb{E}[(I_t)^-]$	0.6	0.91
α	25.9	23.6
C_t	2000.0	500.0
C_p	400.0	1400.0
C_d	265.3	433.2
C_c	89.6	146.4
C	2755.0	2479.6

Table 7.2: Car supply chain example

7.2.5 Further discussion

It appears from the discussion above that the classic inventory under periodic review model constitutes a possible candidate for a microeconomic basis of shippers' behaviour. The role of the demand amount and variability are distinguished, and the transport and logistic decisions of the shipper are presented in a comprehensive framework. However, the approach followed up to this point has disregarded many concrete logistic problematics. Some of them are briefly discussed in this section.

Transport lead-time variability Assume the transport lead-time is not certain and given by l, but a random variable of mean l and variability σ_l. Once again, the problem will not be solved formally; we only provide a schematic approach of the question.

Intuitively, transport lead-time variability means that the shipments may arrive sooner, or later, than initially intended. The information available to the shipper at each period is identical, so that the order-up-to logistic policy is still justified. This tends to increase the destination inventory variance, by an amount proportional to the transport lead time variance σ_l^2 and to the average shipment size d.

Assume the transport lead time variations are independent from the demand variations. The destination inventory standard deviation, which is $\sqrt{l+1}\sigma$ when the transport lead time variance is zero, becomes:

$$\sigma_I = \sqrt{(l+1)\sigma^2 + d^2\sigma_l^2}.$$

In that case, the variation of the destination inventory standard deviation with σ_l is:

$$\frac{\partial\sigma_I}{\partial\sigma_l} = \frac{d^2\sigma_l}{\sigma_I},$$

so that the cost variation is:

$$\frac{d\mathbb{E}(C)}{d\sigma_l} = \frac{\zeta d^2\sigma_l}{\sqrt{(l+1)\sigma^2 + d^2\sigma_l^2}},$$

which can also be written as:

$$\frac{d\mathbb{E}(C)}{d\sigma_l} = \frac{\zeta d}{\sqrt{(\sqrt{l+1}\sigma/d\sigma_l)^2 + 1}}. \tag{7.28}$$

Similarly to the value of time, we can define the value of reliability α_ρ as the transport unit cost reduction that would compensate an increase of the transport lead time variability from the shipper's standpoint. Given Equations (7.22) and (7.28):

$$\alpha_\rho = \frac{\zeta}{\sqrt{(\sqrt{l+1}\sigma/d\sigma_l)^2 + 1}}. \tag{7.29}$$

Unsurprisingly, we find that α_ρ is proportional to the marginal cost of variability ζ. What is more interesting is the behaviour of α_ρ when $\sqrt{l+1}\sigma/d\sigma_l$ varies. This can be interpreted in terms of relative variances: Indeed, if $\sigma/d \gg \sigma_l/\sqrt{l+1}$, *i.e.* the relative demand variance is large with respect to the relative transport lead time variance, the shipper is not very sensitive to the transport lead time reliability. On the contrary, if $\sigma/d \ll \sigma_l/\sqrt{l+1}$, α_r increases up to its maximum value, ζ.

Perishable products The model is based on the hypothesis that the commodity considered can be stocked as long as desired. In some cases, this hypothese is irrelevant. However, the model has no simple solution if it is lifted. The case of perishable products is discussed below, without an analytical treatment.

Perishable products cannot be stocked too long. After a while, they loose part or all of their value. The main change in the model is that since some commodities are lost, the average shipment size is larger than d. The transport lead time l plays an even more important role in such a framework: a decrease in transport lead time, in addition to improving the system's reactivity and reducing the pipeline inventory cost, increases the time the perishable commodity can spend in the retail center before being obsolete, thus the probability it is bought before. Some complex effects can also occur: for example, a diminution of the transport unit cost c_t may allow the shipper to increase the amount of lost commodities, so that the demand for transport may increase even though the final demand is unchanged.

The backlogging hypothesis The model is also based on the hypothesis that all the customers wait until they are served, no matter the delay[8]. However, some customers may not wait. For example, a customer can go to a retail center to buy a commodity of a given brand. If this particular brand is not available at this time, the customer may come back, but he may also buy nothing and leave, or chose another brand.

This implies a lot of difficulties: first, the cost for a shipper of a customer who has such a behaviour is obviously significant, but hard to measure. Second, in the model, this means one should distinguish the amount of commodities requested from the amount of commodities actually served. Inventory models without backlogging, *i.e.* where the customers leave when they are not served, can be solved through dynamic programming, but they are very tedious to address analytically. In particular, an order-up-to policy can never be optimal. Some bounds on the optimal policy and cost are provided in Morton (1969). Furthermore, there is no simple expression for the safety stock and the expected cost and so on, which makes the economic interpretation challenging. Third, they are based on the assumption that all the customers leave as soon as their request is delayed. This corresponds only to a limited amount of case. In fact, the strategic use of product availability by the firms to attract demand, on the basis of the consumers' expectations, is a new and vivid stream of literature (for a review of these works, see e.g. Su and Zhang, 2009). Finally, if customers can choose another brand if the one they wanted is not available, then the economic incentives of the producer and the retail seller depart from one another. Indeed, for the producer, a stock shortage means lost sales, and maybe lost consumers. This is much less the case for the retail seller: if the customer opts for another brand, the dissatisfaction of the customer w.r.t. the retail seller is limited; and the turnover is unchanged.

It should be noted that fortunately, the inventory models with or

[8]This is the backlogging hypothesis, also referred to as the backorder hypothesis. The system where customers don't wait is also called the lost sales system. These two assumptions can be considered as the bounds defining the range in which all real situations take place, either near one of these two extremes, or somewhere in the middle.

without backlogging yield similar results when a is very large with respect to $a_c + a_w$. Indeed, in that case, the safety stock is large. As a consequence, the average waiting time of the customers is very low, which means that it is extremely improbable that a customer has to wait more than one time period. A customer which waits thus generally implies a cost of a, where a can be indifferently interpreted as the customer value of time or the cost for the shipper of losing a customer. In this particular case, an order-up-to logistic where a is replaced by $a + (l+1)(a_c + a_w)$ is proved to be asymptotically optimal (Huh et al., 2009).

Transport policies and warehousing demand The amount of commodity actually stocked at a given time is on average $\mathbb{E}[(I_t)^+]$. If warehouses can be used and left freely, this variable stands for the expected warehousing demand associated to the commodity flow considered.

From Equation (7.18), this demand increases with μ_I and σ_I. This means that it increases with σ, a, and l, whereas it decreases with a_c and, quite intuitively, a_w. c_t and d have no influence on the warehousing demand.

Whereas it is hardly conceivable to influence the majority of the model's parameters by a transport policy, the case of l is particular. Indeed, modifying travel times is the main objective of a significant share of transport policies and infrastructure projects. Although its outreach should not be overestimated, the model implies that these policies may have a significant effect on logistic land use. Besides, it stems from the model that the interaction between freight transport and logistic land use partly finds its source in the demand variability and the strategies shippers develop to address it.

7.2.6 Conclusion

This section is based on a very classic inventory theory model. As a consequence, no innovative theoretical results are presented. However, examining such a model from the microeconomic perspective of freight transport demand analysis yields new highlights on the behaviour of shippers.

The model allows indeed to explicit the transport and other logistic costs of a shipper for a given commodity type. The way the full logistic cost depends on the model parameters illustrates the linkage between the logistic constraints of the shipper and its preferences with respect to freight transport.

In particular, the shipper's value of time derives from its logistic imperatives. The shipper is ready to pay to reduce the transport lead time first because it reduces the time commodities spend in the transport operation, second because it improves the system's reactivity: a lower transport lead time allows the inventory to stick more closely to the demand, so that both the inventory and the customer costs are decreased.

As illustrated by the numerical applications, this framework can explain why a certain type of transport technology can be preferred to another, on the basis of its characteristics of price and lead time. However, it is not able to explain why, under certain circumstances, a shipper would use two different transport technologies jointly. This is the topic of the next section.

7.3 A model with two transport modes

The previous model has illustrated the advantages and disadvantages of the various available transport modes. A light transport mode is usually fast but expensive, but it improves the supply chain's reactivity. On the contrary, a heavy transport mode is less expensive but slower, so that the shipper loses some of its ability to limit the inventory and customer costs.

According to the previous model, when the shipper can choose among a set of transport modes, one of them is optimal, so that it is used exclusively. However, one observes that in some cases, some firms combine two modes in parallel on a given supply chain. Their objective is most probably to take profit of each mode's advantage: by moving a part of the commodities by the heavy mode, the overall transport cost is kept in reasonable limits, whereas the supply chain's reactivity is not affected too much, thanks to the light mode.

Extending the model of the previous section to the two modes case is not an easy task. A usual modeling strategy is to consider a classic mode and an emergency mode, the latter one with a zero transport lead time. This is formally equivalent to the lost-sale inventory model, with the same difficulty that there is no analytic solution and the other drawback that the emergency mode lead time is considered to be null. Considering an emergency mode with a positive lead time has been the object of a large number of papers, the precursor being Barankin (1961). These papers vary by the assumptions they make about the system: periodic or continuous review, vehicle capacity constraint, arbitrary lead times, fixed ordering costs, etc. In most cases, the objective of these studies is to derive a — sometimes approximate — optimal logistic policy.

They generally prove to be quite complicated, even for the approximated cases[9].

As a consequence, a new, very simple periodic inventory model with two transport modes is introduced in this section. The objective of this model is to provide a basis for the microeconomic analysis of freight mode choice. It is based on a heuristic logistic policy mixing the two transport modes which has two qualities: first, this policy is, under certain circumstances, more efficient than if any of the two available modes is used alone; second, using some approximations, it is analytically tractable. As a consequence, the decisions of the shipper can be analysed microeconomically, which is our main objective.

The outline of this section is identical to that of Section 7.2. After the framework modifications are introduced in Section 7.3.1, the model is presented in Section 7.3.2. The economic analysis of the results is provided in Section 7.3.3, and some numerical applications can be found in Section 7.3.4. Section 7.3.5 finally discusses some further points.

7.3.1 Framework

The framework is unchanged, except that the shipper may now use two transport modes, which are arbitrarily called the heavy and the light transport mode. The heavy (resp. light) transport lead time is denoted l_h (resp. l_l). The heavy (resp. light) unit transport cost is denoted c_h (resp. c_l). The heavy mode is assumed slower and less expensive than the light mode:

$$\begin{cases} l_h & > & l_l, \\ c_h & < & c_l. \end{cases} \tag{7.30}$$

Denote s_t^h (resp. s_t^l) the size of the shipment sent at time t by mode h (resp. l).

7.3.2 The model

In the case of a stochastic demand, the sizes of shipments are decided so as to compensate the inventory variations in the retail center. Intuitively,

[9]The closest to a microeconomically workable model is Moinzadeh and Nahmias (1988). The logistic policy is indeed quite simple. There are two re-order points, one for each mode. If the inventory falls below one of these values, a shipment of a given size is sent by the corresponding mode. However, this is a continuous review model, and not a periodic one. Furthermore, there are fixed order costs, which are not taken into account here. Finally, the assumption is made that there is never more than one shipment being carried by one mode at one time, which is not necessarily realistic.

if two modes are used simultaneously, with one fast and expensive and the other slower and cheaper, one expects that compensating the inventory variation will be the role of the faster mode. If the heavy mode is much slower than the light mode, it cannot be used adaptatively: it is reasonable to assume the amount send by this mode will be constant.

In the following, we assume the heavy mode is used as a base mode, carrying a constant amount of commodities, while the demand variations are addressed by the light mode, using an order-up-to logistic policy.

7.3.2.1 The logistic policy

Denote d_h the amount sent by the heavy mode at each period:

$$\forall t, s_t^h = d_h \tag{7.31}$$

In order to define the size of the shipments that should be sent by the light mode s_t^l, Equation (7.10) can be rewritten:

$$I_{t+l_l+1} = I_t - \sum_{i=t}^{t+l_l} D_i + \sum_{i=t-l_l}^{t-1} (s_i^h + s_i^l) + s_t^h + s_t^l,$$

which, given Equation (7.31), is equivalent to:

$$I_{t+l_l+1} = I_t - \sum_{i=t}^{t+l_l} (D_i - d_h) + \sum_{i=t-l_l}^{t-1} s_i^l + s_t^l.$$

It appears that once the heavy mode shipment size is decided, the light mode derives from an order-up-to logistic absolutely similar to the one presented in Section 7.2, except that D_t is replaced by $D_t - d_h$.

Let the remaining demand be denoted by D_t^l:

$$D_t^l = D_t - d_h.$$

The $\{D_t^l\}$ are i.i.d., and their common distribution is:

$$\begin{cases} \mathbb{P}\{D_t^l = d - d_h - \sigma\} &= 1/2, \\ \mathbb{P}\{D_t^l = d - d_h + \sigma\} &= 1/2; \end{cases}$$

their expected value and variance are respectively:

$$\begin{aligned} \mathbb{E}(D_t^l) &= d - d_h, \\ \mathbb{V}(D_t^l) &= \sigma^2. \end{aligned}$$

An important point should be noted here: D_t is never lower than $d-\sigma$. As a consequence, if d_h is comprised between 0 and $d-\sigma$, then D_t^l is always positive. In that case, it is easy to see that the optimal light mode shipment size is $s_t^l = s_t - d_h$, with s_t defined by Equation (7.11). I_s^*, defined by Equation (7.21), is unchanged. The expected full logistic cost function is thus linear in d_h over $[0; d-\sigma]$. As a consequence, the optimum d_h is either zero or higher than $d-\sigma$.

Consider now the case where $d_h > d-\sigma$. The remaining demand D_t^l has a probability 1/2 to be negative, which means that at time t, there have been more commodities delivered than actually ordered. In that case, Equation (7.11) cannot be applied as is. Indeed, if it were applied for all t, from Equation (7.16), it would lead mechanically to:

$$s_t^l = D_{t-1}^l.$$

This is clearly contradictory: from the physical nature of the system considered, s_t^l cannot be negative.

The order-up-to policy is thus a bit more complicated in this framework: We still assume the shipment s_t^l is determined by Equation (7.11) (with D_t^l replacing D_t), except that it must be positive. In other words, if the destination inventory level is too low, the shipper can send more commodities. On the contrary, the shipper cannot take freight back, the only option if the amount of commodities in the destination inventory is too large is to wait for it to decrease. The light mode shipment at time t is thus determined by:

$$s_t^l = \left(I_s + (l_l+1)(d-d_h) - I_t - \sum_{i=t-l_l}^{t-1} s_i^l \right)^+ .$$

This results in the following logistic policy:

Hypothesis 7.4 *The logistic policy of the shipper is the following: the heavy transport mode is used to send a fixed amount of commodities:*

$$s_t^h = d_h,$$

and the light transport mode is used to replenish the destination inventory up to a given level, if necessary:

$$s_t^l = \left(I_s + (l_l+1)(d-d_h) - I_t - \sum_{i=t-l_l}^{t-1} s_i^l \right)^+ . \tag{7.32}$$

This logistic policy is not intended to be optimal. It is the result of a simple, hopefully reasonable heuristic approach applied to a complex problem of operations research. However, even this simple logistic policy is hardly workable. The next section introduces some necessary technical steps.

7.3.2.2 Clarifying the system's dynamics

Presented as is, little can be said about this logistic policy. In particular, the expected value of the destination inventory and the expected time the customers wait are not easily derived. This is problematic, as these two values are of central importance to discuss the optimal values of d_h and I_s. Further analysis is necessary to understand the system's dynamics under this policy. This section presents some intermediary steps which will prove useful in the following.

In order to understand the system's dynamics, one should remember the main difficulty here, which was absent from Section 7.2.2: the commodities can be delivered too early, and it is not possible to take them back. This is problematic because the inventory level can be higher than needed.

In order to keep track of this phenomenon, a new variable is introduced:

Definition 7.5 *the excess inventory I_t^E is defined as:*

$$I_t^E = \left(I_s + (l_l + 1)(d - d_h) - I_t - \sum_{i=t-l_l}^{t-1} s_i^l \right)^- . \tag{7.33}$$

The light shipment size is subsequently:

$$s_t^l = I_s + (l_l + 1)(d - d_h) - I_t - \sum_{i=t-l_l}^{t-1} s_i^l + I_t^E. \tag{7.34}$$

The system dynamics is:

$$I_{t+1} = I_t + s_{t-l_l}^l - D_t^l, \tag{7.35}$$

so that Equation (7.10) remains valid (with D_t^l replacing D_t and s_t^l replacing s_t):

$$I_{t+l_l+1} = I_t - \sum_{i=t}^{t+l_l} D_i^l + \sum_{i=t-l_l}^{t-1} s_i^l + s_t^l.$$

By replacing s_t^l using Equation (7.34) we obtain:

$$I_{t+l_l+1} = I_s + \sum_{i=t}^{t+l_l}(d - d_h - D_i^l) + I_t^E,$$

which can be translated $l_l + 1$ time periods backwards.

Proposition 7.6 *The destination inventory I_t at time t is:*

$$I_t = I_s + \sum_{i=t-l_l-1}^{t-1}(d - D_i) + I_{t-l_l-1}^E. \tag{7.36}$$

This equation is very similar to Equation (7.12), except for the excess inventory term, which comes on top of the safety stock term and the variability term.

The evolution of the inventory is now clearer, but some investigation remains to be done to fully explicit the system's dynamics. In particular, it results straightforwardly from Equation (7.36) that:

$$I_{t+1} = I_t - D_t^l + D_{t-l_l-1}^l + \Delta I_{t-l_l-1}^E, \tag{7.37}$$

where $\Delta I_t^E = I_{t+1}^E - I_t^E$.

The comparison of Equations (7.35) and (7.37) yields:

$$s_t^l = D_{t-1}^l + \Delta I_{t-1}^E.$$

From Equations (7.32) and (7.33), if $I_t^E > 0$, s_t^l is necessarily zero. In that case:

$$I_t^E = I_{t-1}^E - D_{t-1}^l.$$

On the contrary, if $I_t^E = 0$, then $\Delta I_{t-1}^E = -I_{t-1}^E$. Furthermore, $s_t^l \geq 0$, so that $I_{t-1}^E \leq D_{t-1}^l$, and $I_t^E \geq I_{t-1}^E - D_{t-1}^l$.

It is now possible to describe the excess inventory dynamics in a greatly simplified manner:

Proposition 7.7 *The excess inventory dynamics is:*

$$\begin{array}{lcl} I_0^E & = & 0, \\ I_t^E & = & \left(I_{t-1}^E - D_{t-1}^l\right)^+. \end{array} \tag{7.38}$$

Propositions 7.6 and 7.7 provide a pretty clear vision of the evolution of the inventory. Note that these results do not depend on the the distribution of the demand D_t.

Before going into microeconomic considerations, it is now necessary to dispose of a clear vision of the excess inventory distribution, expected value and variance. Unfortunately, we could not answer these questions analytically. The next section is focused on a numerical analysis of the excess inventory behaviour.

7.3.2.3 Destination inventory's expected value and variance

The following notations are introduced for the expected values and variances of the destination inventory and the excess inventory:

$$\begin{array}{rcl} \mu_I & = & \mathbb{E}(I_t), \\ \sigma_I^2 & = & \mathbb{V}(I_t), \\ \mu_{I^E} & = & \mathbb{E}(I_t^E), \\ \sigma_{I^E}^2 & = & \mathbb{V}(I_t^E). \end{array}$$

The destination inventory level at time t stems from Equation (7.36). Given that the $\{D_i\}_{t-l_l-1 \leq i \leq t-1}$ are i.i.d. and that, from Equation (7.38), $I^E_{t-l_l-1}$ is itself a function of I_{t-l_l-2} and $D^l_{t-l_l-2}$, thus independent from the $\{D_i\}_{t-l_l-1 \leq i \leq t-1}$, the destination inventory level has the following expected value:

$$\mu_I = I_s + \mu_{I^E}, \tag{7.39}$$

and a variance of:

$$\sigma_I^2 = (t_t + 1)\sigma^2 + \sigma_{I^E}^2. \tag{7.40}$$

We could not manage to derive a closed formula of μ_{I^E} and σ_{I^E}, not to speak about its distribution. Simulation was used to derive approximations of these statistics as functions of d, d_h and σ.

To do so, we proceed in a series of steps. First, define X_t the following process:

$$\begin{array}{rcl} X_0 & = & 0, \\ X_{t+1} & = & \left\{ \begin{array}{ll} X_t + \delta & \text{with probability } 1/2, \\ (X_t - (1-\delta))^+ & \text{with probability } 1/2, \end{array} \right. \end{array} \tag{7.41}$$

with $\delta \in [0; 1/2[$.

Then the following formula constitutes a satisfying approximation of the expected value of the stationary distribution of X_t (see Appendix D.2 for the derivation of this result):

$$\mathbb{E}(X_t) = 0.30\delta + \frac{0.22\delta}{0.5 - \delta}. \tag{7.42}$$

The expected value increases more than proportionally with δ: linearly when δ is close to zero, then explosively when δ tends towards 1/2. A similar approximation is available for X_t's variance:

$$\mathbb{V}(X_t) = \left(0.43\delta + \frac{0.24\delta}{0.5 - \delta}\right)^2. \tag{7.43}$$

Now consider the demand specification given by Equation (7.1), together with the excess inventory dynamics given by Equation (7.38). For a heavy mode shipment size d_h, the excess inventory increases of $(d_h - d + \sigma)$ with probability 1/2 and decreases of $(d + \sigma - d_h)$ with probability 1/2. As the result of a simple comparison, $I_t^E/2\sigma$ has the same behaviour as X_t with $\delta = (d_h - d + \sigma)/2\sigma$.

On the basis of this discussion, the expected value and standard deviation of the stationary distribution of I_t^E can now be derived from Equations (7.42) and (7.43):

$$\mu_{I^E} = 0.30(d_h - d + \sigma) + \frac{0.44\sigma}{d - d_h}(d_h - d + \sigma), \tag{7.44}$$

and:

$$\sigma_{I^E} = 0.43(d_h - d + \sigma) + \frac{0.48\sigma}{d - d_h}(d_h - d + \sigma). \tag{7.45}$$

From the second equation, $\sigma_{I^E}(d_h, d, \sigma)$ is homogeneous of degree 1 (this is also true of μ_{I^E}, but this property is not useful in the following).

Besides, these equations can now be combined with Equations (7.39) and (7.40) to yield the expected value of the destination inventory as a function of I_s and d_h:

Approximation 7.8 *For a heavy mode shipment size d_h and a safety stock I_s, the destination inventory expected value is:*

$$\mu_I = I_s + 0.30(d_h - d + \sigma) + \frac{0.44\sigma}{d - d_h}(d_h - d + \sigma), \tag{7.46}$$

and its variance is:

$$\sigma_I^2 = (l_l + 1)\sigma^2 + \left(0.43 + \frac{0.48\sigma}{d - d_h}\right)^2 (d_h - d + \sigma)^2. \tag{7.47}$$

As a direct consequence, the partial derivatives of μ_I and σ_I with respect to I_s and d_h exist. Second, given Equation (7.47) and the fact that $\sigma_{I^E}(d_h, d, \sigma)$ is homogeneous of degree 1, $\sigma_I(d_h, d, \sigma, t_l)$ is also homogeneous of degree 1 w.r.t (d_h, d, σ).

This is the basis on which the optimal d_h and I_s will be derived.

7.3.2.4 Optimal safety stock and modal share

By and large, the optimal safety stock I_s and the optimal amount of commodities sent by the heavy mode d_h are derived as in Section 7.2.2, on the basis of the minimisation of the full logistic cost function given by Equation (7.8).

However, the method employed in Section 7.2.2 is not directly applicable. In particular, the transport cost and pipeline inventory cost terms should account for the two transport modes. Therefore, the full logistic cost function is:

$$\begin{aligned} C_t &= c_h s_t^h + c_l s_t^l + a_c \sum_{i=t-l_h}^{t-1} s_t^h + a_c \sum_{i=t-l_l}^{t-1} s_t^l \\ &\quad + (a_c + a_w)(I_t)^+ + a(I_t)^-. \end{aligned}$$

The expected value of this cost is:

$$\begin{aligned} \mathbb{E}(C) &= c_h d_h + c_l(d - d_h) + a_c l_h d_h + a_c l_l (d - d_h) \\ &\quad + (a_c + a_w)\mathbb{E}\left[(I_t)^+\right] + a\mathbb{E}\left[(I_t)^-\right], \end{aligned}$$

and can also be written:

$$\begin{aligned} \mathbb{E}(C) &= c_l d + d_h(c_h - c_l) + a_c\left(l_l d + d_h(l_h - l_l)\right) \\ &\quad + (a_c + a_w)\mathbb{E}\left[(I_t)^+\right] + a\mathbb{E}\left[(I_t)^-\right]. \end{aligned} \tag{7.48}$$

From the observation of this formula, it appears that d_h exerts a twofold influence. First, changing d_h changes the pipeline inventory cost: sending more commodities with the heavy mode implies a lower transport cost, since $c_h < c_l$ by hypothesis, and a higher pipeline inventory cost, since $t_h > t_l$ by hypothesis also. Second, changing d_h changes the inventory and customer costs. Let us focus on the first effect.

A necessary condition on the commodity value of time for the heavy mode to be used. Given the dynamics of the excess inventory in Proposition 7.7 and the daily demand distribution in Equation (7.1), the excess inventory is obviously always null if $d_h < d - \sigma$. In that case, the destination inventory level at time t, given by Equation (7.36), reduces to the single mode Equation (7.12), clearly independent from d_h. Subsequently, the expected full logistic cost is a linear function of d_h for $0 \leq d_h \leq d - \sigma$, as already observed above. Furthermore:

$$\forall d_h \in [0; d - \sigma], \ \frac{\partial \mathbb{E}(C)}{\partial d_h} = (c_h - c_l) + a_c(t_h - t_l).$$

In that case, it is most probable that if the commodity value of time a_c is higher than a given limit $\bar{a_c}$ defined as:

$$\bar{a_c} = \frac{c_l - c_h}{t_h - t_l}, \tag{7.49}$$

then the optimal d_h is zero. In the contrary case, $d_h \geq d - \sigma$. This result will be formally proved later on.

We get here a first intuitive result: if the commodity value of time is too high, the pipeline inventory cost alone is too high for the shipper to use the heavy mode.

First-order condition on the optimal safety stock. The optimal safety stock is determined by minimising the expected full logistic cost. Strictly speaking, the expected full logistic cost depends on the distribution of I_t. Unfortunately, there is no simple expression for this distribution. This is why we take now a drastic hypothesis, thanks to which we focus fully on the mean and variance of the destination inventory, and disregard the distribution shape:

Hypothesis 7.9 *The destination inventory I_t is assumed normally distributed, of mean μ_I and variance σ_I^2 given by Approximation 7.8.*

This approximation is not too unrealistic when d_h is close to $d - \sigma$. Indeed, the distribution of the excess inventory is vaguely exponential, and if d_h is close to $d - \sigma$, its variance is small compared to the variance of the sum of the D_t^l. From the central limit theorem, this sum can be assumed normally distributed. Finally, the sum of a normal and an exponential independent random variables still looks like a normal random variable, provided the variance of the exponential random variable is small. The approximation is less justified when d_h gets close to d.

Now, given Hypothesis 7.9, the first order condition on I_s is easily derived. From Equation (7.48), the transport cost and the pipeline inventory cost do not depend on I_s. The expected value of the destination inventory cost is proportional to $\mathbb{E}[(I_t)^+]$, which is given by Equation (7.18).

From Approximation 7.8, $\partial\mu_I/\partial I_s = 1$, and $\partial\sigma_I/\partial I_s = 0$. As a consequence, the reasoning presented in Section 7.2.2 leading to Equation (7.19) is still valid:

$$\frac{\partial \mathbb{E}(C)}{\partial I_s} = (a_c + a_w + a)\Phi\left(\frac{\mu_I}{\sigma_I}\right) - a.$$

The first order condition on the safety stock is thus:

$$(a_c + a_w + a)\Phi\left(\frac{\mu_I}{\sigma_I}\right) - a = 0. \tag{7.50}$$

Given that μ_I increases with I_s whereas σ_I remains unchanged, the equation above proves that $\mathbb{E}(C)$ is concave in I_s for a given d_h. Furthermore, given that $\partial\mathbb{E}(C)/\partial I_s$ increases strictly from $-a$ to $(a_c + a_w)$ on $\mathbb{R}$, there is a unique $I_s^*(d_h)$, function of d_h, verifying the above equality.

As a consequence, given Equation (7.39):

Lemma 7.10 *For a given d_h, there is a unique optimal safety stock, determined by:*

$$I_s^* = \sigma_I \Phi^{-1}\left(\frac{a}{a_c + a_w + a}\right) - \mu_{I^E}. \tag{7.51}$$

The optimal safety inventory can also be written as:

$$I_s^*(d_h) = \sqrt{(l_l+1)\sigma^2 + \sigma_{I^E}{}^2}.\Phi^{-1}\left(\frac{a}{a_c + a_w + a}\right) - \mu_{I^E},$$

so that the role of the excess inventory appears more clearly. The safety stock level depends on d_h, through σ_{I^E} and μ_{I^E}.

First-order condition on the heavy mode shipment size. Contrary to the previous case, the four cost components depend on d_h. Let us focus on the behaviour of $\mathbb{E}[(I_t)^+]$ and $\mathbb{E}[(I_t)^-]$. Remind, once again, Equation (7.18):

$$\mathbb{E}\left[(I_t)^+\right] = \int_{-\frac{\mu_I}{\sigma_I}}^{+\infty} (\sigma_I v + \mu_I)\varphi(v)\,dv.$$

Then:

$$\frac{\partial\mathbb{E}\left[(I_t)^+\right]}{\partial d_h} = \frac{\partial\sigma_I}{\partial d_h}\varphi\left(\frac{\mu_I}{\sigma_I}\right) + \frac{\partial\mu_I}{\partial d_h}\Phi\left(\frac{\mu_I}{\sigma_I}\right). \tag{7.52}$$

Similarly:

$$\frac{\partial\mathbb{E}\left[(I_t)^-\right]}{\partial d_h} = \frac{\partial\sigma_I}{\partial d_h}\varphi\left(\frac{\mu_I}{\sigma_I}\right) + \frac{\partial\mu_I}{\partial d_h}\left(\Phi\left(\frac{\mu_I}{\sigma_I}\right) - 1\right). \tag{7.53}$$

As a consequence, given Equations (7.48), (7.52) and (7.53), the expected full logistic cost derivative with respect to d_h is, for $d_h < d - \sigma$:

$$\frac{\partial\mathbb{E}(C)}{\partial d_h} = (c_h - c_l) + a_c(t_h - t_l), \tag{7.54}$$

and for $d_h > d - \sigma$:

$$\begin{aligned}\frac{\partial \mathbb{E}(C)}{\partial d_h} &= (c_h - c_l) + a_c(t_h - t_l) \\ &+(a_c + a_w + a)\frac{\partial \sigma_I}{\partial d_h}\varphi\left(\frac{\mu_I}{\sigma_I}\right) \\ &+\left((a_c + a_w + a)\Phi\left(\frac{\mu_I}{\sigma_I}\right) - a\right)\frac{\partial \mu_I}{\partial d_h}. \qquad (7.55)\end{aligned}$$

From Proposition 7.8, both the partial derivatives of μ_I and σ_I w.r.t. d_h increase indefinetely when d_h tends towards d. Combined with the fact, easily proved, that μ_I/σ_I tends towards a finite value, we observe that for any I_s, there exists at least one $d_h \in [d - \sigma; d[$ minimising the expected full logistic cost. However, at this stage, there is little we can say on the unicity of the optimal d_h for a given I_s.

Properties of the expected full logistic cost. The previous results are now combined to derive the optimal logistic policy. Let us introduce the following notation for the expected full logistic for a heavy modal shipment size d_h and a safety stock $I_s^*(d_h)$:

$$\widetilde{C}(d_h) = \mathbb{E}(C)(I_s^*(d_h), d_h).$$

The derivative of $\widetilde{C}$ w.r.t. d_h is:

$$\widetilde{C}'(d_h) = \frac{\partial \mathbb{E}(C)}{\partial I_s}\frac{\mathrm{d}I_s^*}{\mathrm{d}d_h} + \frac{\partial \mathbb{E}(C)}{\partial d_h}.$$

Given the optimality of $I_s^*(d_h)$ we obtain (as a straightforward application of the envelop theorem):

$$\widetilde{C}'(d_h) = \left.\frac{\partial \mathbb{E}(C)}{\partial d_h}\right|_{I_s = I_s^*(d_h)}$$

Using Equation (7.50), which gives the optimal μ_I/σ_I ratio, and Equations (7.54), and (7.55), which give the partial derivative of $\mathbb{E}(C)$ w.r.t. d_h for any value of I_s, $\widetilde{C}'$ becomes, for $d < d_h$:

$$\widetilde{C}'(d_h) = (c_h - c_l) + a_c(l_h - l_l),$$

and for $d_h > d - \sigma$:

$$\widetilde{C}'(d_h) = (c_h - c_l) + a_c(l_h - l_l) + \zeta\frac{\partial \sigma_I}{\partial d_h},$$

where ζ is defined by Equation (7.23).

Note that, σ_I is differentiable in $d_h = d - \sigma$, even though σ_{I^E} is not. Indeed, from Equation (7.40) we have:

$$\frac{\partial \sigma_I}{\partial d_h} = \frac{\sigma_{I^E}}{\sigma_I} \frac{\partial \sigma_{I^E}}{\partial d_h}, \tag{7.56}$$

and $\sigma_{I^E} = 0$ when $d_h = d - \sigma$. This ensures the left and right derivatives of $\widetilde{C}$ in $d - \sigma$ exist and are equal, which implies that $\widetilde{C}$ is differentiable everywhere.

The convexity of $\widetilde{C}$ stems from the convexity of σ_I. Indeed, from Equation (7.40), σ_I can be written as follows:

$$\sigma_I = \sqrt{K + \sigma_{I^E}^2},$$

with K a positive constant. It is clear from Equation (7.45) that σ_{I^E} is a convex function of d_h. Besides, $x \mapsto \sqrt{K + x^2}$ is an increasing and convex function, yielding the claimed result. There exists subsequently a unique d_h minimising $\widetilde{C}$.

The optimal logistic policy Given the previous discussion, two cases should be distinguished. First, the derivative of the full logistic cost is positive at zero: due to its convexity, $\widetilde{C}$ is an increasing function of d_h on $[0; d[$, yielding that the optimal heavy mode shipment size is zero:

$$a_c \geq \bar{a}_c \;\Rightarrow\; d_h^* = 0.$$

(in the specific case where $a_c = \bar{a}_c$, the shipper is indifferent between all $d_h \in [0, d - \sigma[$; without loss of generality we assume $d_h = 0$).

We find again, as discussed above, that in this case, the commodity value of time is so high that the pipeline inventory costs are sufficient to deter the shipper from using the heavy mode. The other parameters are not even taken into account.

Second case, if this condition does not hold, then the optimal heavy mode share d_h^* is larger than $d - \sigma$ and necessarily:

$$\zeta \frac{\partial \sigma_I}{\partial d_h} = (c_l - c_h) + a_c(l_l - l_h).$$

This equation determines uniquely d_h^*, thus I_s^*. Besides, σ_I is homogeneous of degree 1 w.r.t. (d_h, d, σ), so that $\partial \sigma_I / \partial d_h$ is homogeneous of degree 0 with respect to the same variables. $\partial \sigma_I / \partial d_h$ can therefore be considered as a function of d_h/d, σ/d, and l_l.

In order to clarify the notations, define ξ as:

$$\xi : \left(\frac{d_h}{d}, \frac{\sigma}{d}, l_l\right) \mapsto \frac{\partial \sigma_I}{\partial d_h}\left(\frac{d_h}{d}, \frac{\sigma}{d}, l_l\right). \tag{7.57}$$

Its partial derivatives are denoted ξ_d, ξ_σ and ξ_{t_l}.

A close formula for ξ can be derived from Equation (7.56) and by derivating σ_{I^E} in Equation (7.45). It is however quite complex, and provides no possibility to write explicitly the optimal d_h as a function of the model parameters.

These results sum up as follows:

Proposition 7.11 *In the case where two transport modes are used simultaneously, and given all the previously introduced hypotheses, two cases arise. If the commodity value of time is higher than a given threshold value $\bar{a}_c$, then $d_h^* = 0$ and the optimal safety stock is the classical one-mode optimal safety stock.*

Otherwise, the optimal heavy mode share is uniquely determined by the following equation:

$$\xi\left(\frac{d_h^*}{d}, \frac{\sigma}{d}, l_l\right)\zeta = (c_l - c_h) + a_c(l_l - l_h), \tag{7.58}$$

and the optimal safety stock is :

$$I_s^* = \sigma_I(d_h^*)\Phi^{-1}\left(\frac{a}{a_c + a_w + a}\right) - \mu_{I^E}(d_h^*). \tag{7.59}$$

In the bimodal case, the optimal safety stock is decided as in the one case mode: given the variability σ_I of the inventory, I_s is chosen so that μ_I is a trade off between the inventory and the customer costs.

Concerning the optimal heavy mode shipment size, the left hand side of Equation (7.58) is the marginal increase $\partial\sigma_I/\partial d_h$ of the inventory variability times the marginal cost ζ of a variability increase. Its right hand side is the marginal cost of transferring a commodity unit from the heavy mode to the light mode. We find that the optimal heavy mode shipment size is such that the marginal cost due to the increased inventory variability matches the marginal gain of transport and pipeline inventory costs.

The specific role of the commodity value of time has already be noted. If it is too high, the pipeline inventory cost is sufficient for the heavy mode to be ruled out; the optimal heavy mode shipment size is thus zero. It should also be noted that, in some cases, it can be more competitive for the shipper to use the heavy mode alone than to mix the two modes along the above policy, even for an optimal d_h. This is a limit of the model.

Nevertheless, this framework explains why a shipper would use two transport modes simultaneously for a given commodity flow. The modal share is chosen so as to find a trade off between the transport cost savings granted by the heavy mode and the logistic cost savings deriving from the light mode speed, which allows the system to be more reactive. Each transport mode has its own role, and these roles are not symmetric. Indeed, as it will be explained in detail in the next section, the sensitivity of the shipper to the transport mode characteristics is not the same for the heavy and the light mode.

7.3.3 Transport demand analysis

Once determined the optimal variables of the two modes logistic policy, the next stage of the analysis focuses on how these parameters, as well as the full logistic cost, depend on the model parameters. As noted above, it may happen that it is optimal for the shipper to use the heavy mode alone, even if $d_h^* > 0$. In the following, we assume this is not the case.

We first analyse the characteristics of the function ξ introduced in Equation (7.57). Then are analysed respectively the influence of the transport mode characteristics, the logistic parameters, and the demand parameters.

7.3.3.1 Preliminary calculations

The objective of this section is to derive some properties of function ξ, in order to facilitate the interpretation of the model, and to provide approximates of the optimal heavy mode share d_h^*.

Monotonicity of ξ The linkage between d_h^* and the model parameters is closely related to the monotonicity of ξ:

Lemma 7.12 *ξ increases with d_h/d and with σ/d, and decreases with l_l.*

Proof: see Appendix D.3.1.

This result is relatively intuitive: given that the excess inventory expected value and variance increase more than proportionately with d_h, one expects the derivative of the standard deviation to increase with d_h/d. The behaviour of ξ *w.r.t.* σ/d is not much less intuitive. To see it, it is sufficient to consider that a decrease in σ and an increase in d_h

are somewhat symmetric regarding their consequences on the behaviour of the excess inventory.

Approximation of ξ for d_h/d close to $1-\sigma/d$, implications on the optimal heavy mode share In this paragraph, we first provide a Taylor approximation of degree one of $\xi(d_h/d, \sigma/d, l_l)$ along d_h/d, then we calculate the optimal heavy mode share when this approximation is valid.

Lemma 7.13 *the Taylor approximation of degree one of ξ at $d_h/d = 1-\sigma/d$ is:*

$$\xi\left(\frac{d_h}{d}, \frac{\sigma}{d}, l_l\right) = \frac{0.83}{\sqrt{l_l+1}}\frac{d}{\sigma}\left(\frac{d_h - d + \sigma}{d}\right) + o\left(\frac{d_h - d + \sigma}{d}\right). \quad (7.60)$$

Proof: see Appendix D.3.2.

This approximation can be used in Equation (7.58) if the gain from using the heavy mode is small with respect to the marginal cost of variability: $(c_l - c_h) + a_c(l_l - l_h) \ll \zeta$. In that case, Equation (7.58) becomes:

$$\frac{0.83}{\sqrt{l_l+1}}\left(\frac{d_h^*}{\sigma} - \frac{d}{\sigma} + 1\right) = \frac{(c_l - c_h) + a_c(l_l - l_h)}{\zeta},$$

so that:

Proposition 7.14 *When $(c_l - c_h) + a_c(l_l - l_h)$ is positive and much smaller than ζ, the optimal heavy mode shipment size is approximately given by:*

$$d_h^* = d - \sigma + \frac{1.2\sigma\sqrt{l_l+1}\,((c_l - c_h) + a_c(l_l - l_h))}{\zeta}. \quad (7.61)$$

Approximation of ξ for d_h/d close to 1 Symmetrically, ξ can be simplified for high values of d_h.

Lemma 7.15 *When d_h/d is close to 1, $\xi(d_h/d, \sigma/d, l_l)$ is equivalent to:*

$$\xi\left(\frac{d_h}{d}, \frac{\sigma}{d}, l_l\right) \sim \frac{0.48\sigma^2}{(d-d_h)^2}.$$

Proof: see Appendix D.3.3.

As a consequence, the optimal heavy mode shipment size can be approximated in the following way:

Proposition 7.16 *When* $(c_l - c_h) + a_c(l_l - l_h)$ *is positive and much larger than* ζ*, the optimal heavy mode shipment size is approximately given by:*

$$d_h^* = d - 0.69\sigma\sqrt{\frac{\zeta}{(c_l - c_h) + a_c(l_l - l_h)}}. \tag{7.62}$$

The two approximations behave similarly with the model parameters. These parameters can be grouped into three categories: the characteristics of the two transport modes, the logistic costs, and the demand characteristics. Their influence will now be examined.

7.3.3.2 Influence of the transport mode characteristics

The transport modes are described by their unit costs c_h and c_l and their lead times l_h and l_l.

First of all, two cases can already been distinguished: if $a_c \geq \bar{a}_c$ (where $\bar{a}_c$ has been defined by Equation (7.49)), the shipper does not use the heavy mode at all, due to the excessive pipeline inventory cost. On the contrary, if $a_c < \bar{a}_c$, the heavy mode is competitive. It is more profitable for the shipper to combine the two modes or to use the heavy mode than to use the light mode alone.

From Lemma 7.12, ξ is an increasing function of d_h/d. The RHS of Equation (7.58)'s increases with c_l and l_l, and decreases with c_h and l_h, while its LHS does not depend on c_l, c_h, and l_h, and decreases with l_l. As a consequence, the heavy mode share increases with c_l and l_l, and decreases with l_h and c_h, which is quite intuitive: when one mode gets more valuable than the other, its modal share is modified consequently.

The expected value of the destination inventory, $\mu_I = I_s^* + \mu_{I^E}$, is proportional to its standard deviation σ_I. Its behaviour with respect to the cost parameters is thus straightforward: the expected destination inventory level increases when the heavy mode share increases, and decreases when the heavy mode share decreases.

Value of time Let α_h and α_l be the values of time of the heavy and the light mode, respectively:

$$\alpha_i = \frac{dC}{dl_i} \bigg/ \frac{dC}{dc_i} \quad \text{with } i \in \{h, l\},$$

where C is evaluated at the optimum.

From Equation (7.48), and given the envelop theorem:

$$\frac{dC}{dc_h} = d_h, \tag{7.63}$$

and:

$$\frac{dC}{dc_l} = d - d_h. \tag{7.64}$$

The derivatives of the expected full logistic cost w.r.t. the unit transport costs are equal to the average daily amounts of goods carried by the corresponding modes.

As regards the transport lead times, the derivatives are:

$$\frac{dC}{dl_h} = a_c d_h, \tag{7.65}$$

and:

$$\frac{dC}{dl_l} = a_c(d - d_h) + \frac{\partial}{\partial l_l}\left((a_c + a_w)\mathbb{E}[(I_t)^+] + a\mathbb{E}[(I_t)^-]\right),$$

which, from Equation (7.18), is equivalent to:

$$\begin{aligned}\frac{dC}{dl_l} \quad &= \quad a_c(d - d_h) + \frac{\partial \sigma_I}{\partial l_l}\left((a_c + a_w)\int_{-\frac{\mu_I}{\sigma_I}}^{+\infty} v\varphi(v)dv\right.\\ &\left. -a\int_{-\infty}^{-\frac{\mu_I}{\sigma_I}} v\varphi(v)dv\right),\end{aligned}$$

(from Equation (7.50), the $\partial\mu_I/\partial l_l$ term disappears.) Therefore:

$$\frac{dC}{dl_l} = a_c(d - d_h) + \zeta\frac{\partial \sigma_I}{\partial l_l}.$$

Finally, note that:

$$\frac{\partial \sigma_I}{\partial l_l} = \frac{\sigma}{2\sqrt{l_l + 1}}\frac{\sigma_{I^E}}{\sigma_I},$$

so that:

$$\frac{dC}{dl_l} = a_c(d - d_h) + \frac{\sigma}{2\sqrt{l_l + 1}}\frac{\sigma_{I^E}}{\sigma_I}\zeta, \tag{7.66}$$

the transport cost is all the more sensitive to an increase of l_l as the light transport mode is fast (*i.e.* as l_l is small), and as the variability of the final demand is large.

The values of time are then:

Proposition 7.17 *The heavy and light mode values of time are:*

$$\alpha_h = a_c, \tag{7.67}$$

and

$$\alpha_l = a_c + \frac{\sigma}{2\sqrt{l_l+1}} \frac{\sigma_{I^E}}{\sigma_I} \frac{\zeta}{d-d_h}. \tag{7.68}$$

By construction, the two modes have asymmetric roles. This appears clearly when their values of time are compared.

The asymmetry of the two modes roles appears clearly through the comparison of their respective values of time. The benefit the shipper withdraws from a reduction of the heavy mode lead time is a reduction of the pipeline inventory cost. The benefit for the shipper from a reduction of the light mode lead time also encompasses the reduction of the pipeline inventory cost, but not only. Indeed, a reduction of the light mode lead time allows the shipper to improve the reactivity of its supply chain, so that the expected inventory and customer costs are reduced. As a consequence, $\alpha_l > \alpha_h$.

Furthermore, the light mode value of time decreases with l_l: the value of the light mode's flexibility to the shipper directly depends on its speed. If the light mode is fast, a small decline of its speed impacts strongly the full logistic cost. On the contrary, if the light mode is relatively slow, the full logistic cost is less sensitive to a travel time increase.

Substitution effects Contrary to the single mode model where the transport demand was fixed, the shipper can choose between two transport modes here. As a consequence, the modal share depends on the model parameters.

Due to the model specification, it is not possible to assess the ordinary transport demand elasticities. Indeed, the profit function has not been explicitly stated, so that the market size effects are out of this study's scope. The analysis must be limited to the conditional transport demand (*i.e.* the transport demand for a given daily demand d), which catches only the substitution effects (Oum et al., 1992). For the sake of brevity, only the partial derivatives are calculated thereafter. From them, the elasticities derive straightforwardly.

The conditional transport demands are given by Equation (7.58). As a consequence, the heavy mode conditional demand varies as follows with the transport costs (the light mode share varies identically, except for the sign):

$$\frac{dd_h^*}{dc_h} = -\frac{d}{\zeta\xi_d}, \tag{7.69}$$

which is, as expected, negative, and:

$$\frac{dd_h^*}{dc_l} = \frac{d}{\zeta\xi_d}, \tag{7.70}$$

which is positive. These formulae inspire two comments. First, the influences of the transport costs are symmetric. This was to be expected, as they only intervene in the transport cost component of the full logistic cost. Second, in both cases, the modal share varies more slowly when the marginal cost of variability, ζ, is larger: if the logistic costs are large, the transport costs have little influence on the modal share.

Proposition 7.18 *The derivatives of the conditional transport demands with respect to the travel times are:*

$$\frac{dd_h^*}{dl_h} = -\frac{a_c d}{\zeta\xi_d}, \tag{7.71}$$

which is negative and:

$$\frac{dd_h^*}{dl_l} = \frac{a_c d}{\zeta\xi_d} - \frac{\xi_{l_l} d}{\xi_d}, \tag{7.72}$$

which is positive and, given that $\xi_{l_l} < 0$, greater than the absolute value of dd_h^*/dl_h.

Just as above, we find that the two modes don't have the same role. From the observation of the substitution effects, we deduce that the asymmetry between the light and the heavy mode stems from their lead times and not their costs.

Indeed, we observe by comparing Equations (7.63) and (7.64), as well as Equations (7.69) and (7.70), that changes in c_h or c_l have similar impacts on the two modes. On the contrary, the comparison of Equations (7.65) and (7.66) tells the same thing as the comparison of Equations (7.71) and (7.72): while the influence of l_h on the cost and modal share is limited to a classic logic of generalised transport costs, such is not the case of l_l, since a change in the light mode travel time influences the reactivity of the supply chain, with significant impacts on all kinds of costs. A shipper using both transport modes is ready to pay more for an improvement of the light mode speed than for an improvement of the heavy mode speed.

7.3.3.3 Influence of the logistic costs

The second main group of parameters that influence the optimal logistic policies are the logistic costs a_c, a_w and a. Each of them plays its own role. The two first are related to commodities actually owned by the shipper: they incur both a capital cost (both during the transport and in the warehouse) and a warehousing cost. The last one is the customer cost, related to the time the customers may wait before being delivered.

The modal share depends obviously on these costs. Let us examine the substitution effects. From Equation (7.58), the influence of the warehousing cost is:

$$\frac{dd_h^*}{da_w} = -\frac{\zeta_x \xi}{\zeta \xi_d} d,$$

with ζ_x the first derivative of ζ with respect to x. Similarly, the influence of the customer cost is:

$$\frac{dd_h^*}{da} = -\frac{\zeta_y \xi}{\zeta \xi_d} d,$$

with ζ_y the first derivative of ζ with respect to y.

The influence of the commodity value of time is a bit different, because of the pipeline inventory cost:

$$\frac{dd_h^*}{da_c} = \frac{l_l - l_h - \zeta_x \xi}{\zeta \xi_d} d.$$

From Equation (7.30), $l_h > l_l$; from Lemma 7.15, $\xi_d > 0$, and, from Lemma 7.2, $\zeta_x > 0$ and $\zeta_y > 0$. Therefore, the three derivatives are negative: if any of the logistic cost parameters increases, the heavy mode gets less competitive compared to the light mode. The optimal heavy shipment size diminishes. This is all the truer when a_c increases, as the pipeline inventory increases, making the heavy mode, which is slower, even less competitive.

7.3.3.4 Influence of the demand characteristics

The influence of d on the model is trivial: everything is proportional to d, so that there is little to comment on, apart from the lack of returns to scale in the model with respect to the flow intensity (provided σ increases proportionately to d). This is mainly due to the absence of transport fixed costs.

The case of σ is much more interesting. To see it, assume $\sigma = 0$. Then the model is straightforward: either the light mode or the heavy mode is used, and it is used alone[10]. Without variability, the only determinant of the modal choice is Equation (7.49), and the only relevant parameters are the transport modes characteristics and the commodity value of time.

In other words, the demand variability is a seemingly strong reason of why one shipper would use two modes simultaneously, a phenomenon which is not easily explained otherwise. The importance of σ in the model is not disclaimed by the other results: σ has a direct influence on the logistic costs, since the final inventory variability increases with it. Besides, one expects that an increased σ means a lower heavy mode share. This is confirmed by Equation (7.30), from which:

$$\frac{dd_h^*}{d\sigma} = -\frac{\xi_\sigma}{\xi_d},$$

which is negative, from Lemma 7.12.

7.3.4 Numerical applications

Let us come back to the two examples introduced in Subsection 7.2.4. In each case, the possibility that the shipper uses two modes simultaneously, and, should the case arise, the optimal modal share are investigated. These results come with a graphical analysis of the variation of the various costs with the heavy mode shipment size.

Laptop computers Take the same hypotheses as in Section 7.2.4. Consider the possibility that the shipper uses air and sea transport simultaneously. In this specific case, the result is straightforward: the maximum commodity value of time for the sea mode to be competitive is $\bar{a_c} = 47.5$. For the laptops, $a_c = 550$. The pipeline inventory cost of using the sea mode is so high that all the commodities are transported by air.

Figure 7.1 confirms this analysis. A larger heavy mode shipment size implies transport cost savings, but they are outrun by the pipeline inventory costs. Given that the inventory and customer costs also increase with d_h, the heavy mode is not used at the optimum.

[10] And here appears again one of the main weaknesses of the model: the model makes no sense for $d_h^* = d$. We would expect an order-up-to policy with the heavy transport mode used alone and a destination inventory variance of $(l_h + 1)\sigma^2$, we obtain a fixed shipment size policy and the variance of the destination inventory is infinite.

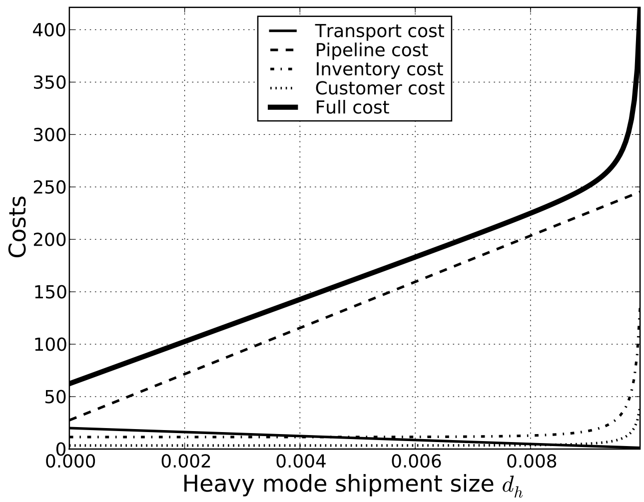

Figure 7.1: Cost components in the laptop case

Cars In the cars case, the limit commodity value of time is $\bar{a}_c = 33.3$ whereas $a_c = 16.5$. The pipeline inventory cost is not high enough to outrun by itself the heavy transport mode. We can expect the optimal heavy mode share to be positive.

Indeed, the full logistic cost minimisation yields $d_h^* = 7.47$, which means the heavy mode share is about 75% of the commodity flow. The economic trends underlying this result are illustrated by Figure 7.2: the heavy mode is more competitive than the light mode in terms of transport and pipeline inventory costs. However, beyond a given heavy mode share, the inventory and customer costs increase quickly, as the supply chain's reactivity worsens. The optimal heavy mode share is a trade off between transport costs and logistic costs.

As illustrated by Table 7.3, mixing the two modes is on the whole more efficient than using one of them alone. The values of time are also indicated, in the three configurations. Their comparison show, once again, the asymmetry in the role the two modes use when they are employed together. It is all the more visible in this case as these values of time are relatively similar for each mode used alone.

Table 7.4 gives the elasticities of the heavy mode shipment size with respect to the various model parameters. Their signs are as expected. The transport cost only has a small influence in that case. The transport lead times' influences are significantly higher, in particular as regards the light mode. This is consistent with the previous discussion about the

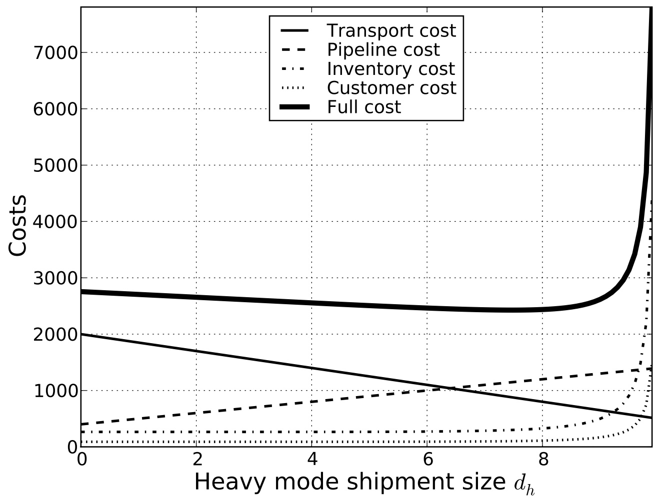

Figure 7.2: Cost components in the cars case

	Road	Rail	Mix
α_h	-	23.6	16.5
α_l	25.9	-	26.7
C_t	2000.0	500.0	878.4
C_p	400.0	1400.0	1147.7
C_d	196.8	433.2	299.7
C_c	89.6	146.4	101.3
C	2754.9	2479.6	2427.2

Table 7.3: Car supply chain with two modes, $l_h = 7$

respective roles of the transport modes when they are used together.

We immediately observe that, apart from the warehousing cost, which is so small that it has no influence, the logistic costs are the most critical parameters. Due to the fact that it impacts both the destination and pipeline inventory costs, the influence of the commodity value of time is even greater. Finally, the elasticity of d_h^* with respect to d is 1, as expected, given that everything is proportional to the flow rate in the model; and the influence of the demand variability on the heavy mode share is significant, but not as much as the logistic cost parameters.

However, the optimal mode share depends strongly on the model parameters. Assume indeed the heavy mode transport lead time is 5 days instead of 7. In that case, it is more competitive for the shipper

Variable	Elasticity
c_h	-0.12
c_l	0.12
l_h	-0.23
l_l	0.32
a_w	-0.00
a	-0.70
a_c	-1.11
d	1.00
σ	-0.25

Table 7.4: Heavy mode shipment size elasticities

to use the heavy mode alone than to apply the logistic policy described above (Table 7.5). Once again, this is a limit of the model.

	Road	Rail	Mix
C_t	2000.0	500.0	805.9
C_p	400.0	1000.0	877.6
C_d	196.8	375.2	324.2
C_c	89.6	126.8	109.5
C	2754.9	2002.0	2117.2

Table 7.5: Car supply chain with two modes, $l_h = 5$

7.3.5 Further discussion

The model introduced in this section is more complex than in the one mode case. The points addressed in Subsection 7.2.5 will therefore not be examined here. However, there is a question specific to the two modes case which should be raised: which determinants of modal choice have been disregarded? From our point of view, two of them deserve discussion: fixed costs, and the uncertainty of the economic environment.

Fixed costs In practise, it seems that few firms use two modes such as road and rail simultaneously for the supply chain of a given commodity. One of their common characteristics is that they are big firms. This is consistent with the fact that having access to a transport mode may imply fixed costs.

Whereas the fixed costs associated to road transport are relatively small, the ones associated to rail transport can be much larger. The shipper needs an access to the rail network, which is expensive. Small shippers which want to send commodities by train without having access to the rail network can resort to piggyback transport, but piggyback transport is much more expensive than rail transport.

The fixed costs cannot explain modal choices where they do not intervene, as for example the simultaneous use of sea and air transport, where the pickup and delivery movements are done by truck.

Note that the fixed costs just mentioned are in general sunk costs. This raises the question of the risk that the economic environment changes, and that seemingly profitable investments finally prove to be wasteful.

Economic environment uncertainty The fixed costs associated with a transport mode, in addition to be quite high, can be irretrievable investments, particularly in the case of rail transport[11]. The decision to make the investment thus relies on future traffic and costs forecast. But some important economic variables, such as the transport costs, are particularly hard to forecast. The recent sharp changes of the oil price illustrate this uncertainty.

Let us now come back to the example of the cars supply chain presented in Section 7.3.4. It seems from Figure 7.2 that the full logistic cost curve is very flat in the vicinity of d_h^*. The optimal heavy mode shipment size may thus be very sensitive to the values of the parameters. Under these circumstances, the decision to proceed to an onerous, irretrievable investment appears to be quite risky.

However, the elasticities of Table 7.4 shed a different light on this question. Indeed, one observes from the comparison of these various elasticities that d_h varies less with the transport costs than with the transport lead times, and even less than with the logistic costs.

The influence of the transport lead times and the logistic costs on d_h^* is more significant. But these parameters are not expected to vary a lot. The transport lead times depend basically on the organisation of carriers, on the infrastructure networks and on the authorised speeds, which are all quite stable. The logistic costs are even less prone to vary: the commodity value of time a_c is related to the shipper's cost of capital and the commodity's tendency to depreciation, which are both structuring parameters, and the customers value of time a is related to the customers

[11] The rail freight transport facilities market is certainly not as active as the corresponding market for road transport.

attitude towards waiting, a behavioural parameter than can be assumed fairly stable over time.

As a consequence, the heavy mode shipment size does not depend much on the transport costs. The risk that the necessary investment has been made in vain is thus limited in the example considered. Note that the full logistic cost depends much on the transport costs, but this is another point. More general investigations should be undertaken to see to what extent the discussion presented here is valid.

7.4 Conclusion

In this chapter, the simple supply chain of a given commodity is modeled as single product single location inventory system under periodic review, with backlogging. The demand for this commodity is assumed stochastic, but with a given, known, distribution. Four types of costs are considered: the direct transport costs, the pipeline inventory costs, due to the time the commodities spend in transport, the destination inventory costs, due to the warehousing as well as the inventory holding costs, and the customer cost, corresponding to the disutility of waiting time for customers.

The optimal logistic policy, when there is only one mode available, is the simple order-up-to logistic policy, associated to a given safety stock. This safety stock is chosen as the result of a trade off between inventory holding and warehousing costs. It increases with the demand variability and with the transport lead time. This implies that the value of time of the shipper, *i.e.* the amount the shipper is ready to pay for an improvement of the transport lead time, does not only stand for the pipeline inventory cost reduction. This is why this value of time increases with the marginal cost of the destination inventory standard deviation, and with the demand's standard deviation. An improvement of the transport lead time means a higher reactivity of the supply chain, which is valuable from the shipper's perspective. It is possible to derive from this model the shipper's preferences on the available transport modes[12], but not to explain why a shipper would use two modes together.

[12] These preferences are similar to the ones derived in Baumol and Vinod (1970), but derive from a different framework. Indeed, the crucial hypothesis in Baumol and Vinod (1970) is the existence of fixed shipment costs. In this study, it is assumed the carriers move a considerable number of distinct commodity types each period, so that the shipment costs, due to handling, loading and unloading operations, are assumed negligible for a given commodity type. However, contrary to Baumol and Vinod (1970), the crucial hypothesis in this study, is the stochasticity of the demand.

To address this deficiency, a logistic policy with two transport modes is then introduced. The shipper can use simultaneously a heavy mode assumed slow and not expensive, and a light mode, fast and onerous. In order to keep the results analytically tractable, the logistic policy is kept simple. In some cases, it is more profitable for the shipper to use one of the two modes alone. This is the case, in particular, when the pipeline inventory cost associated to the heavy transport mode is too large. In other cases, however, it is more profitable for the shipper to follow this policy than to use one of the two modes available alone. The modelling objective is thus reached.

By construction, the roles of the two modes are asymmetric. The heavy mode is used on a regular basis, in order to grant the shipper with transport cost savings. The light, more expensive, mode allows the shipper to keep the destination inventory close to the demand. Subsequently, the benefit for the shipper from a reduction of the heavy transport lead time is simply a reduction of the pipeline inventory cost, whereas a reduction of the light mode transport lead time is both a reduction of the pipeline inventory cost and an improvement of the supply chain's reactivity, which implies a reduction of the destination inventory and customer costs.

The approach used in this chapter illustrates the possibility, by using inventory theory models, to represent explicitly the logistic imperatives of shippers, and their influences on their transport decisions. Among these imperatives, the variability of the demand plays a central role, as well as the warehousing costs, the capital opportunity and depreciation costs, and the customers preferences with respect to waiting time.

From the standpoint of transport economics, our approach yields that transport modes are complementary inputs for a supply chain, but this complementarity may be weak. For most shippers, it is more interesting to use one mode alone. For some shippers though, using two modes simultaneously can be profitable. Interestingly, in that case, the optimal modal share seems to vary more with the logistic parameters quoted above than with the transport modes characteristics. Concerning the transport modes characteristics, the transport lead time seems to have more influence than the transport costs on the choices of the shippers (given the characteristics of the supply chain under study). Given that the transport costs are the only economic variables that can be significantly influenced by transport policies, this could explain why shippers' behaviours are not very sensitive to transport policies targeting transport costs.

Conclusion

This work is purported to provide a microeconomic analysis of the wide issue of the choice of shipment size in freight transport. A set of selected issues have been addressed, yielding specific answers to specific questions. In addition to these results of a somewhat limited generality, more general conclusions can be drawn. These conclusions are outlined here.

Representation of the freight transport system

The classical four stages representation of the freight transportation system suffers strong limitations (Chapter 3), which recent modelling advances have tried to address (Chapter 1). Representing explicitly shipping decisions clarifies a series of points which remain otherwise ambiguous. First of all, the notion of shipment can be considered from a decisional standpoint as the atom of the freight transport system (Chapter 3). Second, it allows for a clear distinction between shippers and carriers. On one hand, shippers consume transport services, *i.e.* door-to-door transport operations concerning clearly defined sets of commodities, under clearly defined conditions of price, transport duration, delivery time reliability, etc, without reference to transport modes or, more generally, to the way the transport operation is concretely achieved. On the other hand, carriers produce door-to-door transport services using an often complex combination of various transport modes and other logistic assets such as warehouses, cross-dock platforms, etc. Door-to-door shipment transport services constitute the object of the freight transport market.

This clarification once done, it is easier to distinguish the preferences of shippers and carriers respectively. Quite obviously, the objective of carriers is to provide transport operations of given characteristics to the lowest cost. Therefore, carriers combine at best the resources at their disposal to provide cost-efficient transport services. This combination process is defined here as the carriers logistic protocol (Chapter 3).

Shippers preferences are a bit less straightforward to derive. Shippers are usually firms selling goods or services to customers. Customers draw utility from goods or services once they have them at their disposal (Chapter 2). Therefore, they dislike a longer delivery time, or a higher risk of stock shortage, etc. Shippers thus try to organise their commodity flows so as to find a trade off between the satisfaction of their customers and the corresponding transport and inventory costs; this process is defined as the shippers logistic protocol (Chapter 3).

Some models have identified a difference between production-consumption (PC) flows and origin-destination (OD) flows in the freight transport system, and have tried to address this distinction explicitly (Chapter 1). The discussion above shows that the logistic stages (e.g. transshipments) which occur between the production location and consumption location of a given commodity flow can result from logistic decisions of shippers or from logistic decisions of carriers, and that these two types of decisions are not underlied by the same microeconomic logic. Modelling them without distinction would be economically unsatisfactory, as the distinction between freight transport demand and supply would be lost.

Freight transport costs

The representation of freight transport costs keeps often simple in spatialised freight transport demand models. Costs are generally assumed proportional to the amount of commodities carried. In fact, their structure is much more complex, as it appears instantly when shipments are explicitly taken into account. Three remarks can be done about this issue.

First, freight carriers extensively use indivisible resources to produce freight transport (e.g. vehicles, platforms). As a consequence, freight transport costs are strongly characterised by economies of scale and scope, so that the costs of different transport operations are often closely related and closely depend on the demand. Notably, the spatial structure of costs and the spatial structure of the freight transport demand are strongly and non trivially related. For example, an unbalanced demand on a given link means asymmetric marginal costs, particularly when there is empty running in one direction (Chapter 6).

Second, accounting for shipment size implies that transport costs must be examined more closely. In particular, some costs (such as haulage costs) are more dependent on shipment size than others (such as access costs). The capacity constraint also implies an interdependency between the marginal costs of carrying shipments of various sizes, depending on the demand (6).

Third, shippers and carriers should be distinguished. Indeed, as in passenger transportation, a transport operation requires inputs from carriers (e.g. driver, vehicle, fuel, etc.), and it also requires inputs from shippers, or user inputs, of which the costs are not borne by carriers. For example, from the perspective of a shipper, a higher travel time in a supply chain means a higher depreciation cost as well as a reduced supply chain reactivity. This explains the preference of shippers for shorter transport lead times (Chapter 7). Taking these user inputs into account yields important insights on the structure of costs: it can happen for example that the total logistic cost of shippers is endowed with economies of scale with respect to commodity flow rate, while this is not the case for carriers (Chapter 6).

Additional remarks on the constraints of motor carriers as well as on their various organisations, and on the methods to observe them are available in Chapter 4. They mainly concern the capacity constraint of heavy good vehicles, in weight and in volume, and specific organisations such as double crews or relays.

Microeconomic modelling of the choice of shipment size

It is necessary, in order to represent explicitly the shipment size variable in spatialised freight transport models, to describe microeconomically how shippers decide the size of a shipment. Models of choice of shipment size are classic in inventory theory. In these models, the choice of shipment size can result from a trade off between cycle inventory costs and transport costs (e.g. the Economic Order Quantity model), or between safety inventory costs and unsatisfied customers (e.g. the newsvendor model), or other combinations of these stocks.

There have been many refinements of these models, depending on the shipper's logistic constraints (notably interactions between transport and production, a point which is disregarded in this study, despite its considerable importance). However, these models are valid under sets of hypotheses which generally limit their outreach. From the perspective of modelling the demand of freight transport, which originates from a large, heterogeneous population of firms, there is no guarantee that one of these models outruns the others, and has the generality required to be included in a large scale, fully-fledged spatialised freight transport model.

As a consequence, designing or identifying a microeconomic shipment size model econometrically valid for all shippers, is a crucial step. This is generally difficult, as the variables on which even the simplest shipment size models are based remain generally unobserved, even in shipment databases. For example, the EOQ model requires that the rate of the commodity flow between the shipper and the receiver is observed, whereas most shipment databases disregard the shipper-receiver relationship.

Fortunately, the French database ECHO describes a set of 10,000 shipments with great accuracy; notably, the shipper, the receiver, and their relationship are described. Therefore, it has been possible to assess econometrically the EOQ model, which has been proved to perform quite correctly (Chapter 5). As a consequence, the EOQ model may constitute a good candidate for modelling the freight transport demand.

Microeconomic modelling of the choice of transport mode

Inventory models are immediately interesting from the perspective of freight transport modelling as they are based on an explicit representation of transport and other logistic costs. It is possible to derive from these models the preferences of shippers with respect to freight transport. In particular, it is possible to model microeconomically the impact of transport lead time and transport lead time reliability on the costs of shippers.

As a consequence, classic ratios of freight transport economics such as the value of travel time savings and the value of reliability are thus given an formal economic interpretation, with an explicit linkage to the logistic imperatives of shippers. It is also possible, by comparing the costs for distinct transport modes, to predict the mode choice of shippers. From this perspective, the EOQ model with distinct transport modes may also be a relevant candidate.

Besides, it is possible, through the use of more complex inventory models, to model the simultaneous use of two distinct transport modes by a given shipper for a given commodity flow is also possible. This point is of particular importance, as this phenomenon, which is sometimes observed in the freight transport system, is inconsistent with classical freight transport demand models.

This approach provides microeconomic insights on how a supply chain works. In particular, this approach illustrates how a faster but more expensive transport mode can be preferred by a shipper to a slower but less onerous transport mode, not because of the depreciation of commodities, but because the faster mode improves the reactivity of the supply chain, so that the additional transport costs are more than outrun by the decrease in the expected inventory levels and the expected number of unsatisfied customers.

This approach also illustrates how two modes can be combined to find a trade off between transport costs, inventory costs, and customer dissatisfaction costs. In both cases, it appears that the transport costs do not necessarily play a strong role in the mode choice by shippers; this indicates that inducing a significant modal shift through taxes might require extremely high tax levels. On the contrary, the mode share seems to be more sensitive to transport lead time and reliability (Chapter 7).

Using inventory theory in a microeconomic framework is not easy. In this study, strong hypotheses have been made so as to obtain usable results. Furthermore, the econometric assessment of such models is particularly demanding concerning data. But it constitutes a first step towards the microeconomic modelling of supply chains.

Equilibrium prices in the freight transport market

A freight transport model requires both a description of the demand and of the supply. When the shipment size variable is absent, transport alternatives are generally described by their prices, durations, and mode used. In a basic way, prices are assumed proportional to flow rates.

This simplification is not valid anymore when shipment size is taken into account. The linkage between transport price and shipment size must be made explicit, and analysed in detail: indeed, the joint shipment size and transport mode decision depends non trivially on the shape of price schedules. It is necessary, to model equilibrium prices, to introduce an additional level of detail in the description of the transport technology. For example, it appears that road transport price schedules are explained first by access costs, including the movement made by vehicles to reach shippers and receivers, and loading and unloading operations; second by haulage costs; third by the capacity constraint of vehicles (Chapter 6). Each of these three elements play an important role in explaining equilibrium freight transport prices.

It also appears the sizes of shipments influence the possibility carriers have to consolidate them. In other words, the transport demand and supply cannot be treated independently. This calls for an equilibrium model, where freight transport prices would be endogenous. However, such a task is intractable from a computation perspective, as it would require to calculate how carriers consolidate shipments in vehicles. Another approach is necessary. One could try to find a statistical linkage between freight transport costs and freight price schedules, then to represent the freight transport supply by these prices, and finally to induce the movements of vehicles used on the basis of average loading factors conditional to the size of shipments (such a data can be derived, for example, from the ECHO database). As much as possible, the explicit representation of the consolidation process should be avoided.

Information availability

The information agents have about their economic environments (*.i.e.* the future consequences of the decisions they take, the behaviours and objectives of their competitors or partners) plays a fundamental role in how the freight transport system works. As a consequence, information availaility should be given a central role, both in qualitative and formal analysis.

Qualitatively, the role of information in a supply chain is twofold (Chapter 2). First, agents must take operational decisions on the basis of imperfect forecasts of what they will be asked to deliver at given times. This is the case of carriers, which must decide fleet sizes, routes, prices, on the basis of a forecast demand; as well as shippers, which must anticipate their demand before taking production, stocking and transport decisions. Addressing uncertainty is at the core of the logistic function: if shippers were able to forecast their demand exactly, whatever the horizon, safety stocks would be useless, warehouses would be smaller, and faster transport modes would lose much of their attractiveness.

Second, many agents usually work together to produce and sell a given good to final consumers. These agents, considered as a whole, constitute a supply chain, inside which coordination is necessary to minimise the production, logistic and transport costs. However, the frontier is thin between cooperation and competition, and if members of a given supply chain have a strong incentive to share information so as to coordinate their decisions, they have in some cases an even stronger incentive to share as little information as they can, if they can withdraw a competitive advantage on their partners by doing so. This strategic dimension is crucial in understanding the freight transport system, and particularly the demand for freight transport.

Taking the information dimension into account in microeconomic models of freight transport demand and supply is not only possible, but also necessary. Technically, a lack of information is naturally taken into account by introducing stochasticity in models. Introducing stochasticity in freight transport models improves their realism: the freight transport system is constituted of a large amount of heterogeneous and uncoordinated agents. Trying to represent it as a centralised system is probably not relevant. It helps understanding how shippers can decide prices given the combinatorial nature of the problem of optimising the loading factor of their fleets (Chapter 6). It makes it possible to model microeconomically the preference of shippers for faster transport modes, which allow them to improve the reactivity of their supply chain (a demand shock is sooner addressed, and the implied inventory costs and customer dissatisfaction are reduced; Chapter 7).

Finally, taking into account the availability of information provides a strong rationale for explaining why some shippers and carriers sign long term contracts, while others don't. Indeed, a shipper with a good knowledge of its future flows is able to guarantee volumes to a motor carrier, for whom a regular transport demand is preferable to an irregular transport demand, and who in return offers preferential prices.

Shipment size in spatialised freight transport models

The four stage representation describes the freight transport system with a relative concision. This concision is a condition to build an usable model. Introducing the shipment size variable in freight transport demand models requires specific strategies to keep the representation of the freight transport system usable. In particular the detailed and accurate representation of each shipment in the system is impracticable, both in terms of data and computing requirements.

As a first step, and given the lack of coordination between the agents of the freight transport system, it seems reasonable to address separately the freight transport demand and supply. As stated above, the supply must be represented in a summarised way. It is impossible to describe the positions and movements of every vehicle, driver, shipment in the freight transport system. A simple strategy can be to search for a statistical linkage between some observable costs (such as fuel, driver, vehicle, etc.), other variables (e.g. the distance between the origin and destination, the urban density in both areas, etc.) and freight price schedules.

Concerning the demand, the EOQ model may constitute a sensible first step to model the demand. Fortunately, the representation of the demand in the frame of the EOQ model is simple: a commodity flow is sufficiently described by its rate and commodity value of time (note that this is a strong distinction between passenger and freight transport modelling: passengers are usually described by their value of time alone).

This constitutes only a rough sketch of the tasks implied by the design of a spatialised freight transport model with an explicit representation of shipment sizes.

Bibliography

Abad, P. L. (2007), 'Buyer's response to a temporary price reduction incorporating freight costs', *European Journal of Operational Research* **182**, 1073–1083.

Abdelwahab, W. M. (1998), 'Elasticities of mode choice probabilities and market elasticities of demand: Evidence from a simultaneous mode choice/shipment-size freight transport model', *Transportation Research Part E* **34**(4), 257–266.

Abdelwahab, W. M. and Sargious, M. A. (1992), 'Modeling the demand for freight transport: New approach', *Journal of Transport Economics and Policy* **26**(1), 49–72.

Abraham, C. (2006), Pour une régulation durable du transport routier de marchandises, Centre d'analyse stratégique.

Académie des Technologies (2009), *Le transport de marchandises*, Editions le Manuscrit.

Ambrosini, C. and Routhier, J. L. (2004), 'Objectives, methods and results of surveys carried out in the field of urban freight transport: an international comparison', *Transport Reviews* **24**(1), 57–77.

Ambrosini, C., Routhier, J. L. and Toilier, F. (2004), How do urban policies work on urban goods transport flows?, *in* 'Proceedings of World Conference on Transport Research', Istanbul.

Anciaux, D. and Yuan, K. (2007), Green supply chain: Intermodal transportation modelling with environmental impacts, *in* 'Proceedings of the European Transport Conference'.

Anderson, S. P., de Palma, A. and Thisse, J.-F. (1992), *Discrete Choice Theory of Product Differentation*, MIT Press, Cambridge, Massachussetts, London, England.

Barankin, E. W. (1961), 'A delivery-lag inventory model with an emergency provision (the single period case)', *Naval Research Logistics Quarterly* **8**, 285–311.

Basso, L. J. and Jara-Diaz, S. R. (2005), 'Calculation of economies of spatial scope from transport cost functions with aggregate output with an application to the airline industry', *Journal of Transport Economics and Policy* **39**(1), 25–52.

Baumol, W. J. (1977), 'On the proper cost tests for natural monopoly in a multiproduct industry', *American Economic Review* **67**(5), 809–22.

Baumol, W. J. and Vinod, H. D. (1970), 'An inventory theoretic model of freight transport demand', *Management Science* **16**(7), 413–421.

Ben-Akiva, M. and Lerman, S. (1985), *Discrete Choice Analysis*, MIT Press, Cambridge.

Beuthe, M., Degrandsart, F., Geerts, J. F. and Jourquin, B. (2002), 'External costs of the Belgian interurban freight traffic: a network analysis of their internalisation', *Transportation Research Part D.* **7**, 285–301.

Bishop, R. L. (1964), 'A zeuthen-hicks theory of bargaining', *Econometrica* **32**(3), 410–417.

Blumenfeld, D., Burns, L. and Daganzo, C. (1991), 'Synchronizing production and transportation schedule', *Transportation Research Part B* **25B**(1), 346–360.

Blumenfeld, D. E., Burns, L. D., Diltz, J. D. and Daganzo, C. (1985), 'Analyzing trade-offs between transportation, inventory and production costs on freight networks', *Transportation Research Part B* **19**(5).

Carbone, V. (2004), Rôle des prestataires logistiques en Europe, intégration des chaînes logistiques et alliances logistiques, PhD thesis, Ecole Nationale des Ponts et Chaussées.

Catalani, M. (2003), Transport competition on a multimodal corridor by elasticity evaluation, *in* 'Proceedings of the European Transport Conference', Strasbourg.

Chopra, S. (2003), 'Designing the distribution network in a supply chain', *Transportation Research Part E.* **39**, 123–140.

Christopher, M. (1992), *Logistics and Supply Chain Management – Strategies for reducing costs and improving services*, Pitman Publishing, London, England, p. 229.

Combes, F. and Leurent, F. (2007), Advances in freight transport demand modelling: an assessment with research perspectives, *in* 'European Transport Conference', Leuwenhoorst, NL.

Combes, P.-P. and Lafourcade, M. (2005), 'Transport costs: Measures, determinants, and regional policy implications for france', *Journal of Economic Geography* **5**, 319–349.

Combes, P. P., Mayer, T. and Thisse, J.-F. (2006), *Economie Géographique*, Economica.

Daganzo, F. F. (2005), *Logistics Systems Analysis*, fourth edn, Springer, Berlin.

Danzanvilliers, P., Carpentier, S. and Leurent, F. (2005), L'intermodalité dans les études de planification des transports en milieu interurbain, *in* 'Proceedings of Congrès ATEC 2005', Paris.

de Jong, G. and Ben-Akiva, M. (2007), 'A micro-simulation model of shipment size and transport chain choice.', *Transportation Research Part B* **41**, 950–965.

de Jong, G., Ben-Akiva, M. and Baak, J. (2007), A micro-model for logistics decisions in norway and sweden calibrated to aggregate data, *in* 'Proceedings of the European Transport Conference'.

de Jong, G., Ben-Akiva, M., Florian, M., Gronland, S. E. and van de Voort, M. (2005), Specification of a logistic model for norway and sweden, *in* 'Proceedings of the European Transport Conference'.

de Jong, G., Kroes, E., Plasmeijer, R., Sanders, P. and Warffemius, P. (2004), The value of reliability, *in* 'Proceedings of the European Transport Conference'.

De Vany, A. S. and Saving, T. R. (1977), 'Product quality, uncertainty, and regulation: The trucking industry', *The American Economic Review* **67**(4), 583–594.

Dejonckheere, J., Disney, S. M., Lambrecht, M. R. and Towill, D. R. (2003), 'Measuring and avoiding the bullwhip effect: A control theoretic approach', *European Journal of Operational Research* **147**, 567–590.

Delvaux, G. and Duhautois, R. (1998), 'Complémentarité des facteurs de production dans le transport routier de marchandises en france', Notes de Synthèse du SESP.

Disney, S. M., Potter, A. T. and Gardner, B. M. (2003), 'The impact of vendor managed inventory on transport operations', *Transportation Research Part E.* **39**, 363–380.

Dornier, P.-P. and Fender, M. (2007), *La logistique globale et le Supply Chain Management*, second edn, Eyrolles, Paris, France.

Erlenkotter, D. (1989), 'An early classic misplaced: Ford W. Harris's economic order quantity of 1915', *Management Science* **35**(7), 898–900.

Evans, C. and Harrigan, J. (2005), 'Distance, time, and specialization: Lean retailing in general equilibrium', *The American Economic Review* **95**(1), 292–313.

Felton, J. R. (1981), 'Impact of ICC rate regulation upon truck back hauls', *Journal of Transport Economics and Policy* **15**(3), 253–267.

Fernandez, A., Arruñada, B. and Gonzalez-Diez, M. (2006), Quasi-integraation in less-than-truckload trucking. Social Science Research Network, Working Paper.

Forrester, J.-W. (1961), *Industrial Dynamics*, Wiley Editions, New York.

Fowkes, A. S., Firmin, P. E., Tweddle, G. and Whiteing, A. E. (2004), 'How highly does the freight transport industry value journey time reliability – and for what reasons?', *International Journal of Logistics: Research and Applications* **7**(1), 33–43.

Fujita, M., Krugman, P. and Venables, A. (1999), *The Spatial Economy*, The MIT Press, Cambridge, Massachussets, London, England.

Garrido, R. A. and Mahmassani, H. S. (2000), 'Forecasting freight transportation demand with the space-time multinomial probit model', *Transportation Research Part B* **34**, 403–418.

Gillen, D., Oum, T. H. and Tretheway, M. (1990), 'Airline cost structure and policy implications, a multi-product approach for Canadian airlines', *Journal of Transport Economics and Policy* **24**(1), 9–34.

Golicic, S. L., Foggin, J. H. and Mentzer, J. T. (2003), 'Relationship magnitude and its role in interorganizational relationship structure', *Journal of Business Logistics* **24**(1), 57–75.

Goyal, S. K. (1977), 'An integrated inventory model for a single product system', *Operations Research Quarterly* **28**(3), 539–545.

Groothedde, B. (2003), Economies of scale through collaboration in an intermodal hub network, *in* 'Proceedings of the European Transport Conference'.

Guerrien, B. (2002), *Dictionnaire d'Analyse Economique*, 3rd edition edn, La Découverte.

Guilbault, M. and Gouvernal, E. (2008), Demand for transport and production systems - shipper surveys in france in 1998 and 2004, *in* 'European Transport Conference', Leuwenhoorst, NL.

Guilbault, M. et al. (2006), Enquête ECHO, premiers résultats d'analyse, Phase 6 rapport final, INRETS.

Guilbault, M. et al. (2008), Enquête ECHO envoi — chargeurs – opérateurs de transport. résultats de référence, Synthèse n°58, INRETS.

Hall, R. W. (1985), 'Dependence between shipment size and mode in freight transportation', *Transportation Science* **19**(4), 436–444.

Hall, R. W. (1989), 'Vehicle packing', *Transportation Research Part B* **23**(2), 103–121.

Harker, P. T. (1986), 'Alternative models of spatial competition', *Operation Research* **34**(3), 410–425.

Harris, F. W. (1913), 'How many parts to make at once', *Factory, the Magazine of Management* **10**(2), 135–136. 152.

Helpman, E. and Krugman, P. R. (1985), *Market Structure and Foreign Trade*, MIT Press.

Hensher, A. D. and Puckett, S. M. (2005), 'Refocusing the modelling of freight distribution: Development of an economic-based framework to evaluate supply chain behaviour in response to congestion charging', *Transportation* **32**, 573–602.

Hill, R. M. (1997), 'The single-vendor single-buyer integrated production-inventory model with a generalised policy', *European Journal of Operation Research* **97**, 493–499.

Holguin-Veras, J. (2002), 'Revealed preference analysis of commercial vehicle choice process', *Journal of Transportation Engineering* **128**(4), 336–346.

Holguin-Veras, J. (2007), Us commodity flow survey – presentation and challenges, *in* 'Journée Spécialisée INRETS – Enquête ECHO – Les Apports des Enquêtes Chargeurs pour la Connaissance des Chaînes de Transport de Marchandises et de leurs Déterminants Logistiques'.

Holguin-Veras, J., Preziosi, M., de Jong, G. and Maurer, H. (2007), An experimental economics investigation of shipper-carrier interactions on the choice of mode and shipment size in freight transport, *in* 'Proceedings of the European Transport Conference'.

Hotelling, H. (1929), 'Stability in competition', *The Economic Journal* **39**(153), 41–57.

Huh, W. T., Janakiraman, G., Muckstadt, J. A. and Rusmevichientong, P. (2009), 'Asymptotic optimality of order-up-to policies in lost sales inventory systems', *Management Science* **55**(3), 404–420.

Ivanova, O. (2007), On the development of a new version of the RAEM model for the netherlands (raem 3.0), *in* 'Proceedings of the Jointed Congress of the European Regional Science Association (47th) and Association de Science Régionale de Langue Française (44th)', Paris.

Jin, Y. and Williams, I. (2005), Integrated regional economic and freight logistics modelling: Results from a model for trans-pennine corridor, uk, *in* 'Proceedings of the European Transport Conference', Strasbourg.

John, S., Naim, M. M. and Towill, D. (1994), 'Dynamic analysis of a wip compensated decision support system', *International Journal of Manufacturing System Design* **1**(4), 283–297.

Jourquin, B. and Beuthe, M. (2000), Multimodal freight network analysis with NODUS, a survey of several applications, *in* 'Proceedings of the 3rd KFB Research Conference', Stockholm.

Jourquin, B. and Beuthe, M. (2005), Multimodal freight network analysis with NODUS, a survey of several applications, *in* 'Proceedings of the European Transport Conference', Strasbourg.

Karlin, S. and Scarf, H. (1958), Inventory models of the Arrow-Harrison-Marschak type with time lag, *in* K. J. Arrow, S. Karlin and H. Scarf, eds, 'Studies in the Mathematical Theory of Inventory and Production', Stanford University Press, Palo Alto, CA, pp. 155–178.

Kettner, M. and Sewcyk, B. (2002), A model for transport planning and railway network evaluation, *in* 'Proceedings of the 9th World Congress on Intelligent Transport Systems', Chicago.

Kettner, M., Sewcyk, B. and Eickmann, C. (2003), Integrating microscopic and macroscopic models for railway network evaluation, *in* 'Proceedings of the European Transport Conference'.

Korte, B. and Vygen, J. (2008), *Combinatorial Optimisation, Theory and Algorithms*, seventh edn, Springer, Berlin.

Krugman, P. and Wells, R. (2006), *Economics*, Worth Pub.

Lancaster, K. J. (1966), 'A new approach to consumer theory', *The Journal of Political Economy* **74**(2), 132–157.

Lee, D.-H., Dong, M., Bian, W. and Tseng, Y.-J. (2007), 'Design of product recovery networks under uncertainty', *Transportation Research Record: Journal of the Transportation Research Board* **2008**, 19–25.

Leontief, W. (1936), *Input-output economics*, 2nd revised edition (1986) edn, Oxford University Press.

Leurent, F. and Meunier, F. (2009), Optimal network estimation of origin-destination flow from link data, *in* 'Transportation Research Board', Washington.

Mariotte, H. (2007), 'L'emploi dans la fonction logistique en france', SESP.

Marzano, V. and Papola, A. (2004), Modelling freight demand at a national level: theoretical developments and application to italian demand, *in* 'Proceedings of the European Transport Conference'.

Maurice, J. and Crozet, Y. (2007), *Le calcul économique*, Economica.

McFadden, D., Winston, C. and Boersch-Supan, A. (1985), Joint estimation of freight transportation decisions under non-random sampling, *in* A. Daughety, ed., 'Analytical Studies in Transport Economics', Cambridge University Press.

McKinnon, A. C. (2005), 'The economic and environmental benefits of increasing maximum truck weight: the british experience', *Transportation Research Part D* **10**, 77–95.

ME&P (2000), 'STREAMS (strategic transport research for european member states)', Final Report.

ME&P (2002), 'SCENES european transport scenarios', Summary Report.

ME&P and WSP (2002), 'Review of freight modelling', Final Report.

Metzler, L. (1941), 'The nature and stability of inventory cycles', *The Review of Economic Statistics* **23**(3), 113–129.

Mohring, H. (1985), 'Profit maximization, cost minimization, and pricing for congestion-prone facilities', *Logistics and Transportation Review* **21**, 27–36.

Moinzadeh, K. and Nahmias, S. (1988), 'A continuous review model for an inventory system with two supply modes', *Management Science* **34**(6), 761–773.

Morton, T. E. (1969), 'Bounds on the solution of the lagged optimal inventory equation with no demand backlogging and proportional costs', *SIAM Review* **11**(4), 572–596.

Nash, J. F. (1950), 'The bargaining problem', *Econometrica* **18**(2), 155–162.

NEI et al. (1999), 'REDEFINE; relationship between demand for freight-transport and industrial effects', Summary Report.

Newell, G. F. (1973), 'Scheduling, location, transportation, and continuum mechanics; some simple approximations to optimisation problems', *SIAM Journal of Applied Mathematics* **25**(3), 346–360.

Olson, J. (1972), 'Price discrimination by regulated motor carriers', *The American Economic Review* **62**(3), 395–402.

Oum, T. H., Waters, W. G. and Yong, J.-S. (1992), 'Concepts of price elasticities of transport demand and recent empirical estimates', *Journal of Transport Economics and Policy* **26**(2), 139–154.

Picard, P. (2007), *Eléments de Microéconomie*, seventh edn, Montchrestien, Paris.

Potter, A. and Lalwani, C. (2007), 'Investigating the impact of demand amplification on freight transport', *Transportation Research Part E* . doi:10.1016/j.tre.2007.06.001.

Quandt, R. (1968), 'Estimation of modal splits', *Transportation Research* **2**, 41–50.

Quinet, E. (1998), *Principes d'économie des transports*, Economica.

Quinet, E. and Vickermann, R. W. (2004), *Principles of Transport Economics*, Edward Elgar Publishing Ltd.

Regan, A. and Garrido, R. A. (2002), Modelling freight demand and shipper behavior: state of the art, future directions, *in* D. Hensher, ed., 'The Leading Edge of Travel Behaviour Research', Pergamon Press.

Routhier, J.-L., Segalou, E. and Durand, S. (2002), Mesurer l'impact du transport de marchandises en ville : le modèle de simulation freturb v1., Report for the french program *Marchandises en Ville*.

Russo, F. and Comi, A. (2004), A state of the art on urban freight distribution at european scale, *in* 'Proceedings of the European Conference on Mobility Management'.

Savy, M. (2006*a*), *Le transport de marchandises*, Eyrolles, Paris, France.

Savy, M. (2006*b*), Logistique et territoire, Technical report, DIACT.

SESP (2001), 'La sous-traitance transport dans l'enquête annuelle d'entreprise 1998'.

SESP (2008), 'Entreprises de transport : résultats de l'EAE 2006'.

Sewcyk, B., Radtke, A. and Wilfinger, G. (2007), Combining microscopic and macroscopic infrastructure planning models, *in* 'Proceedings of the IAROR Congress, 2nd International Seminar on Railway Operations Modelling and Analysis', Hannover.

Shmoys, D. B., Tardos, E. and Aardal, K. (1997), Aproximation algorithms for facility location problems, *in* 'Proceedings of the 29th Annual ACM Symposium on Theory of Computing', pp. 265–274.

Small, K. A. and Verhoef, E. (2006), *The Economics of Urban Transportation*, 2nd revised edition edn, Routledge.

Smith, D. L., Campbell, J. F. and Mundy, R. (2007), 'Modeling net rates for expedited freight services', *Transportation Research Part E* **43**, 192–207.

Su, X. and Zhang, F. (2009), 'On the value of commitment and availability guarantees when selling to strategic consumers', *Management Science* **55**(5), 713–726.

Szpiro, D. (1996), 'Le transport de marchandises : volume, qualité et prix hédoniques', Working paper, Université de Lille I, Modern and Clersé, in collaboration with Gouvernal, E. and Hannape, P.

Tatineni, V. C. and Demetsky, M. J. (2005*a*), 'Supply chain models for freight transportation', University of Virginia. Research Report No. UVACTS-14-0-85.

Tatineni, V. C. and Demetsky, M. J. (2005*b*), 'Supply chain models for freight transportation', University of Virginia, Research Report No. UVACTS-14-0-85.

Tavasszy, L. A. (2006), Freight modelling - an overview of international experiences, *in* 'Transportation Research Board', Washington.

Tavasszy, L., van der VVlist, M. J. M. and Ruilgrok, C. J. (1998), Scenario-wise analysis of transport and logistic systems with a smile, *in* 'Proceedings of World Conference on Transport Research', Antwerp.

Thissen, M. (2003), 'A regional applied general equilibrium model for the Netherlands', WP 1, TNO Inro, Delft.

Tirole, J. (1988), *The Theory of Industrial Organization*, 1st edition edn, MIT Press.

Towill, D. R. (1982), 'Dynamic analysis of an inventory and order-based production control system', *International Journal of Production Research* **20**, 369–383.

Tyworth, J. T. (1991), 'The inventory theoretic approach in transportation selection models: a critical review', *The Logistics and Transportation Review* **27**(4).

Vernimmen, B. and Witlox, F. (2003), 'The inventory theoretic approach to modal choice in freight transport: literature review in case studies', *Brussels Economic Review/Cahiers Economiques de Bruxelles* **46**(2).

Wilson, R. H. (1934), 'A scientific routine for cost control', *Harvard Business Review* **13**, 116–128.

Winston, C. (1981), 'A disaggregate model of the demand for intercity freight transportation', *Econometrica* **49**(4), 981–1006.

Woodburn, A. G. (2003), 'A logistical perspective on the potential for modal shift of freight from road to rail in great britain', *International Journal of Transport Management* **1**, 237–245.

Xu, K., Windle, R., Grimm, C. and Corsi, T. (1994), 'Re-evaluating returns to scale in transport', *Journal of Transport Economics and Policy* **8**(3), 275–286.

Appendix A

Appendix to Chapter 4

A.1 Classic roadside questionnaire

A.2 New questionnaire, RN10 survey

Interviewer name:

Date and station

1 2 Station

3 4 Day — 5 6 Month — 7 8 Hour

9 10 Hour

"GEODE" DATABASE

ROAD FREIGHT TRANSPORT QUESTIONNAIRE

Survey location

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Centre d'Études Techniques de l'Équipement du Sud-Ouest

Number of axles 2 - 2 axles 3 - 3 axles 4 - 4 axles 5 - 5 axles and more	Vehicle registration If French, department number Else, country code	Origin of the trip WHERE ARE YOU COMING FROM? Where was your last compulsory stop? (to load or unload freight, or to take the vehicle) give in this order : department or country code (3 characters), city or province (3 characters), district (2 characters).	Destination of the trip WHERE ARE YOU GOING? Where is your next compulsory stop? (to load or unload freight, or to leave the vehicle) give in this order : department or country code (3 characters), city or province (3 characters), district (2 characters).	Trip length (between the origin and the destination) in Km flush right	Are you running empty or loaded? 1 - empty 2 - loaded	How much freight do you carry (in tons? round to the upper ton	If your vehicle is loaded, what are you carrying? spell out fully the commodity type.	Conditioning 1 - Container 2 - Box 3 - Tanker 4 - Reefer 5 - Dry bulk 6 - Flatbed 8 - Curtain sider 7 - other	Hazardous materials read from the orange descriptive plate the risk class and the material code
11	12 13 14	15 16 17 18 19 20 21 22	23 24 25 26 27 28 29 30	31 32 33 34	35	36 37	38 39	40	41 42 43 44 45 46 47 48
11	12 13 14	15 16 17 18 19 20 21 22	23 24 25 26 27 28 29 30	31 32 33 34	35	36 37	38 39	40	41 42 43 44 45 46 47 48
11	12 13 14	15 16 17 18 19 20 21 22	23 24 25 26 27 28 29 30	31 32 33 34	35	36 37	38 39	40	41 42 43 44 45 46 47 48
11	12 13 14	15 16 17 18 19 20 21 22	23 24 25 26 27 28 29 30	31 32 33 34	35	36 37	38 39	40	41 42 43 44 45 46 47 48

Interviewer | | | Date | | | | | | |
d d m m y y

Road | | | | | Service area | | | Hour | | |

Vehicle characteristics

Q1 Number of axles
2 - 2 axles
3 - 3 axles
4 - 4 axles
5 - 5 axles and more

Q2 Registration plate
Front
Back

Q3 Vehicle type
1 - Container
2 - Rigid sider
3 - Tanker
4 - Reefer
5 - Dry bulk
6 - Flatbed
7 - Others
8 - Curtain sider

Freight

Q4 Is the vehicle empty or loaded? 1 - loaded 2 - empty

Q5 If loaded, what is the commodity type?
if several commodity types, note the main one

Q6 How much freight are you carrying?
Round to the upper ton

Q7 How much of your vehicle's volume is used?
1 - 1/4
2 - 1/2
3 - 3/4
4 - all the volume

Freight route

Q8 Where was the freight loaded (if loaded) or unloaded (if empty)?
Country + city

Q9 Where will the freight be unloaded (if loaded) or loaded (if un
Country + city

Q10 What is the length of this trip? km

Driver route

Q11 How many drivers are they onboard?

Q12 Are you doing the same trip as the freight? 1 - yes 2 - no

If No, you are doing a relay, a part of the freight trip, which one?

Q13 What is your trip's origin?

Country + city

Q14 What is your trip's destination

Country + city

Q15 What is the length of this trip? km

Schedule and breaks

Q16 When did you leave? d d m m h h

Q17 When do you think you will arrive? d d m m h h

Q18 Do you have a specific imperative on the arrival time? 1 - yes 2 - no

Q19 If yes, which one? d d m m h h

Q20 How many breaks did you have since your departure?

Q21 Where was your longest break?

Q22 How much did it last? h h m m

Q23 What was its motive?

1 - sleep (night)
2 - meal
3 - regulatory
4 - other

Q24 Why did you prefer RN10 to A10

1 - It's free
2 - It is the shortest way
3 - the restaurants are better
4 - parking is more convenient
5 - by order of the boss
6 - the fuel is less expensive
7 - other (precise)

If hazardous materials, note the plate numbers

Appendix B

Appendix to Chapter 5

B.1 Validation of the EOQ specification

We keep the validation of the EOQ specification simple by limiting ourselves to a graphical analysis of the residuals. This graphical diagnosis is based on Figure B.1.

From the upper left figure we observe there are strong regularities in the residuals. These regularities have two main reasons. First, the shipments are measured in kg, so that shipments of approximately 1 kg are uniformly indicated as weighting exactly 1kg. There are 298 such shipments, hence the lower left diagonal observed on the upper left graph. The other diagonals result from the various vehicle capacities: many shipments are dimensioned so as to fill freight transport vehicles.

Despite these strong regularities, the residuals are almost exactly normally distributed, as illustrated by the upper right normal Q-Q plot. The lower left graph indicates no heteroscedasticity, and the lower right one indicates no obviously aberrant observation.

B.2 Validation of the extended EOQ specification

The diagnosis of the ordinary least square hypotheses is based on the graphical analysis of the residuals, thanks to Figure B.2.

The comments made about the simple EOQ model remain valid here. The residuals still seem to be homoscedastic and normally distributed, and there seems to be no aberrant observation.

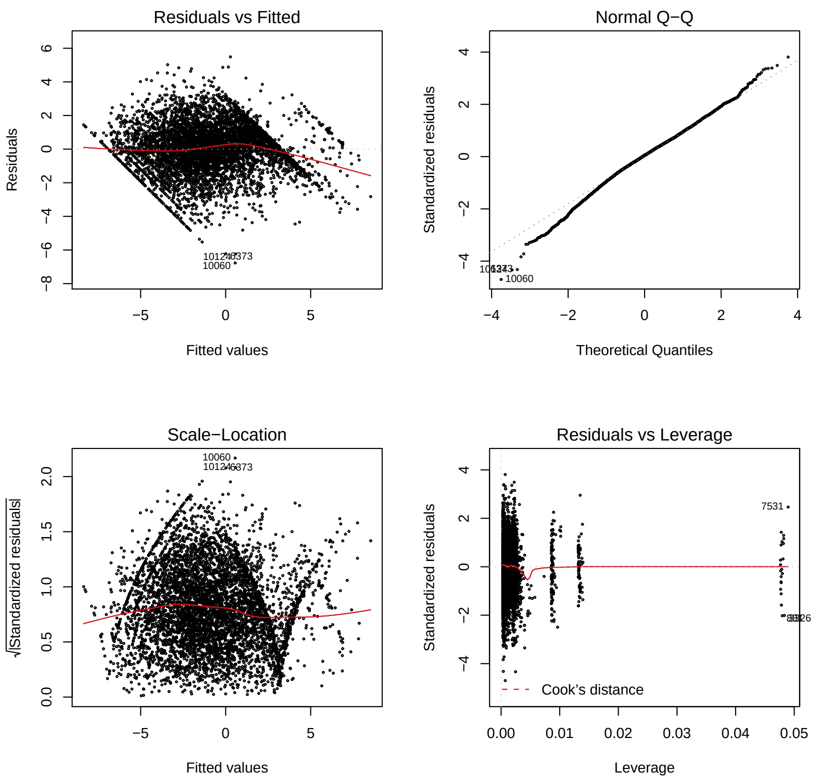

Figure B.1: Residuals of the EOQ model

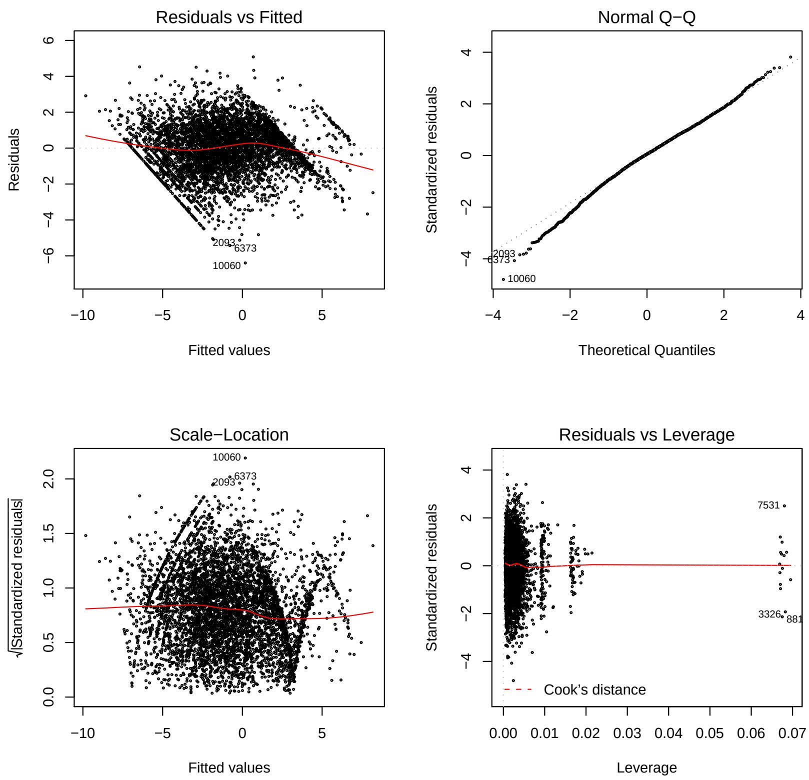

Figure B.2: Residuals of the extended EOQ model

Appendix C

Proofs of Chapter 6

C.1 Proof of Lemma 6.10

Given q_1^s and q_2^s positive, denote f_1 the distribution density of N_1^s and f_2 the distribution density of N_2^s. Denote $I_1 = [(1 - K_s/2)q_1^s, (1 + K_s/2)q_1^s]$, and $I_2 = [(1 - K_s/2)q_2^s, (1 + K_s/2)q_2^s]$. Hypothesis 6.8 implies that:

$$f_1^s(x) = \frac{1}{K_s q_1^s} \quad {}_{I_1}(x),$$

and:

$$f_2^s(y) = \frac{1}{K_s q_2^s} \quad {}_{I_2}(y).$$

Then:

$$\delta(q_1^s; q_2^s) = \frac{1}{K_s^2 q_1^s q_2^s} \int_D (x - y) dx dy,$$

with D defined as:

$$D = (I_1 \times I_2) \cap \{(x; y)/x \geq y\}$$

We distinguish four cases. First, if $(1 + K_s/2)q_1^s \leq (1 - K_s/2)q_2^s$, $D = \emptyset$, as a consequence $\delta(q_1^s; q_2^s) = 0$, and so are its derivatives. Second, if $(1 - K_s/2)/(1 + K_s/2)q_2^s \leq q_1^s \leq q_2^s$:

$$\delta(q_1^s; q_2^s) = \frac{1}{K_s^2 q_1^s q_2^s} \int_{(1-K_s/2)q_2^s}^{(1+K_s/2)q_1^s} A(y) dy,$$

with $A(y) = \int_y^{(1+K_s/2)q_1^s} (x - y) dx$. But:

$$A(y) = \left[\frac{x^2}{2}\right]_y^{(1+K_s/2)q_1^s} - \left(\left(1 + \frac{K_s}{2}\right) q_1^s - y\right) y,$$

which is equivalent to:

$$A(y) = \frac{1}{2}\left(1+\frac{K_s}{2}\right)^2 (q_1^s)^2 - \left(1+\frac{K_s}{2}\right) q_1^s y + \frac{y^2}{2},$$

or:

$$A(y) = \frac{1}{2}\left(\left(1+\frac{K_s}{2}\right) q_1^s - y\right)^2.$$

As a consequence:

$$\delta(q_1^s; q_2^s) = \frac{1}{6K_s^2 q_1^s q_2^s}\left(\left(1+\frac{K_s}{2}\right) q_1^s - \left(1-\frac{K_s}{2}\right) q_2^s\right)^3$$

The derivative of δ with respect to q_1^s when $(1-K_s/2)/(1+K_s/2)q_2^s \leq q_1^s \leq q_2^s$ is then:

$$\frac{\partial\delta}{\partial q_1^s} = \frac{1}{6K_s^2 q_1^s q_2^s}\left(-\frac{1}{q_1^s}B^3 + 3\left(1+\frac{K_s}{2}\right)B^2\right),$$

where $B = (1+K_s/2)q_1^s - (1-K_s/2)q_2^s$. We have:

$$\frac{\partial\delta}{\partial q_1^s} = \frac{B^2}{6K_s^2 q_1^s q_2^s}\left[-\frac{1}{q_1^s}\left(\left(1+\frac{K_s}{2}\right) q_1^s - \left(1-\frac{K_s}{2}\right) q_2^s\right) + 3\left(1+\frac{K_s}{2}\right)\right]$$

$$\Leftrightarrow \frac{\partial\delta}{\partial q_1^s} = \frac{B^2}{6K_s^2 q_1^s q_2^s}\left(\left(1-\frac{K_s}{2}\right)\frac{q_2^s}{q_1^s} + 2\left(1+\frac{K_s}{2}\right)\right).$$

Therefore:

$$\frac{\partial\delta}{\partial q_1^s} = \frac{q_1^s}{6K_s^2 q_2^s}\left(\left(1+\frac{K_s}{2}\right) - \left(1-\frac{K_s}{2}\right)\frac{q_2^s}{q_1^s}\right)^2$$

$$\times\left(2\left(1+\frac{K_s}{2}\right) + \left(1-\frac{K_s}{2}\right)\frac{q_2^s}{q_1^s}\right)$$

or, equivalently:

$$\frac{\partial\delta}{\partial q_1^s} = \frac{1}{6K_s^2}\left(\left(1+\frac{K_s}{2}\right) - \left(1-\frac{K_s}{2}\right)\frac{q_2^s}{q_1^s}\right)^2$$

$$\times\left(2\left(1+\frac{K_s}{2}\right)\frac{q_1^s}{q_2^s} + \left(1-\frac{K_s}{2}\right)\right).$$

We observe that $\partial\delta/\partial q_1^s$ is indeed homogeneous of degree zero (*i.e.* function of q_1^s/q_2^s) and is increasing in q_1^s/q_2^s from 0 to $(6+K_s)/12$ on $[(1-K_s/2)/(1+K_s/2);1]$. Symmetrically, on the same interval:

$$\frac{\partial\delta}{\partial q_2^s} = -\frac{q_1^s}{6K_s^2 q_2^s}\left(\left(1+\frac{K_s}{2}\right)-\left(1-\frac{K_s}{2}\right)\frac{q_2^s}{q_1^s}\right)^2 \times\left(\left(1+\frac{K_s}{2}\right)\frac{q_1^s}{q_2^s}+2\left(1-\frac{K_s}{2}\right)\right),$$

from which we conclude that $\partial\delta/\partial q_2^s$ is also homogeneous of degree zero and is decreasing in q_1^s/q_2^s from 0 to $-(6-K_s)/12$ on $[(1-K_s/2)/(1+K_s/2);1]$.

Consider now the case where $q_1^s/q_2^s \in [1;(1+K_s/2)/(1-K_s/2)]$. Then:

$$\delta(q_1^s;q_2^s) = \frac{1}{K_s^2 q_1^s q_2^s}\left(\int_{I_1\times I_2}(x-y)dxdy - \int_{I_1\times I_2\smallsetminus D}(x-y)dxdy\right)$$

Straightforwardly:

$$\frac{1}{K_s^2 q_1^s q_2^s}\int_{I_1\times I_2}(x-y)dxdy = q_1^s - q_2^s,$$

and:

$$\frac{1}{K_s^2 q_1^s q_2^s}\int_{I_1\times I_2\smallsetminus D}(x-y)dxdy = -\frac{1}{6K_2^2 q_1^s q_2^s}\left(\left(1+\frac{K_s}{2}\right)q_2^s-\left(1-\frac{K_s}{2}\right)q_1^s\right)^3.$$

We adapt the previous calculations and obtain:

$$\frac{\partial\delta}{\partial q_1^s} = 1-\frac{q_2^s}{6K_s^2 q_1^s}\left(\left(1+\frac{K_s}{2}\right)-\left(1-\frac{K_s}{2}\right)\frac{q_1^s}{q_2^s}\right)^2 \times\left(\left(1+\frac{K_s}{2}\right)\frac{q_2^s}{q_1^s}+2\left(1-\frac{K_s}{2}\right)\right)$$

The second term of the right hand side of this equation decreases from $(6-K_s)/12$ to zero when q_1^s/q_2^s increases from 1 to $(2+K_s)/(2-K_s)$. As a consequence, $\partial\delta/\partial q_1^s$ increases from $(6+K_s)/12$ to 1 on the same interval. Symmetrically:

$$\frac{\partial\delta}{\partial q_2^s} = -1+\frac{1}{6K_s^2}\left(\left(1+\frac{K_s}{2}\right)-\left(1-\frac{K_s}{2}\right)\frac{q_1^s}{q_2^s}\right)^2 \times\left(2\left(1+\frac{K_s}{2}\right)\frac{q_2^s}{q_1^s}+\left(1-\frac{K_s}{2}\right)\right)$$

Therefore, $\partial\delta/\partial q_2^s$ decreases from $-(6-K_s)/12)$ to -1 when q_1^s/q_2^s increases from 1 to $(2+K_s)/(2-K_s)$.

The fourth and final case is when $q_1^s/q_2^s \geq (2+K_s)/(2-K_s)$. In that case, $D = I_1 \times I_2$ and $\delta(q_1^s; q_2^s) = q_1^s - q_2^s$. Its first derivative with respect to q_1^s and q_2^s are 1 and -1 respectively. These results ensure first that δ is derivable on $\mathbb{R}_+^2$, second that Lemma 6.10 holds.

C.2 Calculation of the average LTL load factor

Consider a carrier which has to carry N_1^s shipments of size 1 and N_2^s shipments of size 2. Consistently with Hypothesis 6.8, both these variables are random and distributed along uniform distributions of mean q_i^s and variance $(K_s.q_i^s)^2/12$. The objective of this section is to calculate λ_{LTL} the expected value of Λ_{LTL}. From Equation (6.10):

$$\Lambda_{LTL} = \begin{cases} 1 & \text{if} \quad N_1^s \geq N_2^s, \\ \frac{2}{3} + \frac{N_1^s}{3N_2^s} & \text{else.} \end{cases}$$

λ_{LTL} the expected value of Λ_{LTL}, is therefore equal to:

$$\lambda_{LTL} = 1.\mathbb{P}\{N_1^s > N_2^s\} + \left(\frac{2}{3} + \frac{1}{3}\mathbb{E}\left(\frac{N_1^s}{N_2^s}\middle| N_1^s \leq N_2^s\right)\right).\mathbb{P}\{N_1^s \leq N_2^s\} \tag{C.1}$$

Four cases should be distinguished with respect to the relative values of q_1^s and q_2^s. First, assume $(2+K_s)q_1^s < (2-K_s)q_2^s$. In that case, $\mathbb{P}\{N_1^s > N_2^s\} = 0$, so that $\mathbb{E}(N_1^s/N_2^s|N_1^s < N_2^s) = \mathbb{E}(N_1^s/n_2^s)$, which is:

$$\mathbb{E}\left(\frac{N_1^s}{N_2^s}\right) = \frac{1}{K_s^2 q_1^s q_2^s}\int_{(1-K_S/2)q_2^s}^{(1+K_s/2)q_2^s}\int_{(1-K_s/2)q_1^s}^{(1+K_s/2)q_1^s}\frac{x}{y}dxdy.$$

We first integrate along x:

$$\mathbb{E}\left(\frac{N_1^s}{N_2^s}\right) = \frac{1}{K_s^2 q_1^s q_2^s}\int_{(1-K_S/2)q_2^s}^{(1+K_s/2)q_2^s}\frac{K_s(q_1^s)^2}{y}dy,$$

then along y:

$$\mathbb{E}\left(\frac{N_1^s}{N_2^s}\right) = \frac{1}{K_s}\ln\left(\frac{2+K_s}{2-K_s}\right)\frac{q_1^s}{q_2^s}. \tag{C.2}$$

Therefore, $(2+K_s)q_1^s < (2-K_s)q_2^s$ implies:

$$\lambda_{LTL} = \frac{2}{3} + \frac{1}{3K_s} \ln\left(\frac{2+K_s}{2-K_s}\right)\frac{q_1^s}{q_2^s}. \tag{C.3}$$

Consider now the second case: $(2-K_s)/(2+K_s).q_2^s \le q_1^s < q_2^s$. The following identity holds:

$$\begin{aligned}\mathbb{E}\left(\frac{N_1^s}{N_2^s}\right) = {} & \mathbb{E}\left(\frac{N_1^s}{N_2^s}\middle| N_1^s \le N_2^s\right).\mathbb{P}\{N_1^s \le N_2^s\} \\ & + \mathbb{E}\left(\frac{N_1^s}{N_2^s}\middle| N_1^s > N_2^s\right).\mathbb{P}\{N_1^s > N_2^s\},\end{aligned}$$

so that Equation (C.1) can be rewritten as follows:

$$\begin{aligned}\lambda_{LTL} = {} & \frac{2}{3} + \frac{1}{3}\mathbb{P}\{N_1^s > N_2^s\} + \frac{1}{3}\mathbb{E}\left(\frac{N_1^s}{N_2^s}\right) \\ & - \frac{1}{3}\mathbb{E}\left(\frac{N_1^s}{N_2^s}\middle| N_1^s > N_2^s\right).\mathbb{P}\{N_1^s > N_2^s\}.\end{aligned} \tag{C.4}$$

It follows from a simple geometric argument that the second term of the right-hand side of Equation (C.4) is:

$$\mathbb{P}\{N_1^s > N_2^s\} = \frac{\left(\left(1+\frac{K_s}{2}\right)q_1^s - \left(1-\frac{K_s}{2}\right)q_2^s\right)^2}{2K_S^2 q_1^s q_2^s}.$$

The third term is given by Equation (C.2). In order to calculate the fourth term, let Y the following random variable:

$$Y = \begin{cases} \frac{N_1^s}{N_2^s} & \text{if } N_1^s > N_2^s, \\ 0 & \text{else.} \end{cases}$$

The expected value of Y is:

$$\mathbb{E}(Y) = \frac{1}{K_s^2 q_1^s q_2^s}\int_{(1-K_s/2)q_2}^{(1+K_s/2)q_1^s}\int_{y}^{(1+K_s/2)q_1^s}\frac{x}{y}dxdy.$$

By integrating along x, this equation becomes:

$$\mathbb{E}(Y) = \frac{1}{2K_s^2 q_1^s q_2^s}\int_{(1-K_s/2)q_2}^{(1+K_s/2)q_1^s}\left(\left(1+\frac{K_s}{2}\right)^2\frac{(q_1^s)^2}{y} - y\right)dy,$$

which yields:

$$\mathbb{E}(Y) = \frac{1}{2}\left(\frac{1}{K_s}+\frac{1}{2}\right)^2 \frac{q_1^s}{q_2^s}\ln\left(\frac{2+K_s}{2-K_s}.\frac{q_1^s}{q_2^s}\right) - \frac{1}{4}\left(\frac{1}{K_s}+\frac{1}{2}\right)^2\frac{q_1^s}{q_2^s} + \frac{1}{4}\left(\frac{1}{K_s}-\frac{1}{2}\right)^2\frac{q_2^s}{q_1^s}. \tag{C.5}$$

But from the definition of Y we have:

$$\mathbb{E}(Y) = \mathbb{E}\left(\frac{N_1^s}{N_2^s}\middle| N_1^s > N_2^s\right).\mathbb{P}\{N_1^s > N_2^s\}.$$

As a consequence, we may substitute Equation (C.5) into Equation (C.4) and obtain that when $(2-K_s)/(2+K_s).q_2^s \le q_1^s < q_2^s$:

$$\begin{aligned}\lambda_{LTL} = {} & \frac{2}{3} + \frac{\left(\left(1+\frac{K_s}{2}\right)q_1^s - \left(1-\frac{K_s}{2}\right)q_2^s\right)^2}{6K_S^2 q_1^s q_2^s} + \frac{1}{3K_s}\ln\left(\frac{2+K_s}{2-K_s}\right)\frac{q_1^s}{q_2^s} \\ & - \frac{1}{6}\left(\frac{1}{K_s}+\frac{1}{2}\right)^2\frac{q_1^s}{q_2^s}\ln\left(\frac{2+K_s}{2-K_s}.\frac{q_1^s}{q_2^s}\right) + \frac{1}{12}\left(\frac{1}{K_s}+\frac{1}{2}\right)^2\frac{q_1^s}{q_2^s} \\ & - \frac{1}{12}\left(\frac{1}{K_s}-\frac{1}{2}\right)^2\frac{q_2^s}{q_1^s}.\end{aligned} \tag{C.6}$$

We now proceed to the third case: $q_2^s \le q_1^s < (2+K_s)/(2-K_s).q_2^s$. Calculating λ_{LTL} is slightly easier in this case than in the previous one. Indeed, Equation (C.1) may be written as:

$$\lambda_{LTL} = 1 - \mathbb{P}\{N_1^s \le N_2^s\} + \left(\frac{2}{3}+\frac{1}{3}\mathbb{E}\left(\frac{N_1^s}{N_2^s}\middle| N_1^s \le N_2^s\right)\right).\mathbb{P}\{N_1^s \le N_2^s\}.$$

λ_{LTL} is then calculated as in the previous case. We directly provide the result. When $q_2^s \le q_1^s < (2+K_s)/(2-K_s).q_2^s$ we have:

$$\begin{aligned}\lambda_{LTL} = {} & 1 - \frac{\left(\left(1+\frac{K_s}{2}\right)q_2^s - \left(1-\frac{K_s}{2}\right)q_1^s\right)^2}{6K_S^2 q_1^s q_2^s} + \frac{1}{12}\left(\frac{1}{K_s}+\frac{1}{2}\right)^2\frac{q_2^s}{q_1^s} \\ & - \frac{1}{12}\left(\frac{1}{K_s}-\frac{1}{2}\right)^2\frac{q_1^s}{q_2^s} - \frac{1}{6}\left(\frac{1}{K_s}-\frac{1}{2}\right)^2\frac{q_1^s}{q_2^s}\ln\left(\frac{2+K_s}{2-K_s}.\frac{q_2^s}{q_1^s}\right).\end{aligned} \tag{C.7}$$

The fourth and final case proves to be the easiest. Indeed, when $q_1^s \ge (2+K_s)/(2-K_s).q_2^s$, the probability that N_1^s is larger than N_2^s is 1, so that:

$$\lambda_{LTL} = 1 \tag{C.8}$$

The expected load factor of the trucks carrying LTL shipments is given by Equations (C.3), (C.6), (C.7) and (C.8), depending on the values of q_1^s and q_2^s.

Q.E.D.

C.3 Proof of Proposition 6.17

Remind that the equilibrium is uniquely determined by Equation (6.14):

$$r^* = f_r(r^*).$$

f_r is continuous and derivable in b. From the implicit function theorem, so is r^*, and its derivative with respect to b verifies:

$$\frac{\mathrm{d}r^*}{\mathrm{d}b} = \frac{\partial f_r}{\partial r}.\frac{\mathrm{d}r^*}{\mathrm{d}b} + \frac{\partial f_r}{\partial b}$$

As a consequence:

$$\frac{\mathrm{d}r^*}{\mathrm{d}b} = \left(\frac{\partial f_r}{\partial b}\right) \Big/ \left(1 - \frac{\partial f_r}{\partial r}\right)$$

We know from the proof of Proposition 6.14 that $\partial f_r/\partial r$ is negative, which implies first that the derivative of r^* with respect to b always exists, second that its sign is the sign of $\partial f_r/\partial b$. Furthermore, remind that:

$$f_r(r) = \frac{f_{n_1^s}(p(1);p(2))}{f_{n_2^s}(p(1);p(2))},$$

In the following, we keep the same notations as in the remaining of the chapter. Anyway, the prices and transport demands should be considered as functions of r and b, even when not stated explicitly; it is from this standpoint that we write partial derivatives of these functions with respect to r and b. Furthermore, to simplify even further the notations, the previous equation is rewritten as follows:

$$f_r(r) = \frac{n_1^s}{n_2^s},$$

In order to calculate the first derivative of f_r with respect to b, we start from the most disaggregate level before consolidating. We begin with prices. From Equation System (6.9):

$$\begin{cases} p(1) &= b + \frac{\partial \delta}{\partial q_1^s}(n_1^s;n_2^s)c/3, \\ p(2) &= b + c + \frac{\partial \delta}{\partial q_2^s}(n_1^s;n_2^s)c/3, \\ p(3) &= b + c, \end{cases}$$

The partial derivatives of these functions with respect to b are constant and equal to 1. We may thus calculate $\partial n_i^s/\partial b$. Indeed, Equation (6.11) yields:

$$n_1^s = N.(1 - F_a(2p(1) - p(2))),$$

so that:

$$\frac{\partial n_1^s}{\partial b} = -2.N.f_a(2p(1) - p(2)) + N.f_a(2p(1) - p(2)),$$

or, equivalently:

$$\frac{\partial n_1^s}{\partial b} = -N.f_a(2p(1) - p(2)).$$

The first derivative of n_2^s with respect to r is calculated similarly:

$$\frac{\partial n_2^s}{\partial b} = \frac{N}{2}.f_a(2p(1) - p(2)) - \frac{N}{6}.f_a(p(2) - 2p(3)/3).$$

The last step is now to calculate the first derivative of f_r with respect to b:

$$\frac{\partial f_r}{\partial b} = \frac{(\partial n_1^s/\partial b).n_2^s - n_1^s.(\partial n_2^s/\partial b)}{(n_2^s)^2}.$$

This equation can be rewritten as follows:

$$\frac{\partial f_r}{\partial b} = \frac{(\partial n_1^s/\partial b) - r^*.(\partial n_2^s/\partial b)}{(n_2^s)^3},$$

so that the sign of $\partial f_r/\partial b$ is the sign of numerator, which is of the same sign as:

$$-\left(1 + \frac{r^*}{2}\right) f_a(2p(1) - p(2)) + \frac{r^*}{6}.f_a(p(2) - 2p(3)/3).$$

Q.E.D.

C.4 Proof of Proposition 6.18

This proof is similar to the proof of Proposition 6.17. Indeed:

$$\frac{\mathrm{d}r^*}{\mathrm{d}c} = \left(\frac{\partial f_r}{\partial c}\right) \Big/ \left(1 - \frac{\partial f_r}{\partial r}\right),$$

which implies that the sign of the derivative of r^* with respect to c is the sign of $\partial f_r/\partial c$.

The partial derivatives of the prices with respect to c are derived from Equation System (6.9) as follows:

$$\frac{\partial p(1)}{\partial c} = \frac{1}{3}\frac{\partial \delta}{\partial q_1^s},$$

and:

$$\frac{\partial p(2)}{\partial c} = 1 + \frac{1}{3}\frac{\partial \delta}{\partial q_2^s},$$

and:

$$\frac{\partial p(3)}{\partial c} = 1.$$

This straightforwardly implies that:

$$\frac{\partial n_1^s}{\partial c} = N.f_a(2p(1) - p(2))\left(1 - \frac{2}{3}\frac{\partial \delta}{\partial q_1^s} + \frac{1}{3}\frac{\partial \delta}{\partial q_2^s}\right),$$

and

$$\frac{\partial n_2^s}{\partial c} = -\frac{N}{2}f_a(2p(1) - p(2))\left(1 - \frac{2}{3}\frac{\partial \delta}{\partial q_1^s} + \frac{1}{3}\frac{\partial \delta}{\partial q_2^s}\right)$$
$$- \frac{N}{2}f_a(p(2) - 2p(3)/3)\left(\frac{1}{3} + \frac{1}{3}\frac{\partial \delta}{\partial q_2^s}\right)$$

From Lemma 6.10, $\partial n_1^s/\partial c$ is positive, and $\partial n_2^s/\partial c$ is negative. This ensures that $\partial f_r/\partial c$ is positive. Q.E.D.

Appendix D

Proofs of Chapter 7

D.1 Proof of Lemma 7.2

Remind that ζ is defined by Equation (7.23). The reason why ζ is symmetric is that:

$$\Phi^{-1}(x) = -\Phi^{-1}(1-x),$$

due to the fact that $\Phi(x) = 1 - \Phi(-x))$. As a consequence:

$$\zeta(y,x) = (y+x)\varphi \circ \Phi^{-1}\left(\frac{y}{y+x}\right)$$

is equivalent to:

$$\zeta(y,x) = (x+y)\varphi \circ \Phi^{-1}\left(\frac{x}{x+y}\right),$$

because first: $1 - x/(x+y) = y/(x+y)$, second: $\phi(x) = \phi(-x)$. Thus the symmetry of ζ.

Let us now consider the partial derivative of ζ with respect to x. First, observe that:

$$\frac{d\varphi}{dx} = -x\varphi(x),$$

so that:

$$\frac{d}{dx}\left(\varphi \circ \Phi^{-1}\right) = \frac{1}{\varphi \circ \Phi^{-1}}\left(-\Phi^{-1}\right)\varphi \circ \Phi^{-1},$$

which simplifies itself into:

$$\frac{d}{dx}\left(\varphi \circ \Phi^{-1}\right) = -\Phi^{-1}.$$

As a consequence:

$$\frac{\partial \zeta}{\partial x} = \varphi \circ \Phi^{-1}\left(\frac{x}{x+y}\right) - (x+y)\Phi^{-1}\left(\frac{x}{x+y}\right)\frac{y}{(x+y)^2},$$

which, due to the properties of φ and Φ presented just above, is equivalent to:

$$\frac{\partial \zeta}{\partial x} = \varphi \circ \Phi^{-1}\left(\frac{y}{x+y}\right) + \Phi^{-1}\left(\frac{y}{x+y}\right)\frac{y}{x+y},$$

Let $h = \Phi^{-1}(y/(x+y))$. Then note first that $h \in]0;1[$, second that:

$$\frac{\partial \zeta}{\partial x} = \varphi(h) + h\Phi(h),$$

so that $\partial\zeta/\partial x$ is positive. ζ is thus increasing in x, and, from its symmetry, in y. Q.E.D.

D.2 Approximation of the excess inventory expected value and variance

The evolution of the Dyck walk process over time is illustrated by Figure D.1 for $\delta = 0.2$ and by Figure D.2 for $\delta = 0.45$. The excess inventory is a kind of asymmetric Dyck walk (a Dyck walk is a random process which can increase of 1 or decrease of 1 at each time step with the same probability, but is always positive). While there are results on the expected value, distribution and other properties of Dyck walks, nothing could be found about processes behaving like the excess inventory.

The approximation of I_t^E's expected value and standard deviation are therefore based on a series of simulations. For δ going from 0.0 to 0.45 by steps of 0.05, the process is simulated over 10000 steps to reach its stationary distribution, then on 200000 steps on the basis of which the observed expected value and standard deviation are observed. The results are illustrated by Figures D.3 and D.4.

Given the shape of the solution and the theoretical property that $\mathbb{E}(I_t^E) = 0$ when $\delta = 0$, the following function was fitted to the expected value and standard deviation of the excess inventory, by minimising the sum of the differences square:

$$f(x) = A\delta + B\frac{\delta}{(0.5-\delta)^C}.$$

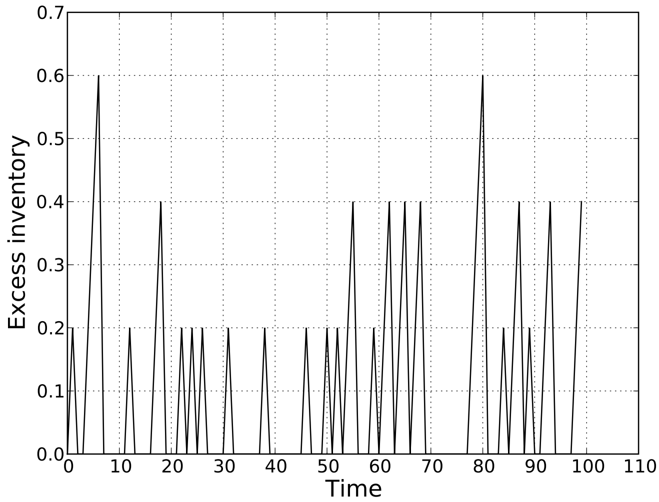

Figure D.1: Evolution of the excess inventory, $\delta = 0.2$

The estimation gave in both cases $C = 1.01$. As a consequence, $C = 1$ has been fixed to 1. The estimations obtained for A and B for the expected value and standard deviation of I_t^E are given in Equations (7.42) and (7.43). Graphically, they are hardly distinct from the simulations.

D.3 Proofs of Subsection 7.3.3.1

D.3.1 Proof of Lemma 7.12

First of all, note that from Equation (7.45), σ_{I^E}/d is given by:

$$\frac{\sigma_{I^E}}{d} = 0.43\left(\frac{d_h}{d} - 1\right) - 0.05\frac{\sigma}{d} + 0.48\frac{(\sigma/d)^2}{1 - d_h/d}, \tag{D.1}$$

and $\partial\sigma_{I^E}/\partial d_h$ is positive and given by:

$$\frac{\partial\sigma_{I^E}}{\partial d_h} = 0.43 + \frac{0.48(\sigma/d)^2}{(1 - d_h/d)^2}, \tag{D.2}$$

which implies they are both increasing and convex with d_h/d.

Then, note that from Equation (7.56), we have:

$$\xi\left(\frac{d_h}{d}, \frac{\sigma}{d}, t_l\right) = \frac{\sigma_{I^E}/d}{\sqrt{(t_l + 1)(\sigma/d)^2 + (\sigma_{I^E}/d)^2}}\frac{\partial\sigma_{I^E}}{\partial d_h}.$$

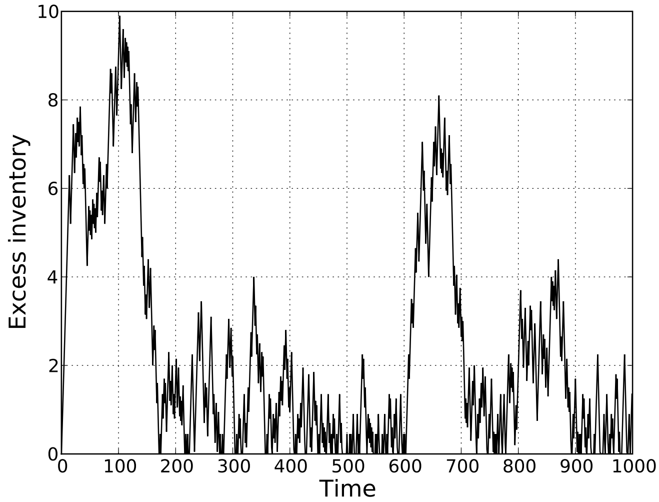

Figure D.2: Evolution of the excess inventory, $\delta = 0.45$

This observation, combined with the fact that $x \mapsto x/\sqrt{K+x^2}$ with $K > 0$ is an increasing function of x implies that ξ is an increasing function of d_h/d.

Let us now consider the variation of ξ with σ/d. From Equation (D.2), $\partial\sigma_{I^E}/\partial d_h$ increases with σ/d. A more complicated task is to identify the behaviour of σ_{I^E}/σ_I. To do so, first observe that:

$$\frac{\sigma_{I^E}}{\sigma_I} = \frac{\sigma_{I^E}/\sigma}{\sqrt{(t_l+1)+(\sigma_{I^E}/\sigma)^2}}, \tag{D.3}$$

which implies that σ_{I^E}/σ_I and σ_{I^E}/σ vary together with σ/d.

Consider now σ_{I^E}/σ. From Equation (7.45) we have:

$$\frac{\sigma_{I^E}}{\sigma} = 0.43\left(\frac{d_h}{\sigma} - \frac{d}{\sigma} + 1\right) + \frac{0.48}{d/\sigma - d_h/\sigma}\left(\frac{d_h}{\sigma} - \frac{d}{\sigma} + 1\right),$$

this equation is rewritten as follows, so that σ_{I^E}/σ is written as a function of d_h/d and σ/d:

$$\begin{aligned}\frac{\sigma_{I^E}}{\sigma} &= 0.43\left(\left(\frac{d_h}{d} - 1\right)\frac{d}{\sigma} + 1\right) \\ &\quad + \frac{0.48}{d/\sigma\,(1 - d_h/d)}\left(\left(\frac{d_h}{d} - 1\right)\frac{d}{\sigma} + 1\right).\end{aligned}$$

Given that $d_h < d$, we deduce that σ_{I^E}/σ increases with σ/d. From Equation (D.3), this implies that σ_{I^E}/σ increases with σ/d. Finally,

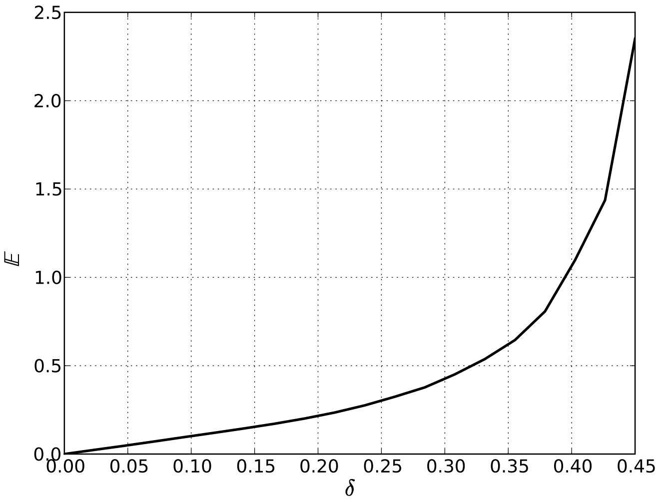

Figure D.3: Estimation of $\mathbb{E}\left(I_t^E\right)$

given that $\partial\sigma_{I^E}/\partial d_h$ also increases with σ/d, and is positive, ξ increases with σ/d.

The monotonicity of ξ w.r.t. t_l is simply derived: σ_{I^E} thus $\partial\sigma_{I^E}/\partial d_h$ do not depend on t_l, and σ_I increases with t_l. As a consequence, given Equation (7.56), ξ decreases with t_l. Q.E.D.

D.3.2 Proof of Lemma 7.13

First of all, let $\delta = d_h - d + \sigma$, so that $d_h = d - \sigma + \delta$. From Equation (D.1), σ_{I^E}/d is given by:

$$\frac{\sigma_{I^E}}{d} = 0.43\frac{\delta}{d} - 0.48\frac{\sigma}{d} + \frac{0.48(\sigma/d)^2}{\sigma/d - \delta/d},$$

or, for small values of δ/d:

$$\frac{\sigma_{I^E}}{d} = 0.91\frac{\delta}{d} + o\left(\frac{\delta}{d}\right). \tag{D.4}$$

Given that the derivative of $x \mapsto \sqrt{K+x}$ is zero for $x = 0$ and $K > 0$, and considering Equations (7.40) and (D.4), the derivative of σ_I when d_h/d is close to $1 - \sigma/d$ is zero, so that we simply have:

$$\frac{\sigma_I}{d} = \frac{\sqrt{t_l+1}\,\sigma}{d} + o\left(\frac{\delta}{d}\right).$$

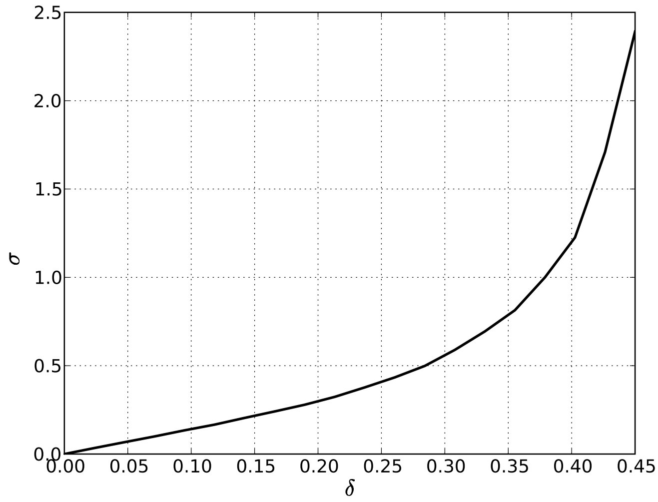

Figure D.4: Estimation of $\sigma\left(I_t^E\right)$

From Equation (D.2), the partial derivative of σ_{I^E} can be written:

$$\frac{\partial \sigma_{I^E}}{\partial d_h} = 0.43 + \frac{0.48\sigma^2}{(\sigma-\delta)^2},$$

so that its Taylor approximation of degree one is:

$$\frac{\partial \sigma_{I^E}}{\partial d_h} = 0.91 + 0.96\frac{\delta/d}{\sigma/d} + o\left(\frac{\delta}{d}\right).$$

These intermediate steps are quickly combined using Equation (7.56):

$$\begin{aligned} \xi\left(\frac{d_h}{d}, \frac{\sigma}{d}, t_l\right) &= \left(\frac{0.91\delta/d}{\sqrt{t_l+1}\,\sigma/d} + o\left(\frac{\delta}{d}\right)\right) \\ &\quad \times \left(0.91 + 0.96\frac{\delta/d}{\sigma/d} + o\left(\frac{\delta}{d}\right)\right), \end{aligned}$$

The final step consists in replacing δ and using a straightforward simplification. Q.E.D.

D.3.3 Proof of Lemma 7.15

When d_h tends toward d, σ_{I^E} increases undefinitely. As a consequence:

$$\sqrt{(t_l+1)\sigma^2 + \sigma_{I^E}^2} \sim_d \sigma_{I^E},$$

where $f(x) \sim_a g(x) \Leftrightarrow \lim_{x\to a} f(x)/g(x) = 1$. so that:

$$\frac{\sigma_{I^E}}{\sigma_I} \sim_d 1.$$

Besides, from Equation (D.2):

$$\frac{\partial \sigma_{I^E}}{\partial d_h} \sim_d \frac{0.48\sigma^2}{(d-d_h)^2},$$

so that from Equation (7.56) we have the result. Q.E.D.