TØI report 1538/2016

Elise Caspersen
Bjørn Gjerde Johansen
Inger Beate Hovi
Gerard de Jong

Institute of Transport Economics
Norwegian Centre for Transport Research

Norwegian Logistics Model:

Moving from a deterministic framework to a random utility model

TØI-report
1538/2016

Norwegian Logistics Model:

Moving from a deterministic framework to a random utility model

Elise Caspersen, Bjørn Gjerde Johansen, Inger Beate Hovi, Gerard de Jong

Front page photography: Shutterstock

ISSN 0808-1190
ISBN 978-82-480-1814-8 Electronic version
Oslo, December 2016

Tittel:	Nasjonal godstransportmodell: Fra et deterministisk rammeverk til en stokastisk modell	**Title:**	Norwegian Logistics Model: Moving from a deterministic framework to a random utility model
Forfattere:	Elise Caspersen, Bjørn Gjerde Johansen, Inger Beate Hovi, Gerard de Jong	**Authors:**	Elise Caspersen, Bjørn Gjerde Johansen, Inger Beate Hovi, Gerard de Jong
Dato:	12.2016	**Date:**	12.2016
TØI-rapport:	1538/2016	**TØI Report:**	1538/2016
Sider:	43	**Pages:**	43
ISBN elektronisk:	978-82-480-1814-8	**ISBN Electronic:**	978-82-480-1814-8
ISSN:	0808-1190	**ISSN:**	0808-1190
Finansieringskilder:	Statens vegvesen Vegdirektoratet, Jernbaneverket, Kystverket	**Financed by:**	Norwegian Public Road Administration Norwegian Railway Administration
Prosjekt:	4137 Stokastisk godsmodell	**Project:**	4137 Stochastic logistics model
Prosjektleder:	Inger Beate Hovi	**Project Manager:**	Inger Beate Hovi
Kvalitetsansvarlig:	Kjell Werner Johansen	**Quality Manager:**	Kjell Werner Johansen
Fagfelt:	Næringslivets godstransporter	**Research Area:**	Industry and freight
Emneord:	Langtransport Nasjonal godstransportmodell Diskret valgmodell Transportmiddelvalg	**Keywords:**	Long distance freight transport Logistics model Discrete choice model Mode choice

Sammendrag:

I foreliggende rapport beskrives arbeid knyttet til estimering av økonometriske modeller for det simultane valget av transportmiddel og forsendelsesstørrelse, samt hvordan slike modeller kan brukes til å forbedre den nasjonale godstransportmodellen. Arbeidet har bestått i å estimere diskrete valgmodeller for tre forskjellige varegrupper. Datagrunnlaget har vært forsendelsesdata fra den svenske varestrømsundersøkelsen for 2009. Estimerte koeffisienter og tilhørende funksjoner for utvalgskriterier har blitt implementert i det norske godstransportmodellsystemet via en ny versjon av applikasjonen ChainChoice. Vi har validert modellen ved å sammenligne etterspørselselastisiteter for den stokastiske modellen med tilsvarende for den deterministiske modellen.

Summary:

In the present report we describe work related to estimating an econometric model for the choice of transport mode and shipment size, and how this econometric model can be used to improve the Norwegian Logistics Model. We have estimated discrete choice models for three different commodity groups using data from the 2009 Swedish Commodity Flow Survey. The estimated coefficients and their corresponding selection criteria functions have been implemented in the Norwegian Logistics Model by implementing a new version of the executable ChainChoice program. We validated the model by comparing the deterministic model to the stochastic model through the demand elasticities for the three transport modes road, sea and rail, with respect to changes in time and distance-based link costs.

Language of report: English

Transportøkonomisk Institutt
Gaustadalleen 21, 0349 Oslo
Telefon 22 57 38 00 - www.toi.no

Institute of Transport Economics
Gaustadalleen 21, 0349 Oslo, Norway
Telefon 22 57 38 00 - www.toi.no

Preface

When the logistics model for Norway was first conceived, the idea was to estimate a deterministic model based on data from the Swedish Commodity Flow Survey and a couple of logistics providers in Norway and later to develop a stochastic model. However, since a deterministic logistics module is complex and estimation of disaggregate models take a significant amount of time, a 'preliminary' or 'prototype' version of the logistics model was developed. Despite several improvements, the model has remained deterministic, relying on cost minimization. Hence, it seems to be a sensible time to move from a deterministic framework to a random utility model for the Norwegian Logistics Model.

This project is conducted as a pre-project for the National Transport Authorities with the purpose of introducing a stochastic element in the National Logistics Model. In the project we estimated econometric models for the choice of transport mode and shipment size based on basic data from the Swedish Commodity Flow Survey, for transports between Norway and Sweden.

The estimation work has been carried out by Elise Caspersen, with supervision from Gerard de Jong (director at Significance in the Netherlands). She has also written the major part of the report. Bjørn Gjerde Johansen has carried out the validation of the model and written chapter 6. Jaap Baak (Significance) has calculated and facilitated transportation costs for alternative transport chains and done the implementation of the stochastic element in the logistics module. Inger Beate Hovi has been project leader and has obtained the basic data, organized the work and been discussion partner during the work. Oskar Andreas Kleven has been the client's contact person.

Oslo, December 2016
Institute of Transport Economics

Gunnar Lindberg
Managing Director

Kjell Werner Johansen
Deputy Managing Director

Content

Summary

Norwegian Logistics Model: Moving from a deterministic framework to a random utility model

TØI Report 1538/2016

Authors: Elise Caspersen, Bjørn Gjerde Johansen, Inger Beate Hovi, Gerard de Jong

Oslo 2016 43 pages English language

In the present report we describe work related to estimating an econometric model for the choice of transport mode and shipment size, and how this econometric model can be used to improve the Norwegian Logistics Model. We have estimated discrete choice models for three different commodity groups using data from the 2009 Swedish Commodity Flow Survey. The estimated coefficients and their corresponding selection criteria functions have been implemented in the Norwegian Logistics Model by creating a new version of the executable ChainChoice program. We validated the model by comparing the deterministic model to the stochastic model through the demand elasticities for the three transport modes road, sea and rail, with respect to changes in time and distance-based link costs. The comparison showed that all own elasticities and all cross elasticities for sea and rail have the expected sign. For two of the commodity groups, the own elasticities in the stochastic case are consistently lower than the own elasticities in the deterministic case, while there is no clear pattern for the third commodity group.

Introduction

The purpose of this report is twofold. First, we describe the work related to estimating an econometric model for the choice of transport mode and shipment size in Norway, and then how this econometric model can be used to improve the Norwegian Logistics Model. The models presented in this report are estimated for three of the commodity groups in the current Norwegian Logistics Model. These are commodity group 13 “Iron and steel”, 17 “Plastic and rubber” and 30 “Consumables”. The models are estimated on shipment level data from the 2009 Swedish commodity flow survey (CFS).

When the logistics model for Norway was first conceived, the idea was to estimate a model based on data from the Swedish Commodity Flow Survey and a couple of logistics providers in Norway. However, since a deterministic logistics module is complex and estimation of disaggregate models take a significant amount of time, a ‘preliminary’ or ‘prototype’ version of the logistics model was developed (see de Jong and Ben-Akiva, 2007, section 8) in 2005-2006. Despite several improvements, the model has remained deterministic, relying on a cost minimisation procedure per firm-to-firm (f2f) to find preferred transport mode and shipment size. Recently, studies have shown that this assumptions is not always valid. Hence, it seems to be a sensible time to move from a deterministic framework to a random utility model for the Norwegian Logistics Model.

Econometric specification of a stochastic logistics model

One of the main reasons to move away from a deterministic model is that a deterministic model has weak empirical foundation in observed behaviour. In addition, given a deterministic cost-minimizing model it is difficult to get full information about all cost elements and other factors transport agents consider when making their choice. Also, if the relevant part of the logistics costs function is rather flat, only a small change in logistics costs can result in a shift to a completely different optimal shipment size and transport chain. Some of these issues can be solved by estimating disaggregate random utility models with available RP data. By their nature, such models are probabilistic models, because they include a stochastic component to account for the influence of omitted factors. We start the process of moving to a random utility model by estimating a joint decision model for three of the commodities in the Norwegian Logistics Model. Every commodity type for which we can base the choice mechanism on observed RP data, constitutes an improvement relative to the deterministic model.

The chosen estimation approach is a joint model with discrete mode and discrete shipment size choice. The advantage with this model specification is that it does not require a combination of techniques from discrete choice and regression analysis, but can be estimated fully within the discrete choice framework. This is also the chosen estimation technique in most of the existing studies in the field. The drawback is that if the choice of shipment size in reality is a continuous one, the approach may lead to measurement errors. In this case, an alternative model specification is a model with discrete transport mode and continuous shipment size. For our chosen model, one might overcome this problem by estimating a nested or a cross-nested logit model.

Data analysis and model approach

The dataset used for estimations is shipments between Norway and Sweden registered in the 2009 Swedish CFS. In total the data set covers 105,533 shipments between Norway and Sweden. For each shipment, the data contains information regarding size, value, mode of transportation, commodity group, and geographical location of senders and receivers of goods, with municipality as the lowest geographical level. Information about transport cost is missing, and needs to be calculated. This is done for both each observed shipment mode choice and its alternatives using the Norwegian Logistics Model. To estimate cost data, we need information about sending and receiving zones in both Norway and Sweden. Observations lacking information about one or both of these variables are excluded from the dataset.

The choice set used in the estimation is determined by all chosen combinations of transport chains and shipment size categories in the CFS, under the restriction that cost data can be estimated. For commodity group 13 Iron and steel 44 % of the shipments are transported by road, and 55 % by road-rail-road. Majority of the shipments weight more than 50 ton. For commodity group 17 Plastic and rubber 61 % is transported by road, while 38 % is transported by rail. Less than 1 % is transported by air or waterborne transport. Commodity group 17 is transported in smaller shipments than iron and steel. The choice set for commodity 30 Consumables consists of 12 available chain choice alternatives. This is the only (out of three) commodity group for which we have a sufficient number of observations to include water as a transport mode choice (in combination with road/rail).

Exogenous variables are transport cost, time use, degradation and capital cost, value density, inventory in transit and region specific dummies.

Discrete choice model estimation for the joint decision of transport mode and shipment size

Estimations are carried out for the three commodity groups using the 2009 Swedish Commodity Flow Survey, the defined choice set and a set of explanatory variables. All models are estimated with the software Biogeme (Bierlaire, 2003). In addition to the three multinomial logit models, we extend the estimation of commodity group 17 to a model where the utility functions are non-linear in cost and time. This extended model is estimated as both a multinomial logit and a nested logit model. This was not done for the other two commodity groups. Commodity 13 lacks observations for the alternative with small shipments transported by rail, while commodity 30 contains so many observations that extending the model led to a severe slowdown of the estimation in Biogeme.

We were not able to estimate a reliable nested logit model with non-linearity in the utility function. Since the multinomial logit model is a restricted version of the nested logit model, this also gives reasons to doubt the estimation of the multinomial model with non-linear variables. For this reason, we chose to keep the simpler models for the discrete choice, where time and cost are linear in the utility function, instead of proceeding with extended model versions.

Results from the Norwegian Logistics Model

The estimated coefficients and their corresponding selection criteria functions was implemented in the Norwegian Logistics Model by creating a new version of the executable ChainChoice program. This executable uses the same input data, but applies new selection criteria according to a logistic specification based on the estimated coefficients. The output of this executable is also slightly different; instead of one (deterministic) mode/weight combination for each freight flow, the model now outputs each potential (considered) choice alternative (i.e. each potential combination of mode choice and shipment size) as well as the predicted probabilities that these particular modes and shipment sizes are chosen jointly. Summing over the potential freight flows, multiplied by their corresponding probabilities for each mode, will therefore give the expected mode split predicted by the model. If, for each freight flow, the probabilities of all but one alternative are zero, the result of the stochastic model will be equivalent to the deterministic model.

We attempt to validate the model by comparing the deterministic model to the stochastic model through the demand elasticities for the three transport modes road, sea and rail, with respect to changes in time and distance-based link costs. The main findings are:

- All own elasticities and all cross elasticities for sea and rail have the expected signs. This is not the case for cross-elasticities for road transport, since road also is a part of the transport chains defined as “sea” and “rail”. For all commodities and both model types, changing the costs for sea will to a large extent lead to a mode shift to/from rail.
- For commodity groups 17 and 30 own elasticities in the stochastic case are consistently lower than the own elasticities in the deterministic case while there is no clear pattern for commodity group 13. We would expect the model to be less responsive.
- For virtually all commodity groups, all transport modes and both types of models, for both own elasticities and cross elasticities, the absolute value of the elasticity increases when the magnitude of the cost change increases. This relationship is stronger than expected.

Discussion and further work

Reducing the data set from the Swedish commodity flow survey to cover only the transport between Norway and Sweden, result in a minor number of commodities that fulfil the requirements regarding number of shipments and variation across transport modes and geographical areas. A possible way to validate the estimated choice models is to estimate the models using other commodity groups. A drawback with this approach is that most of the other commodity groups in the dataset have either too few observations and/or too little spread of transport modes and/or geographical areas to perform estimations. To increase the number of observations and the spread of mode choice and geography, the dataset from the Swedish commodity flow survey must either be increased to cover more than transport between Norway and Sweden, or be replaced with another data source. An alternative data source recently available is the Norwegian foreign trade statistics, that TOI has access to at shipment level, with detailed information about origin, destination, shipment size and value, and border crossing mode choice.

As discretising the choice of shipment size can be seen as a form of measurement error an alternative approach is to estimate a simultaneous discrete-continuous structural model with a joint choice of discrete transport mode and continuous shipment size.

Sammendrag

Nasjonal godstransportmodell: Fra et deterministisk rammeverk til en stokastisk modell

TØI rapport 1538/2016
Forfattere: Elise Caspersen, Bjørn Gjerde Johansen, Inger Beate Hovi, Gerard de Jong
Oslo 2016, 43 sider

I foreliggende rapport beskrives arbeid knyttet til estimering av økonometriske modeller for det simultane valget av transportmiddel og forsendelsesstørrelse, samt hvordan slike modeller kan brukes til å forbedre Nasjonal godstransportmodell. Arbeidet har bestått i å estimere diskrete valgmodeller for tre forskjellige varegrupper. Datagrunnlaget har vært forsendelsesdata fra den svenske varestrømsundersøkelsen for 2009. Estimerte koeffisienter og tilhørende funksjoner for utvalgskriterier har blitt implementert i det norske godstransportmodellsystemet via en ny versjon av applikasjonen ChainChoice. Vi har validert modellen ved å sammenligne etterspørselselastisiteter for den stokastiske modellen med etterspørselselastisiteter for den deterministiske modellen. Elastisitetene er beregnet med hensyn til endringer i tid og distansebaserte kostnader. Sammenligningen viste at alle egenelastisiteter og krysselastisiteter for sjø og bane har forventet fortegn. For to av varegruppene finner vi at egenelastisitetene i den stokastiske modellen er gjennomgående lavere enn egenelastisitetene i den deterministiske modellen. For den tredje varegruppen finner vi ingen klare mønstre.

Introduksjon

Hensikten med denne rapporten er todelt. Først beskriver vi arbeidet knyttet til estimering av økonometriske modeller for valg av transportmiddel og forsendelsesstørrelse. Deretter beskriver vi hvordan modellene kan brukes til å forbedre nasjonal godstransportmodell. Modellene som presenteres i denne rapporten er beregnet for tre av varegruppene i dagens godstransportmodell. Dette er varegruppene 13 «Jern og stål», 17 «Plast og gummi» og 30 «Forbruksvarer». Modellene er estimert på forsendelsesdata fra den svenske varestrømsundersøkelsen fra 2009, for transporter mellom Sverige og Norge.

Ved utviklingen av den norske godstransportmodellen tidlig på 2000-tallet, var den opprinnelige planen å estimere en transportvalgmodell basert på data fra den svenske varestrømsundersøkelsen og noen samlasterselskaper i Norge. Ettersom en deterministisk modell er kompleks og estimeringen av disaggregerte modeller er tidkrevende, ble en «foreløpig» versjon eller en «prototype» av godsmodellen utviklet (se de Jong og Ben-Akiva, 2007, seksjon 8) i 2005-2006. Prototypen ble utviklet som en deterministisk modell, hvor transportmiddel og forsendelsesstørrelse ble valgt i henhold til prinsippet om kostnadsminimering. Til tross for flere videreutviklinger og forbedringer i senere tid har modellen forblitt deterministisk. Nokså nylige studier viser derimot at forutsetningene som hviler bak en deterministisk modell ikke alltid er gyldige for godstransportmodeller. Det synes derfor å være på tide at godsmodellen utvikles fra et deterministisk rammeverk til en stokastisk modell.

Økonometrisk spesifikasjon av en stokastisk logistikkmodell

Noen av de viktigste grunnene til å bevege seg bort fra en deterministisk modell som utelukkende tar hensyn til kostnader ved valg av transportmiddel og forsendelsesstørrelser er at; en deterministisk modell har svakt empirisk grunnlag i observert atferd; det er vanskelig å få full informasjon om alle kostnadselementer og andre faktorer aktører tar hensyn til når de vurderer valg av transportmiddel og forsendelsesstørrelse; dersom den relevante delen av kostnadsfunksjonene er tilnærmet «flat», kan kun en liten endring i logistikkostnadene føre til et skift som resulterer i en annen optimal løsning enn før endring.

Valgt estimeringsmetode er en økonometrisk valgmodell som muliggjøre simultane valg av transportmiddel og forsendelsesstørrelse, hvor vi behandler både transportmiddelvalg og forsendelsesstørrelse som diskrete variabler. Fordelen med denne modellspesifikasjonen er at den ikke krever en kombinasjon av ulike teknikker fra diskrete valg- og regresjonsanalyse, men kan estimeres innenfor rammeverket for diskrete valgmodeller. Ulempen med nevnte metodikk er at dersom aktøren egentlig står overfor en kontinuerlig variabel for forsendelsesstørrelse, kan tilnærmingen gi målefeil.

Dataanalyse og modelltilnærming

Datasettet som brukes til modellestimering er forsendelser mellom Norge og Sverige registrert i den svenske varestrømsundersøkelsen fra 2009. Totalt inneholder datasettet 105 533 forsendelser mellom Norge og Sverige. For hver forsendelse inneholder dataene informasjon om størrelse, verdi, transportmiddel, varegruppe, og geografisk plassering av avsendere og mottakere av varer, med kommune som laveste geografiske nivå. Transportkostnadene mangler, og må dermed beregnes. Dette gjøres både for observerte valg og dets alternative transportløsninger. For å estimere transportkostnadene trenger vi informasjon om avsender- og mottakersone i Norge og Sverige. Observasjoner som mangler denne informasjon ekskluderes fra datasettet.

Valgsettene gis av alle valgte kombinasjoner av transportmiddelvalg og forsendelsesstørrelse som finnes i datasettet, gitt at kostnadsdata kan beregnes. For varegruppe 13 «Jern og stål» transporteres 44% av forsendelser med bil, og 55% av transportkjeder med bil og jernbane. Resterende 1 % transporteres med fly eller sjøtransport. Flertallet av forsendelsene veier mer enn 50 tonn. For varegruppe 17 «Plast og gummi» transporteres 61% med bil, mens 38% fraktes med bil og jernbane. Mindre enn 1% transporteres med fly eller sjøtransport. Varegruppe 17 blir transportert i mindre forsendelser enn jern og stål. For varegruppe 30 «Forbruksvarer» består valgsettet av 12 alternativer. Dette er den eneste (av tre) varegrupper hvor transport på sjø inkluderes i valgsettet (i kombinasjon med bil/jernbane).

Eksogene variable er transportkostnader, tidsbruk, degraderings- og kapitalkostnader, verditetthet (SEK/kg), kostnader ved å ha varer i transitt og regionspesifikke dummyvariabler.

Estimering

Estimeringen er gjort for de tre varegruppene 13, 17 og 30 basert på data fra den svenske varestrømsundersøkelsen for 2009, definerte valgsett og forklaringsvariabler. All estimering er gjort i programvaren Biogeme (Bierlaire, 2003). For varegruppe 17 utvidet vi modellen ved å innføre ikke-linearitet i nyttefunksjonene via variablene for tidsbruk og transportkostnader. Den utvidede modellen ble estimert både som en multinomisk og som en nestet logit-modell. Vi var ikke i stand til å estimere en pålitelig nestet logit-modell for denne modellspesifikasjonen. Dette gir også grunn til å tvile på estimeringen av den tilhørende multinomiske modellen, ettersom denne er en forenkling av nested logit-modellen.

Resultater fra implementering

Estimerte koeffisienter og tilhørende funksjoner for utvalgskriterier ble implementert i Nasjonal godstransportmodell ved å lage en ny versjon av den kjørbare applikasjonen ChainChoice. Applikasjonen bruker samme inndata som før, men har nye utvalgskriterier i henhold til en logistisk spesifikasjon som igjen baserer seg på estimerte koeffisienter. Resultater fra bruk av nye ChainChoice blir noe annerledes enn tidligere; i stedet for én valgt (deterministisk) kombinasjon av transportmiddel og forsendelsesstørrelse for hver varestrøm, presenterer modellen hver kombinasjon av transportmiddel og forsendelsesstørrelse som er aktuell for de enkelte varestrømmene, samt predikerte sannsynligheter for at hvert alternativ velges. Dersom hver varestrøm multipliseres med tilhørende valgsannsynligheter for hvert alternativ og summeres, får man transportmiddelfordelingen som predikert av modellen. Hvis modellen, for enhver varestrøm, beregner sannsynligheter lik null for alle alternativer bortsett fra ett, vil resultatet av den stokastiske modellen være ekvivalent med den deterministiske (opprinnelige) modellen.

Modellen er forsøkt validert ved å sammenligne etterspørselselastisiteter fra den deterministiske modellen med etterspørselselastisiteter fra den stokastiske modellen. Etterspørselselastisiteter er beregnet for de tre transportformene vei, sjø og bane, med hensyn til endringer i tid og distansebaserte kostnader. De viktigste funnene er:

- Alle egenelastisiteter og krysselastisiteter for sjø og bane har forventede fortegn. Dette er ikke tilfellet for krysselastisiteter for veitransport, ettersom «vei» også er en del av transportkjeder definert som «sjø» og «bane».
- Egenelastisiteter for varegruppene 17 og 30 i den stokastiske modellen er gjennomgående lavere enn egenelastisitetene i den deterministiske modellen. Dette er ikke like entydig for varegruppe 13.
- For nesten alle varegrupper, transportformer og begge modellversjoner, finner vi at absoluttverdien av både egen- og krysselastisiteten øker når omfanget av kostnadsendringer øker. Denne sammenhengen er sterkere enn forventet.

Diskusjon og videre arbeid

Ved å redusere datasettet fra den svenske varestrømsundersøkelsen til kun å dekke transporter mellom Norge og Sverige, sitter vi igjen med få varegrupper som oppfyller kravene til antall forsendelser og variasjon både på tvers av transportformer og geografiske områder. For å øke antall observasjoner og spredning over transportformer og geografi, kan vi enten bruke data fra den svenske varestrømsundersøkelsen utover transporter mellom Norge og Sverige, eller erstatte den svenske varestrømsundersøkelsen med andre datakilder. En alternativ datakilde er den norske utenrikshandelsstatistikken, som TØI har tilgang til på forsendelsesnivå i tilknytting til arbeidet med nye varestrømsmatriser. Dataene inneholder detaljert informasjon om opprinnelse, destinasjon, forsendelsesstørrelse og verdi, samt transportmiddel ved grensepassering.

1 Introduction

The purpose of this report is twofold. First, we describe the work related to estimating an econometric model for the choice of transport mode and shipment size in Norway, and then how this econometric model can be used to improve the Norwegian Logistics Model. The models presented are estimated for three of the commodity groups in the current Norwegian Logistics Model. These are commodity group 13 “Iron and steel”, 17 “Plastic and rubber” and 30 “Consumables”. We have estimated the models on shipment level data from the 2009 Swedish commodity flow survey.

When estimating a model for choice of transport mode and shipment size we assume that the choices are made simultaneous. Acknowledging that we do not possess full information regarding the factors that influence the choice of each shipper, we also assume that the joint logistics cost minimization choice of transport mode and shipment size are affected by a (mean zero) random error term. This approach leads to an econometric model specifications of discrete choice that needs to be estimated. The resulting econometric model can predict the probabilities for choosing each of the available combinations of transport modes and shipment sizes.

1.1 Background

When the logistics model for Norway was first conceived, the idea was to estimate a model based on data from the Swedish Commodity Flow Survey (CFS) and a couple of logistics providers in Norway. However, since a deterministic logistics module is complex and estimation of disaggregate models take a significant amount of time, a ‘preliminary’ or ‘prototype’ version of a cost minimizing deterministic logistics model was developed (see de Jong and Ben-Akiva, 2007, section 8) in 2005-2006. During the following years, the prototype of the Norwegian Logistics Model has been improved through numerous iterations, including a calibration that aggregate data to a base year. However, it has remained a deterministic model relying on a cost minimisation procedure per firm-to-firm (f2f).

Quite recently, studies have shown that the assumption that shipment size is chosen to minimize transport costs, given the chain choice, is not always valid: the choice of shipment size will often affect the transport chain as well. Johnson and de Jong (2011) point out that mode and shipment size are closely linked decisions, for which large shipment sizes generally coincide with higher market shares for non-road transport. Based on the 2004/2005 Swedish Commodity Flow Survey, Windisch et al. (2011) and Abate et al. (2014) find similar results through the estimation of joint econometric models of freight transport for mode choice and shipment size. These findings indicate a need to investigate the possibilities for updating Logistics Models to consider the relationship between shipment size and transport mode when allocating freight flows in a network. For further information, we refer to some of the most relevant papers for this research:

- Windisch (2009) that includes discrete-discrete models for the Swedish parts of domestic and export flows, estimated on the CFS 2004/2005 (also described in Windisch et al., 2010).

- Johnson and de Jong (2011) with discrete-discrete models and discrete-continuous models, both for the Swedish parts of domestic and export flows, estimated on the CFS 2001 (this paper also reports some earlier work by Johnson and de Jong, presented at earlier conferences: de Jong, 2007 and de Jong and Johnson, 2009).
- Abate et al. (2014) also includes both types of models, but now for the full outgoing domestic and international chains, and estimated on the CFS 2004/2005, for specific commodities (metal products, chemical products).
- Abate et al. (2016), building on the previous paper, but also discussing implementation of the estimation results into a network model, as well as calculating cost elasticities.

Note that these works (de Jong, 2007; Windisch, 2009; Abate et al, 2014) estimated both mode and shipment size as discrete choices, but clearly stated that shipment size is a continuous variable[1]. Johnson and de Jong (2011) noted that both assuming independence between mode and shipment size choice, and discretizing the continuous information on shipment size, may be interpreted as forms of specification error. Potential errors are discussed more in depth in chapter 2.

The work presented in this document is part of a pilot project; if deemed successful, it is recommended to carry out a larger project, in which transport mode and shipment size decisions in the Norwegian Logistics Model are updated based on econometric model estimates.

1.2 Report outline

The structure of this report is as follows: chapter 2 elaborates on theoretical aspects related to specifying and estimating econometric models of mode choice and shipment size. Section 2.1 give a brief introduction to the Norwegian Logistics Model; section 2.2 argues why we need a Logistics Model based on econometric estimates; section 2.3 contains suggestions for various model specifications; section 2.4 describes potential problems related to these specifications; and finally, section 2.5 reviews the various steps required in order to implement econometric estimates as an operational part of the Norwegian Logistics Model. Chapter 3 describes the utilized dataset. Section 3.1 discusses the main dataset used for estimation, i.e. the Swedish commodity flow survey from 2009. Section 3.2 discusses the cost data used as main explanatory variables; this data is not observed but rather predicted using the logistics model. Chapter 4 elaborates on the model approach, with section 4.1 including a description of the choice sets for each commodity in the model and section 4.2 describing the explanatory variables used for estimation. Chapter 5 presents the estimation results. Section 5.1 presents result from the final model, while section 5.2 contains a discussion of extensions of model 17. Chapter 6 presents results from the Norwegian Logistics Model after inclusion of the estimation. Section 6.1 describes implementation of the estimated coefficients and their corresponding selection criteria functions in to Norwegian Logistics Model, while section 6.2 presents a calculation and discussion of cost elasticities. Finally, chapter 7 gives a discussion of model validation on other commodity groups and provides suggestions for further work.

---

[1] Assuming that the error terms are independently and identically Generalized Extreme Value Type I distributed leads to the well-known multinomial logit formulation.

2 Econometric specification of a stochastic logistics model

In this chapter, we provide a conceptual description of the econometric models. Section 2.1 briefly introduces the Norwegian Logistics Model and its history, section 2.2 argues why using econometric estimates of transport agents' preferences can improve the Norwegian Logistics Model; section 2.3 describes two potential estimation frameworks for an econometric model; section 2.4 discusses the aforementioned problem of error term correlation; and finally, section 2.5 describes the next steps that need to be taken before estimation results from an econometric model can be implemented as an operational part of the Norwegian Logistics Model. The actual estimation of these models is described in Chapter 5.

2.1 A brief introduction to the Norwegian Logistics Model

When the logistics model for Norway within the aggregate-disaggregate-aggregate (ADA)[2] framework was first conceived, the idea was to estimate the model based on shipment level data from the Swedish Commodity Flow Survey (CFS), and on data from one or more of the largest logistics service providers in Norway (see de Jong and Ben-Akiva, 2007, section 7). Since a deterministic logistics module is complex, and estimation of disaggregate models take a significant amount of time, a 'preliminary' or 'prototype' version of the logistics model was developed (see de Jong and Ben-Akiva, 2007, section 8) in 2005-2006. This version did not require disaggregate estimation. Instead, the deterministic model relied on a cost minimisation procedure per firm-to-firm (f2f): only the alternative for transport mode and shipment size with lowest total logistics cost was chosen for each f2f flow.[3] During the following years, the prototype of the Norwegian Logistics Model has been improved through numerous iterations, including a calibration that aggregate data to a base year. However, it has remained a deterministic model, relying on a cost minimisation procedure[4].

---

[2] Aggregate-disaggregate-aggregate model system implies that input data and the network model are given at an aggregate level, while the logistics model operates at a disaggregated level. To secure transferability between the applications, the aggregated input data is disaggregated into flows between individual producers and consumers at firm level. The disaggregated flows are then used to predict logistics decisions in the logistics model deterministically, and the flows are aggregated for network assignment. For detailed information regarding the Norwegian logistics model and Logistics Model we refer to Significance (2008) and Significance and SITMA (2013)

[3] The same development took place in Sweden.

[4] A deterministic setup is also used in other logistic models, like the logistics models within the ADA framework for Sweden and Flanders (Ben-Akiva and de Jong, 2013, section 4.6). The Danish National Logistics Model, however, that is currently being developed and follows the ADA setup, contains a module for the choice of mode to cross the Fehmarn Belt screen line. This module uses a random utility model estimated on disaggregate data (including stated preference SP surveys in the Fehmarn Belt corridor). Other transport chains, for example in Denmark, are handled by a deterministic logistics model (Ben-Akiva and de Jong, 2013, section 4.6).

Conceptually, this method is almost equivalent to the all-or-nothing assignment method often used for allocating traffic flows to a network (de Jong and Ben-Akiva, 2007, section 8); the difference is that in the logistics model, changing the shipment size might lead to different cost minimizing transport chains for given PC-relations and commodity groups as well. The outcome of the model is an optimal shipment size for every given commodity flow, between each sender and receiver. In addition to different shipment sizes, the Norwegian Logistics Model includes different available transport solutions for different f2f-flows, which yields diversity between the choices. Still, it follows the cost minimisation setup, where the choice of transport is given by minimisation of total transport cost.

Today the ADA-version of the Norwegian National Logistics Model system consists of freight flow matrices (representing flows between producers (P) and consumers (C)), a set of cost functions, and a logistics module[5]. The application allocates the producer-consumer (PC) specific freight flows from the freight flow matrices between transport modes, along routes and terminals. The allocation is based on Level of Service-data (LoS), calculated from an international freight transport network. The calculation of LoS-data is done in Cube Voyager (Madslien et al, 2012), while the cost functions are developed by Grønland (2012; 2015).

2.2 From a deterministic to a random utility model

The models we present in this paper deal with the choice of shipment size and transport chains (which is a series of modes, possibly only one, used in the transport from production location P to consumption location C). As pointed out by Johnson and de Jong (2011), transport mode and shipment size are closely linked decisions. Large shipment sizes usually coincide with higher market shares for non-road transport, whereas there is a high correlation between road transport and small shipment sizes.[6] If one is unable to capture this dependence through explanatory variables, there will be correlation in the structure of the error terms. Not taking this correlation into account will lead to endogeneity (simultaneity) bias.

There are several arguments for going from the current deterministic model to a random utility logistics model in Norway:

1) A deterministic model has a weak empirical foundation: the way transport agents behave in the model is not based on observed data, but on the assumption that cargo owners will choose the shipment size and transport chain that minimize costs under certain conditions (and on data relating to transport networks, possible transhipment locations, and expert knowledge of cost functions). Instead of observed behaviour, such models represent normative behaviour: what would be the outcome if all freight transport agents behaved entirely according to standard economic theory (see also Tavasszy and de Jong, 2014, chapters 6 and 10). This

---

[5] The logistics module is an independent computable application (Madslien et al, 2012), developed by Significance (2008).

[6] In this respect, Norway may differ from other countries. The main customers of the *railway* are the big logistics service providers (Schenker, Bring and PostNord) which consolidate shipments from many firms into containers in consolidation centres close to the rail terminals. The main rail services in Norway (in tonne kms) consist of container transport between the main Norwegian cities (Oslo, Stavanger, Trondheim, Bergen, Bodø and Narvik, with Oslo as the main hub).

assumption is not completely without merit, however, in particular for professional buyers that dominate a large part of the market (in particular the bulk market). [7]

2) While the previous assumption may be justified, it is problematic that we lack access to full information regarding all cost elements that transport agents take into account when making their choices. Instead of utilizing actual (observed) transport prices, the logistics model assumes that transport costs can be calculated from bottom-up cost coefficients and network data. Therefore, we need a way to take into account that we lack information regarding the actual cost. Doing this is impossible within a deterministic framework, and calls for econometric methods. So far, calibration of the logistics model has been done at the level of aggregate data (data on ports and rail terminals, road traffic counts on the main road network). Despite this, empirical revealed preference (RP) data are available for Norway, also at the individual shipment level. This can provide a basis for an econometric analysis.
3) Bullet point number two implies that explanatory factors not part of the calculated logistics costs are excluded from the logistics model. This is an issue as the choice of transport chain also depends on factors such as reliability and flexibility of modes.[8] While these factors may affect the perceived costs of the transport agents, it is difficult to determine to what extent this is the case. To include e.g. reliability as an element of the logistics cost, it is necessary to have a meaningful translation from some metric of measured reliability to monetary units. Here the available RP data only provides limited information. At least, the RP data contains observed choice information that is the result of all relevant factors together, which makes it possible to estimate constants per transport chain alternative. To estimate separate coefficients for factors as reliability and flexibility of modes, one would must collect additional data, presumably stated preference data, and estimate a joint model with SP and RP data[9] (it might be possible to make use of the Norwegian SP study on the value of reliability in freight transport (Halse et al., 2010). However, the need to include these other factors applies further improvements of the Norwegian logistics model. For now, every commodity type for which we can base the choice mechanism on observed RP data, constitutes an improvement relative to the deterministic model. Moreover, moving from a deterministic framework to a random utility logistics model lays the foundation for being able to incorporate data from SP surveys in a later phase. Another benefit of SP surveys is that they can lead to better-founded values of reliability for use in cost-benefit analyses.
4) A well-known disadvantage of deterministic models (which can also be a feature of all-or-nothing traffic assignments) is that the impact of changes in scenario variables (e.g. oil prices) or policy variables (e.g. a new road, railway or terminal) can lead to implausibly large responses, or so-called 'overshooting' or 'flip-flop' behaviour. This occurs when the relevant part of the logistics costs function is rather flat, so that a small change in logistics costs can result in a shift to a

---

[7] For forecasting purposes a second assumption is also made indirectly; namely that long term prices move in parallel with long term costs. The last assumption has some merit as well (e.g. through minor empirical comparisons).

[8] On the other hand, the importance of non-cost factors should not be overstated. In negotiations on transport contracts, the major factor for choosing a partner in Norway is the cost for the buyer, given that the supplier delivers with a minimum quality. Due to the competition in the market, at least in Norway, the quality offered also seem to be very much on the same level for the various major transport suppliers.

[9] It is not necessary for a random utility model to include SP evidence; for this purpose, RP shipment data is sufficient.

completely different optimal shipment size and transport chain. This phenomenon does not always arise, and can to some degree be controlled for by using many different f2f flows in the model, which do not have to move in the same direction. Also, if the optimal alternative has much lower logistics costs than the second-best alternative, the model behaviour could be very stable. But the possibility of flip-flop behaviour is present, and could lead to cases that behave too shakily to properly compare reference cases and project cases in a cost-benefit analysis.

Issues 1, 2 and 4 above can be resolved by estimating disaggregate random utility models with available RP data: the observed shipment level data will then form the empirical basis for the behavioural coefficients of the model. By their nature, these are probabilistic models, because they include a stochastic component to account for the influence of omitted factors. A deterministic model effectively assumes that the stochastic component can be ignored – in other words, that the researcher has full knowledge of all the drivers of behaviour, and that there is no randomness in actual behaviour. As a result of adding the stochastic component from the random utility modelling, the response functions (now expressed in the form of probabilities) become smooth, instead of lumped at 0 and 1, as in a deterministic model. Regarding the third issue, the available RP data do not contain information that includes the softer factors that may also influence the choice of shipment size and transport chain. However, the RP data contains information on choices really made, hence one can estimate constants per transport chain alternative and provide the average influence of these factors based on observed choices (besides the influence of the stochastic component).

2.3 Options for the econometric specification

In the following, we present the notation used in this report and then discuss two potential models. The first model is the chosen estimation approach, and will therefore be described more in detail. The second model is meant as an alternative approach, and is further discussed in section 7.2.

The I available modes are indexed by i; the Q *discrete categories* of shipment sizes are indexed by q; while variable SS_i refers to a *continuous* choice of shipment size given mode i. We formulate the following two disaggregate models, where we drop commodity specific subscripts and observational subscripts for notational simplicity, specified as:

1. A joint model with discrete mode and discrete shipment size choice:

$$U_{iq} = \beta_1 X_{iq} + \phi_1 G_{iq} + \varepsilon_{iq} \tag{1}$$

2. A joint model with discrete mode and continuous shipment size choice:

$$U_i = \beta_2 X_i + \phi_2 G_i + \varepsilon_i \tag{2.1}$$

$$SS_i = \beta_3 G_i + \epsilon_i \tag{2.2}$$

where:

- U_{iq} is the utility derived from a discrete combination of mode i and shipment size category q;
- U_i is the utility derived from mode i;

- X_{iq} is a column vector of independent mode- and shipment size category- specific variables assumed to affect mode choice;
- X_i is a column vector of independent mode-specific variables assumed to affect mode choice;
- G_{iq} and G_i are column vectors of shipment category and mode specific independent variables, and mode specific independent variables respectively, assumed to affect shipment size;
- SS_i is the continuous shipment size for mode i;
- $\beta_1, \beta_2, \beta_3, \phi_1, \phi_2$ are row vectors of parameters to be estimated; and
- $\varepsilon_{iq}, \varepsilon_i$ and ϵ_i are random error terms.

Model 1 (eq. 1) is a formulation that accounts for shipment sizes. Here U_{iq} is the utility derived from the choice of a discrete combination of mode i and shipment size choice q. The dimension of this model, i.e. the number of discrete choice alternatives faced by the transport agents, will be IQ. As well as explanatory variables that affect mode choice (X_{iq}), it controls for an additional set of explanatory variables that affect shipment size choice decisions (G_{iq}). Note that grouping explanatory variables into the two categories X and G is purely conceptual and not meant as being restrictive; explanatory variables can affect both mode choice and shipment size at the same time.

Indexing the vectors of explanatory variables by both i and q is the most general formulation, as it allows the set of explanatory variables to vary between modes and/or shipment categories. However, most explanatory variables (e.g. cost) are usually included in all IQ alternative specific utility functions. Coefficients can be restricted to be equal for all or a subset of utility functions, or they can be allowed to vary depending on mode of transport and/or shipment size category.

Following the model 1 (eq. 1) specification has the advantage that it does not require a combination of techniques from discrete choice and regression analysis, but can be estimated fully within the discrete choice framework. It can be estimated as a multinomial logit, but also as a nested logit or mixed logit, depending on the assumptions made about the distribution of the error terms ε_{iq} (see next section). This is the chosen estimation technique in most of the existing studies on the field.

If the choice of shipment size in reality is a continuous one, the approach used in model 1 (i.e. grouping shipment sizes in discrete categories) may lead to measurement errors. Model 2 (eq. 2.1 and eq. 2.2) is a formulation that overcomes this problem by taking into account that shipment size is a continuous choice. This is a more complicated approach than estimating model 1, since it involves estimation of two dependent equations of a different nature (discrete and continuous). Since this is a pilot estimation project, we chose to start with the estimation of model 1. For further work, we suggest looking at the possibility of estimating a model of the same structure as model 2. See section 6.2 for more information regarding a discrete-continuous approach.

2.4 Potential correlation of error terms

Decisions on the optimal shipment size and mode are generated from the same optimisation problem, which implies that the error terms are likely to be correlated. In model 1, it is natural to believe that correlation exists between $\varepsilon_{1q}, \varepsilon_{2q}, \ldots, \varepsilon_{Iq}, \forall q$ as well as between $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iQ}, \forall i$. In other words, it is likely that error terms from utility functions within the same discrete categories of either mode or shipment size are correlated. This will be the case if unobserved factors vary between observations and increase the probability of either choosing a certain transport mode or a certain shipment size. In model 2, there will most likely be correlation between ε_i and $\epsilon_i, \forall i$. In other words, unobserved factors may influence the shipment size as well as the mode choice.

Ignoring this correlation will lead to a specification bias. For model 1, one might overcome this problem by estimating a cross-nested logit model and grouping alternatives with the same mode, as well as alternatives with the same shipment size category in the same nests. Alternatively, one could estimate a nested logit model, nesting either alternatives with the same modes or the same shipment sizes together. For model 2, the problem can be solved by estimating both equation 2.1 and 2.2 simultaneously, and explicitly considering the correlation between the error terms (see, e.g. Holguin-Veras, 2002. See also the discussion in section 7.2).

It is important to emphasize that correlation is only an issue when it is unobserved; if all relevant correlation patterns are captured by explanatory variables, the correlation between the error terms will be zero. This can be illustrated by a simple example: if shippers indeed are solely acting as cost minimizers, and all relevant cost elements are captured in an explanatory variable for "cost", that varies between transport modes and shipment sizes, it is safe to assume that covariances between error terms are zero, as there are no remaining variables that explain the shipper's decision. However, say that there is an actual cost related to uncertain transport time that is not captured in the explanatory (observed) variable "cost". This can for instance be potential stockout costs. Since we do not have access to any explanatory variable capturing (unwanted) variation in transport time, the effect of this uncertainty on the alternative specific utility functions will be captured by the error term. Firms with a high cost of uncertainty might want to choose the mode with the most predictable transport time. However, if they are risk averse, they might also want to increase the number of transport and reduce the size of each shipment, in order to minimize the negative effect of each potential delay. Hence, there will be a negative correlation between the shipment size and modes of transport with predictable transport times, that is not captured by explanatory variables. This may potentially bias all estimates, reducing the validity of the estimation results.

2.5 Proposed procedure for implementation

The previous sections only dealt with the *estimation* of a logistics model. To establish a version of the Norwegian Logistics Model that is based on random utility modelling, the following further steps are required:

a. **Extend coverage to all commodities**

The estimated models described in this report do not cover all commodities that should be covered in the logistics model (see Significance and SITMA, 2013). Therefore, one must determine which behavioural rules should be used for any non-estimated commodities. The alternatives are further estimation, relating the remaining commodities to a similar commodity for which a model was estimated, or keeping the NGM deterministic for these commodities. For the demo version that is developed in this project, we chose to replace the deterministic rules by the estimated models for the three commodities for which we now have such estimates, and to keep using deterministic behaviour for the other commodities.

b. Determine the annual firm-to-firm flows

For the probabilistic models, the routines to generate firm-to-firm flows (from the zone-to-zone flows) can remain as they are. The new models will be applied at the level of these firm-to-firm (f2f) flows, so as input for the logistics choices, we will know the annual firm-to-firm demand Q, as well as the transport time and cost for all available transport chain, and shipment size alternatives.

c. Determine the input for applying the utility functions

The utility functions that are estimated in this paper are similar to the current total logistics costs formulations. An important difference is that some of the cost components and some of the parameters (e.g. order costs, implied discount rate on the inventories in transit and in the warehouse, or a value of time; see de Jong and Ben-Akiva, 2007, section 4) have been estimated (as described above) instead of being assigned an assumed value by logistics cost experts. However, for the application of the estimated utility functions, we will still need transport distance and time for the available shipments size and transport chain combinations. In the current logistics model, this is done by the BuildChain routine. The new BuildChain program can remain similar to the current version, but needs to be adapted to reflect the available chain types from the RP data. In principle, this work has already been carried out in order to provide input data for estimation of the models described in earlier sections of this paper. However, this refers to the base year. If one wants to apply the model for later years (including future years), one have to make new networks and new assumptions on the transhipment locations, new network assignment to determine the optimal routes per transport chain, and new assumptions regarding the magnitude of various components of the logistics cost functions.

d. Implementation of the utility functions and their coefficients to determine shipment size and transport chain choice probabilities

In the current logistics model, the routine for determining the choice of shipment size and transport chain, is called ChainChoi.exe. This is the part of the model that needs to be considerably re-programmed when moving to a random utility model. In a random utility setting, ChainChoi.exe needs to determine the expected *shipment size* and probabilities for each of the *transport chain* alternatives for each annual f2f flow in the model, and then sum potential shipments weighted by probabilities over f2f flows.

In equation form, what the logistics model does, is to determine the sequence (for each flow of commodity k from firm m to firm n):

$$\{q,l\}=\{q,(h_1,t_1),(h_2,t_2),\dots,(h_i,t_i),\dots,(h_{I_l})\}, \tag{3}$$

where q is the shipment size (the same over the whole transport chain, though it can be consolidated with other shipments), l is the transport chain, h is a mode used on a leg of the chain and t is the next transhipment location. The index $i=1,\dots,I_l$ denotes a leg of a transport chain (chain l has I_l legs).

Since we do not observe the transhipment locations t_i in our data from the Swedish CFS, we cannot include this choice in estimation. Therefore, we keep the split between the determination of the optimal transhipment points and the choice of transport chain separate, as in the current model. The determination of the optimal transhipment locations for each available chain type from the set of available locations will be done in BuildChain, as in the present model. The random utility model in the new ChainChoi.exe program will refer to the problem:

$$\{q,l\} = \{q,(h_1|t_1),(h_2|t_2),\dots,(h_i|t_i),\dots,(h_{I_l})\}, \tag{4}$$

The deterministic version of ChainChoi.exe solved this problem by finding a single (least cost) transport chain and shipment size alternative for each annual flow of commodity k from m to n. The probabilistic model then replaces this for specific commodities. We will now calculate a number of *probabilities* for the f2f flow, one *for every available alternative.*

Let the utility functions for each alternative be written as $U_{ql} = V_{ql} + \varepsilon_{ql}$. Here, V_{ql} is the deterministic part of the utility function, i.e. the part of the utility of alternative $\{q,l\}$ that can be explained by a function of the observed variables and their estimated coefficients. Hence, this formulation is equivalent to equation (1). Let ε_{ql} be an independent and identically distributed random error term from the Extreme Value Type I distribution. This ensures that probabilities can be calculated according to a multinomial logit specification. Then, for an alternative j (say an alternative with shipment size (class) q_0 and direct road transport as transport chain), the predicted probability can be calculated as:

$$P(\{q,l\} = j) = \frac{exp(V_j)}{\sum_{\{q,l\}} \exp(V_{ql})}, \forall\{q,l\} \in C, \tag{5}$$

Where C is the choice set for the particular f2f flow, i.e. all available combinations of shipment sizes and transport chains. The numerator is the exponentiated deterministic part of the utility function of alternative j, whereas the denominator is the sum of exponentiated deterministic utilities over all available alternatives.

These probabilities can be summed at the level of the origins and the destinations of the individual legs (e.g. from the sender m to the first transhipment location) over all f2f flows (and weighted by the volume of the annual flows), and by commodity type to get the origin-destination (OD) matrices. This procedure is rather similar to the sample enumeration procedure used in the Norwegian national passenger transport model, where application also involves summing probabilities. In calculating the transport costs, consolidation can be handled in the usual way, except that the volume using a certain mode between two transhipment points from a previous iteration will now also be based on a summation of probabilities. What becomes more difficult is to get a vehicle load factor at the level of the f2f flow (since we only have a probability per transport chain). However, it is still possible to calculate vehicle load factors per OD pair and commodity.

In terms of its dimensions, the stochastic NGM may need to be simplified, since it might be necessary to aggregate commodities for estimation and distinguish fewer transport chain alternatives in the estimated models than in the current model, mainly because of a lack of observations for each category in the available RP data. If one wishes to include more transport chains or vehicle/vessel types than can be managed in model estimation, this could be made possible by combining the estimated random utility models with deterministic models for the allocation to finer categories given the outcomes of the former models.

e. Testing and validating the implemented model

Finally, a test and validation of the resulting OD flows by mode against observed aggregate data for a (new) base year is recommended, since a model estimated on one data set (RP data in individual shipments) will not necessarily match with other data, such as traffic counts.

3 Data analysis

To identify a relationship between mode choice for freight transport and shipment size, we need consistent data over actual shipments, including information on the chosen transport chain and shipment size, in addition to other variables explaining logistic choice. A commodity flow survey is a possible data set for such an analysis, as it captures transport between geographical zones, and includes a set of freight specific variables. In Norway, such a survey was carried out in 2009. A drawback of this survey is that it does not include information about transport modes used. For that reason, we have chosen to use the 2009 Swedish Commodity Flow Survey (CFS) as our main data source. Data from the CFS is chained with cost-data, calculated by means of the Norwegian Logistics Model.

3.1 The Swedish Commodity Flow Survey (CFS)

Several of the previous works on discrete choice models for freight transport use Swedish CFS data. Both Windisch (2009) and Abate et al. (2014) estimate a discrete-discrete model for freight transport in Sweden, and base their analyses on the Swedish CFS data from 2004/2005. We perform similar estimations as both Abate et al. (2014) and Windisch (2009), but use the Swedish Commodity Flow Survey from 2009. However, we only used data on shipments between Norway and Sweden. A drawback of the 2009 survey is that freight transport might have been affected by the 2008 financial crisis, illustrated by different registered weights and values between different CFS. The CFS 2004/2005 registered shipments with a value of 2,093 billion SEK, with a weight totalling 282 million tons. In comparison, the value of the shipments in the CFS 2009 is 1,832 billion SEK, while the weight amounts to 190 million tons in total. We believe this difference mainly influence aggregate transport levels, not shipments at f2f level. Hence, the choice of transport mode(s) and shipment size are less influenced by the crisis.

From Trafikanalys we received data over shipments between Norway and Sweden, registered in the 2009 CFS. For each shipment, the data contains information regarding size, value, mode of transportation, commodity group, and geographical location of senders and receivers of goods, with municipality as the lowest geographical level. As the direction of the shipment is of little importance in this analysis, imports and exports are merged. This gives 105,533 registered shipments between Norway and Sweden, of which 102,882 shipments are from Sweden to Norway and 2,671 shipments are from Norway to Sweden. These numbers exclude shipments of timber, but include petroleum products[10]. The number, size and value of the registered shipments between Norway and Sweden are displayed in Table 1.

---

[10] Timber between Norway and Sweden is only shipped by vehicles. Even though it constitutes a large share of total shipments, it is not suited for estimating mode choice.

Table 1. Value, weight and number of shipments between Norway and Sweden. Data: Swedish CFS (2009).

	Million tons	Billion SEK	Nr. of shipments
To Norway	4.57	65.29	102,882
From Norway	7.66	35.04	2,671
In total	12.23	100.32	105,553

The table shows that the aggregated value of shipments between Norway and Sweden amounts to 100.32 billion SEK, with weight amounting to 12.23 million tons. While shipments from Norway to Sweden make up 63% of the weight, they only make up 35% of the value. This indicates that shipments from Norway to Sweden consist of voluminous goods, while shipments from Sweden to Norway have a higher value density on average.

The Swedish CFS contains information about chosen transport mode for each shipment, so that mode choice connected to both direct and intermodal transports are captured. The drawback is that only main transport modes are included. The CFS distinguish between road (V), rail (J), sea (S), air (L) and unknown (X), and hence lacks information about the type of vehicle/vessel used. For this reason, we can only analyse mode choice between main transport modes.

The shipments between Norway and Sweden in the CFS 2009 are divided into aggregated categories based on which mode in the chain is likely to be the main transport mode.[11] For the transport chains vehicle-ship/rail-vehicle, we assume that ship/rail is the main mode, and that vehicle was used as a feeder. In some cases, the mode of transport is missing. This is originally denoted *X* in the dataset. If mode X (missing) is registered together with another transport mode, besides vehicles, that other mode is chosen as the main mode. This gives us the distribution and frequency of number of shipments, weight and value for the four aggregated groups of transport modes, shown in Table 2. We also included a column for the average value density of the shipments, measured in SEK/kg.

Table 2. Distribution over main transport modes for shipments between Norway and Sweden. Data: Swedish CFS (2009).

Mode	Nr. of shipments	In %	Weight (mill tons)	In %	Value (bill SEK)	In %	SEK/kg
Air	2,179	2%	0.00	0%	1.17	1%	1,007
Rail	5,063	5%	0.29	2%	7.38	7%	26
Water	20,383	19%	9.14	75%	35.43	35%	4
Road	77,928	74%	2.80	23%	56.34	56%	20
Sum	105,553	100%	12.23	100%	100.32	100%	8

Table 2 shows that road transport has the highest market share measured both in number of shipments, and in SEK. When measured in weight, water transport has the highest market share. Air transport has the overall lowest weight share. The distribution of the main modes shown in Table 2 suggests that a negligible amount of freight is transported by air. However, when looking at the value density measured in SEK/kg we see that goods transported by air are on average far more valuable than goods transported by other modes. The mode with the least valuable shipments is water transport, where the average

---

[11] More information on the choice set in chapter 4.

ton transported is worth only 4 SEK. Rail and road are more similar in terms of value of the goods transported, with an average value density of respectively 26 and 20 SEK per kg.

In addition to transport mode, we are interested in the distribution of shipment size. This is shown in Table 3. The table also displays the distribution of shipment sizes over the main transport modes. We see that around 77 % of the shipments weigh less than 200 kg. Air transport mostly consists of small shipments, while the other transport modes have a higher spread. The categories are based on the categories from Abate et al. (2014).

Table 3. Size categories for transports between Norway and Sweden, as stated in the Swedish CFS (2009), for all commodity groups.

Size categories	Air	Rail	Water	Road	Total
0-25	1,175	639	15,494	30,683	47,991
26-50	352	459	2,894	14,185	17,890
51-200	337	1,200	881	13,282	15,700
201-800	172	870	439	6,668	8,149
801-3000	78	746	180	4,497	5,501
3,001-7,500	34	327	49	2,042	2,452
7,501-12,500	10	94	26	817	947
12,501-20,000	13	61	28	598	700
20,001-30,000	3	55	22	626	706
30,001-35,000	2	11	11	192	216
35,001-40,000	1	9	2	157	169
40,001-45,000	1	7	5	132	145
45,001-100,000	-	40	43	904	987
100,001-200,000	-	408	24	707	1,139
200,001-400,000	1	49	21	833	904
400,001-800,000	-	33	37	809	879
800,001-	-	55	227	796	1,078
Total	2,179	5,063	20,383	77,928	105,553

The analysis of the discrete choice of shipment size and transport mode is restricted to the three commodity groups 13 Iron and Steel, 17 Plastic and rubber and 30 Consumables. The reason for restricting the analysis to these commodity groups is the number of registered shipments, as well as the distribution over transport modes. This can be seen from Table 4, which shows the number of observations for the commodity groups shipped between Norway and Sweden, as well as the frequency distribution on the transport modes rail, road, water, and a combination of these.

Table 4. Number of observations and frequency distribution over transport modes for commodities shipped between Norway and Sweden, and registered in the 2009 Swedish CFS. Ra=rail, Ro=road, Wa=water, R-R=road-rail, R-W=road-rail, R-W-R=road-water-rail.

Commodity group		Observ.	Ra	Ro	Wa	R-R	R-W	R-W-R
1	Agricultural products	430	0%	100%	0%	0%	0%	0%
3	Gen cargo, living animals	470	0%	100%	0%	0%	0%	0%
5	Fresh fish and seafood	129	0%	89%	0%	0%	11%	0%
7	Thermo, consumption	229	0%	98%	0%	2%	0%	0%
8	Food, consumption	464	0%	98%	0%	0%	1%	0%
10	Animal food	378	1%	99%	0%	0%	0%	0%
11	Organic inputs/materials	511	0%	97%	0%	3%	0%	0%
12	Other inputs/materials	375	27%	53%	0%	0%	20%	0%
13	Iron and steel	845	51%	43%	0%	0%	5%	0%
14	Other metals	454	4%	91%	0%	2%	4%	0%
15	Metal products	3,765	0%	97%	0%	0%	3%	0%
16	Chemical goods	1,000	2%	89%	0%	1%	8%	0%
17	Plastic and rubber	3,339	0%	62%	0%	38%	0%	0%
18	Timber and products from forestry	8	0%	100%	0%	0%	0%	0%
19	Products from lumber and wood	1,095	0%	96%	0%	0%	3%	0%
20	Wood and pulp	76	0%	100%	0%	0%	0%	0%
21	Paper	925	7%	91%	0%	1%	1%	0%
22	Printed matters, software and film	154	0%	99%	0%	1%	0%	0%
23	Coal, peat and ore	42	0%	100%	0%	0%	0%	0%
24	Mass commodity	346	0%	100%	0%	0%	0%	0%
26	Machines and equipment	867	0%	100%	0%	0%	0%	0%
27	Electric articles	11,193	2%	98%	0%	0%	0%	0%
28	Gen cargo, building materials	1,918	0%	97%	0%	3%	0%	0%
29	Cement and concrete	44	0%	98%	0%	2%	0%	0%
30	Consumer goods	69,255	0%	71%	0%	4%	24%	2%
32	Vehicles/transport equipment	1,475	1%	96%	0%	1%	1%	1%
33	Unrefined petroleum	8	0%	0%	100%	0%	0%	0%
35	Refined petroleum products	1,371	0%	84%	12%	0%	4%	0%
36	Bitumen	3	0%	100%	0%	0%	0%	0%
37	Waste and recyclables	73	0%	100%	0%	0%	0%	0%
38	Processed fish	32	0%	100%	0%	0%	0%	0%
Sum		101,274	1%	77%	0%	4%	17%	0%

3.2 Cost data

The data in the CFS lacks information about transport costs. Since transport cost is expected to be a vital part of the transport mode choice, information about commodity, shipment size, origin and destination from the CFS is chained with cost-data calculated from the Norwegian Logistics Model. Since we want to study the choice of both shipment size and mode chain, we need transport costs not only for the specific shipments size and mode chosen, but also for alternative shipment sizes and mode chains.

For each shipment and origin and destination relation, transport costs for different shipment sizes and transport chains are calculated, using information from the model's cost functions, combined with information from the network model about distance for road and road/rail/ship combinations. The program calculating costs utilizes all available transport chains in the "Build chain program", which is a part of the Norwegian Logistics Model.

The transport costs for different transport chains are utilised in the estimation model for the choice between shipment size and mode choice, from a range of possible available mode chains. In order to estimate cost data, a vehicle is added at the start/end of a chain.

4 Model approach

In this chapter, we present the estimation approach, as well as the endogenous (dependent) and exogenous (explanatory) variables in the estimated model. We have chosen to estimate a discrete-discrete model, treating both the choice of transport chain and shipment size as discrete endogenous variables. The advantage of this methodology is that it does not require a combination of techniques from discrete choice and regression analysis, but can be done fully within the discrete choice framework. One drawback is that assuming independence between the error terms in the utility functions for mode and shipment size choice, and discretizing the continuous information on shipment size, can be interpreted as forms of specification errors (Johnson and de Jong, 2011). However, as this analysis is a preliminary project, where estimation is limited to three commodity groups, we think that the discrete-discrete model is sufficient.

Estimations are carried out for commodity groups 13 Iron and steel, 17 Plastic and rubber, and 30 Consumable goods. The goal is to determine robust coefficients for a selection of explanatory variables in the Norwegian Logistics Model, and to introduce the choice of transport chain and shipment size as the outcome of a random utility model, not solely from deterministic cost minimization. The methods used for estimating discrete choice models for the three commodities is similar to the analysis by Windisch (2009) and Abate et al. (2014).

4.1 Choice set

The discrete potential outcomes of the dependent variable, also referred to as the choice set, are the different combinations of transport chain and shipment size available for each shipment. The universal choice set is made up of all existing shipment alternatives in the network, and is therefore equal for all commodity groups. Presenting the universal choice set is tedious, and only a small part of the universal choice set is actually available or chosen. Instead of presenting the universal choice set, we present the available choice set for each commodity group, given by the observed combination of transport chain choice and shipment size choice in the CFS. Our classification of the choice set is inspired by the choice sets defined by Windisch (2009), Johnson and de Jong (2011) and Abate et al. (2014), who all estimate discrete-discrete choice models, using the Swedish commodity flow survey from 2004/2005. We present their grouping of shipment size and transport chain in chapter 4.1.1.

4.1.1 Choice sets in existing literature

The endogenous variables in Windisch (2009) are shipment size category (see Table 5) and transport chain (defined as a series of modes from P to C; see Table 6). The transport chains are given in the CFS, where the respondents were asked to list all modes used for each shipment in the right sequence. In many cases, Windisch added lorry transport to the start and/or end of a chain, to allow for a calculation of transport costs from the networks. For the transport chain lorry-lorry-lorry, the middle lorry is a larger vehicle type and the others are smaller. The choice alternatives are combinations of the two variables shipment size category and transport chain type. Note that not all combinations are available.

Table 5. Shipment size categories (weight in kg) in Windisch (2009).

Category	From (kg)	To (kg)
1	-	50
2	50	100
3	100	200
4	200	500
5	500	1,000
6	1,000	2,000
7	2,000	5,000
8	5,000	10,000
9	10,000	20,000
10	20,000	50,000
11	50,000	100,000
12	100,000	200,000
13	200,000	400,000
14	400,000	600,000
15	600,000	800,000
16	800,000	1,000,000
17	1,000,000	1,500,000
18	1,500,000	or more

Table 6. Transport chain alternatives in Windisch (2009).

#	Transportation Chain			
	Leg 1	**Leg 2**	**Leg 3**	**Leg 4**
1	lorry			
2	lorry	vessel	lorry	
3	lorry	rail	vessel	lorry
4	lorry	rail	vessel	
5	lorry	air	lorry	
6	lorry	lorry	lorry	
7	lorry	ferry	lorry	
8	lorry	vessel		

Johnson and de Jong (2011) also predict the choice of shipment size category and mode in a discrete-discrete model. They estimate their model using more aggregated groups of shipment size and chain choice categories than Windisch. Their shipment size categories are displayed in Table 7.

Table 7. Shipment size alternatives in Johnson and de Jong (2011).

Category	From (kg)	To (kg)
1	-	3,500
2	3,500	15,000
3	15,000	30,000
4	30,000	100,000
5	100,000	or more

When it comes to mode choice, Johnson and de Jong operate with only the main modes of transport. The shipments' main mode results from an assumed hierarchy of modes where air transport is ranged highest, followed by water, rail and road transport. This means that whenever air transport is present in a transport chain air transport is the chosen mode alternative for this shipment irrespective of whether other modes are used in the transport chain or not. Similarly, will water be the preferred mode if the transport chain contains water transport, but not air transport, etc. Road will only be the allocated transport modes if no other mode are present in the transport chain (Johnson and de Jong, 2011). Johnson and de Jong distinguish between four main transport modes in their discrete choice model:

Table 8. Mode alternatives in Johnson and de Jong (2011).

#	Modes
1	Road transport
2	Rail transport
3	Water transport
4	Air transport

Finally, the discrete-discrete model in Abate et al. (2014) distinguishes between 16 weight categories and five transport chains. Like Windisch (2009), Abate et al. add lorry transport to the start and/or end of a chain, to allow the calculation of transport costs from the networks.

Table 9. Shipment size categories (weight in kg) in Abate et al. (2014).

Category	From (kg)	To (kg)	Freq. in CFS	%
1	-	50	703.9	24
2	50	200	153.2	5
3	200	800	160.4	6
4	800	3,000	157.9	5
5	3,000	7,500	136.9	5
6	7,500	12,500	127.6	4
7	12,500	20,000	161.7	6
8	20,000	30,000	210.9	7
9	30,000	35,000	207.6	7
10	35,000	40,000	344.7	12
11	40,000	45,000	340.5	12
12	45,000	100,000	153.9	5
13	100,000	200,000	10.8	0
14	200,000	400,000	7.2	0
15	400,000	800,000	6.4	0
16	800,001	-	5.6	0
Total			2,889.2	100

Table 10. Chain choice alternatives in Abate et al (2014).

#	Transport chains		
1	Lorry		
2	Rail		
3	Lorry	Rail	Lorry
4	Lorry	Ferry	Lorry
5	Lorry	Vessel	Lorry

4.1.2 Choice set based on the Swedish CFS 2009

Our grouping of shipment sizes and transport modes used in the estimation of discrete choice models is a combination of the grouping used by Windisch, Abate et al., and Johnson and de Jong. Since we are estimating our models based on shipments between Norway and Sweden, registered in the Swedish CFS, we have a limited number of observations, and therefore a limited distribution of the dependent variables. This has an impact on the choice set.

The choice set used in the estimation is determined by all the chosen combinations of transport chains and shipment size categories in the CFS, under the restriction that cost data can be estimated. To estimate cost data, we need information about sending and receiving zones in both Norway and Sweden. Observations lacking information about one or both variables are excluded from the dataset, and choice categories that consequently lack observations will thus be excluded from the choice set.

The available choice set for transport chains is also restricted by the format of the CFS. For each observation, the CFS includes information regarding the transport modes that were used. As in the Swedish CFS from 2004/2005, we started out assuming that shippers reported the transport modes in sequence, and thus provided information about the transport chain. After careful investigation of the reported transport modes, it seemed that the reporting of transport modes is random. For example, we find observations where the transport modes reported are road-rail-water-air. This seems unlikely to be a chain between Norway and Sweden. Some shippers also report "unknown transport mode", denoted by X, in a sequence. Because of this uncertainty, we assume that the modes reported in the CFS are a list of modes used, not the transport chain for the shipment. We present the resulting transport mode choice in the CFS for each of the three chosen commodity groups in Table 11.

Like Johansen and de Jong (2011), we assume a hierarchy of the reported modes. Based on the combinations of modes from Table 11, we assumed that where rail is reported, rail is the main mode for the observations. The same goes for sea. Road is the main mode only when it is reported as the sole transport mode. The reason for these assumptions (as can be seen from Table 11) is that we have a) too few observations to be able to estimate a model based on all the observed transport modes in the CFS, and b) road is the dominating transport mode, so we make some adjustment to increase the frequency of the other transport modes (rail and sea). Air would be a preferred mode, but we have few observations of shipments using air transport and hence chose to exclude air transport from estimation.

Table 11. Observed transport mode choice in the CFS, for commodity group 13, 17 and 30. Per cent of number of trips. Data: Swedish CFS.

Categories	Iron and steel	Plastic and rubber	Consumable goods	Sum
Air	0%	0%	0%	0%
Rail	55%	0%	0%	1%
Road	44%	61%	67%	66%
Road - air	0%	1%	2%	2%
Road - rail	0%	38%	3%	5%
Road - water	0%	0%	22%	21%
Road - water - air	0%	0%	3%	3%
Road - water - rail	0%	0%	2%	2%
Water	0%	0%	0%	0%
Sum	100%	100%	100%	100%

From Table 11, we see that the CFS in total contains nine valid combinations of observed mode choice. In total, road (in isolation) is the most preferred mode, followed by a combination of road and water. Shipments that use transport on water, air or rail in isolation, are less common. For all nine combinations of the mode choice given in Table 11, road transport might be used as a feeder, either at the beginning or the end of the transport, or both. When rail, water or air are the first or last mode in a chain, producers and receivers must be located at a rail terminal, quay or an airport. There is no information in the Swedish CFS 2009 regarding whether one of the operators is located close to a terminal or a quay, or if feeder transport is required at one or both ends of the shipment. To check this, the logistics model was used to investigate whether there was a registered terminal or quay at the sending/receiver zone. If not, truck was added as a feeder in the transport chain, and in the resulting cost data.

In addition to transport modes, the choice set depends on the chosen shipment size. The spread of the shipment size varies highly within the three commodity groups. Based on the shipment size groups in chapter 3.1, we obtain the following distribution, presented in Table 12.

Table 12. Frequency of the observed shipment size in the CFS, in total and for each of the commodities 13, 17 and 30. Data: Swedish CFS.

Categories	Iron and steel	Plastic and rubber	Consumables	Sum
0-25	1%	**17%**	**57%**	54%
26-50	0%	**12%**	**20%**	20%
51-200	1%	**29%**	**14%**	14%
201-800	1%	**15%**	4%	5%
801-3,000	1%	**12%**	3%	3%
3,001-7,500	1%	5%	1%	1%
7,501-12,500	1%	2%	0%	0%
12,501-20,000	2%	2%	0%	0%
20,001-30,000	**8%**	1%	0%	0%
30,001-35,000	2%	0%	0%	0%
35,001-40,000	3%	0%	0%	0%
40,001-45,000	1%	0%	0%	0%
45,001-100,000	**19%**	2%	0%	0%
100,001-200,000	**53%**	1%	0%	1%
200,001-400,000	3%	1%	0%	0%
400,001-800,000	1%	0%	0%	0%
800,001-	**4%**	1%	0%	0%
Sum	100%	100%	100%	100%

In Table 12, we have highlighted the largest shares for each shipment size category and commodity group. For commodity group 13, "Iron and steel", most of the shipments are within the range of 45,000-200,000 kg. For commodity group 17, "Plastic and rubber", and 30, "Consumables", most shipments lie within the range of 0-3,000 kg and 0-200 kg respectively. Observations outside of the mentioned clusters are dispersed between the other shipment size groups. The variation is particularly large for commodity group 13 and 17.

Due to the high spread in shipment size, both within and between the commodity groups, we have chosen to aggregate shipment size groups even further in our estimation. We have chosen to use four shipment size groups for the three commodities. With four shipment size groups, we capture some diversity in shipment size and manage to keep the number of observations in each choice set above 30, which were set as a requirement to be included as an alternative. Due to different frequency distributions for the observed alternatives, choice sets will differ between the commodity groups. Below, we present the choice sets for the three commodities used in estimation, which is 13 Iron and steel, 17 Plastic and rubber and 30 Consumables.

4.1.3 Commodity 13: Iron and steel

The Swedish CFS registered 791 valid observations[12] between Norway and Sweden for commodity group 13 Iron and steel. Most of these shipments are transported by rail or road, and have a weight above 10 tons. The chosen transport chain and shipment size categories are shown in Table 13, together with the frequency distribution of the choices.

Table 13. Frequency distribution of transport chain choice and shipment size for commodity group 13 Iron and steel. Data: Swedish CFS 2009.

Categories	Air	Rail	Road	Road - air	Road - water	Water	Sum
0-10,000 kg	0.3%	0.0%	4.7%	0.3%	0.1%	0.0%	5.4%
10,001-50,000 kg	0.0%	3.9%	13.0%	0.1%	0.0%	0.0%	17.0%
50,001-150,000 kg	0.0%	20.7%	18.2%	0.0%	0.1%	0.0%	39.0%
150,001 kg -	0.0%	30.5%	7.8%	0.0%	0.0%	0.3%	38.6%
Sum	0.3%	55.1%	43.7%	0.4%	0.2%	0.3%	100.0%

From Table 13, we see that 44 % of the shipments are transported by road, and 55 % by road-rail-road (where we can assume that rail is the main mode, and transport by road consists of feeder transport to and from rail terminals). Only 4 shipments are transported on sea, which is approximately 0 % of total shipments of iron and steel between Norway and Sweden in 2009. For this reason, transport by sea is excluded as a mode alternative. The same reasoning applies to air transport.

Looking back at Table 12, we see that the weight groups defined are too wide for iron and steel. Few of the weight categories will produce sufficient observations to be estimated in a discrete choice model. We therefore reduced the number of weight categories, so that each category contains at least 30 observations per transport mode. The chosen shipment size categories are shown in Table 13 and Table 14. The combinations of shipment size categories and transport modes that have enough observations form the choice set for commodity 13. This is shown in Table 14. The smallest shipments (measured in kg) are only transported by road. For the other weight groups, we have observations for both road and road-rail-road, with rail as the main mode.

Table 14. The choice set for commodity group 13 Iron and steel.

Categories	Air	Rail	Road	Road - air	Road - water	Water
0-10,000 kg			X			
10,001-50,000 kg		X	X			
50,001-150,000 kg		X	X			
150,001 kg -		X	X			

---

[12] A valid observation implies that the observation includes all necessary information about weight, value, transport mode and sending and receiving zone. Based on this information we can calculate cost attributes necessary for estimation.

4.1.4 Commodity 17: Plastic and rubber

In total, we have 3,329 registered shipments of plastic and rubber between Norway and Sweden, of which air and water is registered as the main transport mode 21 times. Distributions over commodity chains are shown in Table 15.

Table 15. Frequency distribution of transport chain choice and shipment size for commodity group 17 Plastic and rubber. Data: Swedish CFS 2009.

Categories	Air	Road	Road - air	Road - rail	Road - water	Road - water - rail
0-25 kg	0%	8%	0%	10%	0%	0%
26-50 kg	0%	4%	0%	7%	0%	0%
51-200 kg	0%	11%	0%	19%	0%	0%
201 kg -	0%	39%	0%	3%	0%	0%

From the table, we see that road and rail are the most frequent mode choices also for rubber and plastic. More than half of the shipments, 61 %, is transported by road, while 38 % is transported by rail, with road as feeder transport. Less than 1 % is transported by air or waterborne transport. These chain choice alternatives contain too few observations to be included in the choice set. Combined with four shipment size categories, we get 8 alternative combinations of transport modes and shipments size choices. These alternatives make up the choice set for commodity group 17 Plastic and rubber between Norway and Sweden. The choice set is presented in Table 16.

Table 16. The choice set for commodity group 17 Plastic and rubber.

Categories	Air	Road	Road - air	Road - rail	Road - water	Road - water - rail
0-25		X		X		
26-50		X		X		
51-200		X		X		
201-		X		X		

Table 15 and Table 16 also indicate that commodity group 17, plastic and rubber, is transported in smaller quantities than iron and steel, and that transport chains with rail as main mode are preferred for the smallest shipments, while road is preferred for relatively large shipments of over 200 kg. This is a surprising correlation, as rail is usually preferred for heavy goods transported over longer distance. One explanation is that small shipments are consolidated before they are shipped by rail and over longer distances.

4.1.5 Commodity 30: Consumables

The Norwegian commodity group 30 "Consumables" did originally consist of three Swedish commodity groups; 100 "Textile products, clothing, fur and leather products", 109 "Furniture and it's like" and 111 "Other leftover consumable goods". Together, these three Swedish commodity groups make up almost 73,000 registered shipments. When estimating discrete choice models for commodity 30, we had trouble getting the model to converge. After some careful investigation of the registered shipments in the Swedish CFS, we found that the Swedish commodity group 100 "Textile products, clothing, fur and leather products" constituted nearly 60 % of the registered shipments in commodity group 30, all transported by road. We decided to exclude commodity 100 from estimation of commodity group 30.

After exclusion, commodity 30 consists of 29,200 registered shipments. For some of these registered shipments, we could not provide cost data for the chain choice alternative "Road-Water-Rail" because the chain choice was not available according to the Norwegian network data. These observations were also excluded from the data set for estimation. This means that we, for commodity group 30, lose another 399 observations. The remaining data set consists of 28,801 observations. The frequency distribution for the alternative chain choice is presented in Table 17.

Table 17. Frequency distribution of transport chain choice and shipment size for commodity group 30 Consumable goods. Data: Swedish CFS 2009.

Categories	Air	Road	Road-air	Road-rail	Road-water	Road-water-air	Road-water-rail
0-25 kg	0.0%	2.4%	2.2%	0.5%	47.5%	4.1%	2.0%
26-3,000 kg	0.0%	12.9%	1.5%	6.9%	9.0%	3.2%	2.4%
3,001-12,500 kg	0.0%	2.1%	0.0%	1.0%	0.0%	0.0%	0.0%
12,501 kg -	0.0%	1.9%	0.0%	0.5%	0.0%	0.0%	0.0%

In Table 17, we see that some of the alternatives have a low frequency. Chain choice alternatives with a frequency below 30 are excluded from estimation. Furthermore, we have few observations and lack data on costs for the transport mode air. These observations are therefore excluded as well. This leaves us with the following choice set, given in Table 18.

Table 18. The choice set for commodity group 30 Consumable goods.

Categories	Air	Road	Road-air	Road-rail	Road-water	Road-water-air	Road-water-rail
0-25 kg		X		X	X		X
26-3 000 kg		X		X	X		X
3 001-12 500 kg		X		X			
12 501 kg -		X		X			

The choice set for commodity 30 consists of 12 available chain choice alternatives for estimation, distributed over 25,631 registered shipments. This is the only commodity group for which we are able to include transport on water as an alternative (in combination with road/rail).

4.2 Explanatory variables

Compared to the 2004/2005 Swedish CFS the 2009 Swedish CFS lacks some information, like access to rail terminal or quay at sender, seasonal variation and, it seems, information about transport chain choice, not only transport modes used. Information like access to terminals at sender, and seasonal variation, have been used as explanatory variables in previous work with the Swedish CFS, by Windisch (2009), Johnson and de Jong (2011) and Abate et al. (2014). We have not been able to include these variables in our estimation, but tried to compensate by including regional specific dummies for rail transport and degradation cost in the estimation. An overview of the exogenous variables in the three most relevant models, along with the exogenous variables in our analysis, is given in Table 19.

Table 19. Exogenous variables in the most relevant literature over joint transport chain/mode models and exogenous variables used in our estimation.

Windisch (2009)	Johnson and de Jong (2011)	Abate et al. (2014)	Caspersen, Johansen, Hovi and de Jong (2016)
Transport chain constants	Modal constants	Transport chain constants	Transport chain constants
Shipment size constants		Shipment size constants	Shipment size constants
Cost (per shipment) – all modes	Cost (per shipment) – all modes	Cost (per shipment) – all modes	Transport cost (per shipment) – all modes
Value density dummies - for specific (small) shipment sizes	Value density dummy – for smaller shipment sizes	Value density dummy – for smaller shipment sizes	Value density dummy – for smaller shipment sizes
Dummy for summer – other modes than road only			
Dummies for bulk/pallet/boxes – for specific shipment sizes	Dummies by type of commodity – for specific modes		
Dummy for proximity of rail or quay at sender	Dummies for rail and sea access at sender	Dummies for rail and sea access at sender	
Nesting coefficients			
	Dummy for big company – for rail		
	Transport time times value of the goods – for air	Transport time times value of the goods – for road	Transport time times value of the goods (inventory costs)
			Region specific dummies for Norway
			Degradation costs

From Table 19 we see that there are some differences between the exogenous variables included in the four model estimations. Both Windisch, Johnson and de Jong and Abate et al. estimated their models using the 2004/2005 Swedish CFS, which makes their models comparable with respect to available (and included) variables and estimation results. Our model is estimated using the 2009 Swedish CFS. In the following we give a more detailed description of the explanatory variables included in the estimation.

Transport cost

Transport cost is not included in the 2009 Swedish CFS, and must be calculated from an external source and added to the data. This external source is the logistics model from the Norwegian freight transport model. To obtain the correct transport costs for the different observations in the CFS, we extracted information about each shipments' weight, origin and destination zone, as well as the direction of the shipment (import or export). This information was used to calculate the associated transport cost for each shipment from the logistics model. Costs were calculated for all available joint mode and shipments size choice for each of the three commodity groups. After calculating transport cost for all observations, we chained the cost-data from the logistic model with the original data from the CFS.

Deriving cost data from the logistics model makes it possible to distinguish between different cost components. In our analysis, we assume that all transport cost components have the same impact on the probability to choose a transport alternative. The transport cost that is used as explanatory variable in the estimation is the total transport cost, i.e. the sum of link-based costs, initial loading, final unloading and transhipment costs. In addition to transport cost, time use and degradation costs are included as explanatory variables.

Transport cost is a continuous variable, and changes with transport mode and shipment size. We expect transport cost to have a negative impact on the choice of different transport alternatives.

Time use

Time used for transportation is also expected to be of interest when looking at the choice between different transport alternatives. Variables for time use for each of the observations in the CFS are also calculated by means of the logistics model, and chained to the original data in the CFS.

Time use differs between transport modes. For the different shipment sizes, time use only differs if a change in shipment size induces a change in the vehicle (but still the same mode) used for transportation, which further leads to a change in transport time.

Time is treated as a continuous variable, and we expect an increase in time to have a negative impact on the choice of each alternative. We expect small shipment sizes and road transport to be most time sensitive.

Degradation and capital cost

Degradation costs are the costs associated with degradation of the value of the goods during transport, making it "less fresh" when reaching consumer. This is especially relevant for fresh fish, fruits and vegetables, but also for medication and electrical equipment that are time sensitive. Degradation cost for each observation is calculated by the logistics model, and chained to the original data in the CFS.

For the commodity types analysed here, degradation costs play a smaller role than for the commodity groups mentioned above. Degradation costs are also highly correlated to time use. We expect that goods with high degradation cost will have a high shipment frequency, reflected through transport in small shipments.

Degradation cost is treated as continuous. We expect that goods with high degradation costs will tend towards faster and more reliable transport chains, like road transport.

Value density

The CFS provides information about shipment size (in tons) and value (in 1000 SEK) for the registered shipments. From this information, we calculate value density in SEK/kg by dividing registered shipment weight by registered shipment value. This gives us a variable for the value density for all shipments. This variable is continuous.

A problem with continuous variables with a large interval of observed values, is that it is cumbersome to identify the true relationship between the variable and the choice probabilities. This is especially the case if a variable only affects the choice indirectly and to a small extent, as with "value density". By examining the data and testing various specifications, we concluded that the variance in this variable is too large, and that the relationships between the variable and the choice probabilities are too weak for us to be able to assign a meaningful functional form that can translate the continuous value density into alternative specific utility, by means of an estimated coefficient.

We have chosen to use dummy variables denoting whether the registered shipment value density is above or below the median value. Since outliers affect average value, we chose to use the median value to arrive at a more robust model. We expect value density to affect the size of the shipments negatively.

Inventory in transit

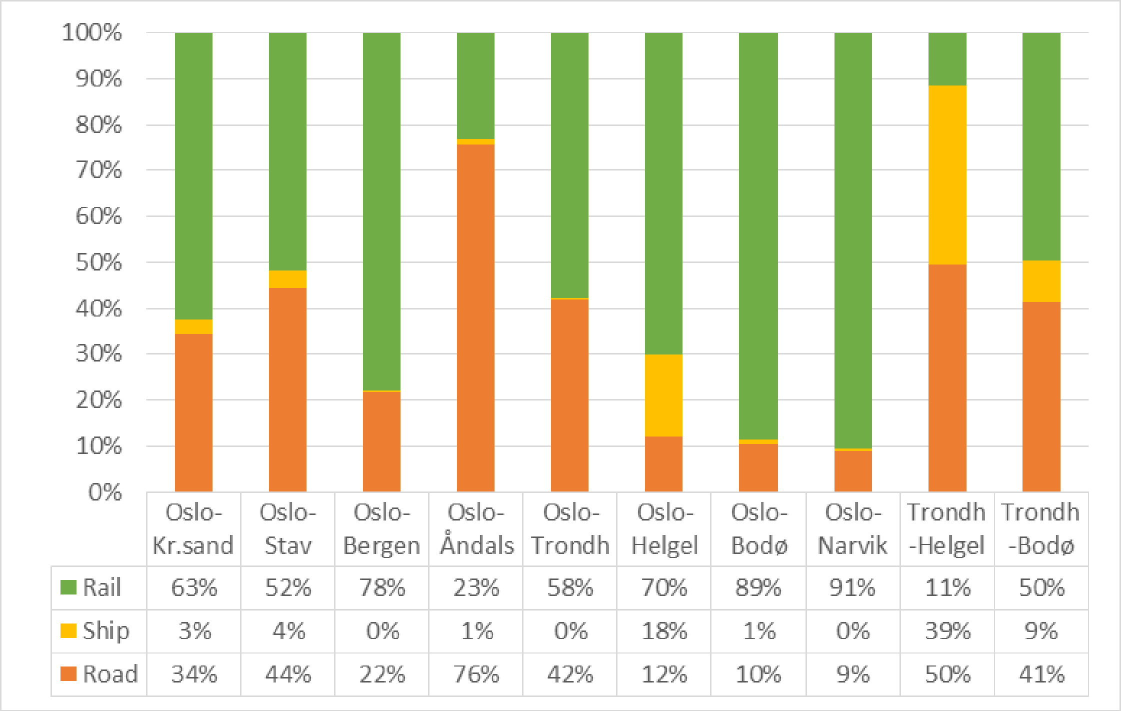
Inventory costs are defined by multiplying transport time with the value of the goods. Large inventories reduce the risk of not being able to serve demand or use the required inputs in a production process, but will increase the inventory cost. Small and frequent shipments, on the other hand, will result in lower inventory costs, but increase the transport and stockout costs (de Jong and Ben-Akiva, 2007). This trade-off between transport and keeping inventories is part of the logistics model.

The influence of the non-mobile inventory cost can be included more indirectly through value density variables. Inventory in transit captures the time cost of shipping, and is the sum of the value of the goods, adjusted for the different shipment sizes and transport time for the different chain choices. Inventory in transit will increase with the value of the goods and/or time used. If high value goods are transported with slow transport modes, inventory in transit will be high. At the same time, using a slow transport chain has less importance for low value goods.

We expect that high value goods are time sensitive, and will be more responsive to transport time. For the same reason, we expect high value goods to be transported in small quantities. Inventory costs are expected to play a larger role for the mode choice for smaller shipment size classes, than for larger. On the other hand, since inventory cost is a combination of the shipment time, and the value of goods, and the value of the goods in units is the same for all shipment size, though adjusted up or down to fit the size categories, there will be no variation between the different shipment sizes. The only variation that can be captured is the variation between the transport chains. We will therefore include the impact of inventory in transport for the transport modes.

Region specific dummies

For freight transport, each transport mode's market share differs between routes, depending on the available infrastructure (especially for transport by rail), length of the route (in km) and the commodity group shipped. The Norwegian transport network can be divided into eight main corridors. The distribution between road and rail as the main transport mode for these corridors is given in Figure 1.



	Oslo-Kr.sand	Oslo-Stav	Oslo-Bergen	Oslo-Åndals	Oslo-Trondh	Oslo-Helgel	Oslo-Bodø	Oslo-Narvik	Trondh-Helgel	Trondh-Bodø
🟩 Rail	63%	52%	78%	23%	58%	70%	89%	91%	11%	50%
🟨 Ship	3%	4%	0%	1%	0%	18%	1%	0%	39%	9%
🟧 Road	34%	44%	22%	76%	42%	12%	10%	9%	50%	41%

Figure 1. Market shares for road, rail and sea for the main transport corridors in Norway, for dry goods, industrial goods and general cargo. Feeder transport is not included. Source: Caspersen and Hovi (2014).

From Figure 1 we can see that rail overall has the highest market shares for long transports to and from Oslo, but smaller market shares on shorter distances, such as transport between Trondheim and Helgeland and Bodø. Moreover, we see that the market shares for rail are around 90 % for transport between Oslo and Bodø, and between Oslo and Narvik. The market share for rail is also relatively high between Oslo and Bergen. For this reason, we are interested in investigating whether there are any differences in the preferred chain choice and shipment size choice for different areas in Norway.

We defined three dummy variables. One for the Northern area (Nordland, Troms and Finnmark), where we expect to find a preference for rail, and if any, sea transport; one for the Western area with Bergen and Stavanger (Rogaland and Hordaland), where we expect a somewhat lower preference for rail and sea, and higher for road transport; and one for the Oslo area (Oslo, Akershus and Vestfold). We expect transports between Sweden and the Oslo area to be relatively short compared to transports to and from the other regions, hence we expect a preference for road transport.

5 Discrete choice model estimation for the joint decision of transport mode and shipment size

The estimation work described is done for the sake of estimating discrete choice models predicting the shipper's joint decision of transport chain and shipment size for each shipment. We have estimated discrete choice models for the three commodity groups 13, 17 and 30 using the 2009 Swedish Commodity Flow Survey, the defined choice set for the three commodity groups (see sections 4.1.3-4.1.5), as well as the explanatory variables previously described (see section 4.2). The model formulations are based on Windisch (2009) and Abate et al. (2014), but we have also added some extensions (e.g. inclusion of other explanatory variables). Similar to Abate et al., we estimate one model per commodity group. However, the utility functions in all three discrete choice models are affected by the same set of explanatory variables. All three models are based on linear additive utility functions and are estimated as multinomial logit models.

In addition to the three multinomial logit models, we extend the estimation of commodity group 17 to a model where the utility functions are non-linear in the variables cost and time. This extended model is estimated as both a multinomial logit and a nested logit model.

All models are estimated with the software Biogeme (Bierlaire, 2003).

5.1 Results from estimation

The results from the main estimation, with estimated coefficients for the three discrete choice models of a joint decision of transport mode and shipment size choice, are displayed in Table 20. The explanatory variables are presented in column one, followed by an explanation of the variables' relevance in the estimated model and the estimated coefficients for each of the three commodity groups in column two to seven. We only present estimation results for explanatory variables that are significant at a significance level $\alpha=0.05$, as well as the alternative specific constants (ASC).

The starting point for estimation was to introduce models consisting of (a mix of) the explanatory variables introduced in chapter 4.2, and testing their relevance in the models. From Table 20 we see that this method has led to the exclusion of inventory costs from all three models. The time variable is excluded in model 13, and only included for rail transport in model 17 (for rail). Variables for time and inventory costs are excluded because they were found insignificant in the estimated models and/or led to misspecification errors.

Variables for transport costs are included in all three models. We estimated one coefficient for each transport alternative in the models, except for transport cost for transport alternatives "water" and "road-water-rail" in model 30, where we found a similar effect of an increase in transport cost.

Table 20. Estimation results for discrete choice model estimation for commodity 13, 17 and 30. Transport alternative R-W-Ra denotes transport chain Road-Water-Rail, Area denotes regional specific dummy for sender/receiver area in Norway, ASC denotes alternative specific constants.

Explanatory variables	Commodity groups					
	13: Iron and Steel		17: Plastic and Rubber		30: Consumables	
	Relevant alternative	*Coeff. estimate*	*Relevant alternative*	*Coeff. estimates*	*Relevant alternative*	*Coeff. estimates*
Cost	Rail	-0.000185***	Rail	-0.000227**	Rail	-0.000459**
	Road	-0.000292***	Road	-0.000207**	Road	-0.000802**
					Water R-W-Ra	-0.00279***
Time			Rail	-0.122***	Rail	-0.104***
					Road	-0.0859***
					R-W-Ra	-0.0734***
Degrad. cost	Road	0.136***			Road	0.00738**
Value density	Rail	-2.17***	Road, Size 1 and 2	-0.602***	Size 1	0.458***
Area: north			Rail	2.67***	R-W-Ra	-1.66***
Area: east	Rail	-5.82***	Rail	-1.1***	R-W-Ra	-0.489***
Area: west			Rail	0.404***	R-W-Ra	-0.249**
ASC	Road	0 (fixed)	Road	0 (fixed)	Road	0 (fixed)
	Rail	11.3***	Rail	3.32***	Rail	-2.10***
					Water	-3.72***
					R-W-Ra	-5.87***
	Size 1	0 (fixed)	Size 1	0 (fixed)	Size 1	0 (fixed)
	Size 2	-1.81***	Size 2	-0.716***	Size 2	1.56***
	Size 3	-1.39***	Size 3	-0.269	Size 3	1.53***
	Size 4	-1.46***	Size 4	-0.636**	Size 4	10.4***
Number of choices		7	8		12	
Estimated parameter		10	11		17	
Observations		781	3,307		25,631	
Final LL-value		-1,052.419	-5,528.798		-29,531.929	
Adjusted ρ^2		0.301	0.194		0.536	

From the table we see that the model for commodity 13 (model 13) is estimated using only 781 observations. It also has a higher explanatory power than the estimated model for commodity 17 (model 17). This is mainly due to the exclusion of degradation cost (for road) in model 17. We excluded degradation cost because it resulted in implementation errors in the logistics model. Exclusion of the degradation cost variable led to a reduction in the model's explanatory power and a reduction of the explained variance in the dependent variable by the variance in the explanatory variables. In addition, the cost variables for road and rail transport lost significance and have a lower impact on the discrete choice when excluding degradation costs from the estimation. Another difference is that the cost coefficient for road is higher than the cost coefficient for rail in the model with degradation cost, but reverse in the model without degradation cost. These differences, resulting from restricting the effect of degradation cost to zero, give us reasons to believe that the variables are correlated.

Excluding degradation cost might also lead to omitted variable bias in the cost variables. The effect of this is that the coefficient estimated for cost also include some of the negative effect of degradation. We experienced some difficulties estimating the cost variables due to correlation between the cost variables for the different transport alternatives, as well as between the time variables and degradation cost variables. Even though correlation between explanatory variables is not problematic in general (the correlation between variables is considered in estimation, such that coefficient estimates are still unbiased and consistent), it generally leads to larger standard errors. The correlation might be artificially high because we use variables for cost, time and degradation from the logistics model (see chapter 4.2). Taking this correlation structure into account, we tried different model structures to improve the model. Given that explanatory variables have a linear impact on the discrete choice, we arrived at the models presented in Table 20 (see chapter 5.2 for a discussion of non-linear effects of (some of) the explanatory variables in the utility function).

The transport time is only included as a variable in model 17 and model 30. In model 17, time has a highly significant negative impact on rail transport. This means that we estimate a negative change in the probability of choosing rail transport when time of rail transport increases. We have not successfully estimated a time coefficient for road transport in model 17. A reason for this might be that time and cost are more collinear for road than they are for rail transport. For model 30, time is a significant variable for three out of four transport modes. For the transport mode “water” time is not found to be significant. For the other transport modes, the coefficient for time is negative and significant. It may seem surprising that transport time has a higher impact on the probability of choosing rail transport than it has for other modes. This can be explained by the fact that in the 2009 Swedish CFS, shipments by rail have a higher value density than road on average (see Table 2). Another explanation may be that for other modes of transport, the negative effect of time can partially be captured by the coefficient for cost due to a higher degree of collinearities between variables.

The value density variable in model 30 is as expected. The variable is implemented for all the smallest shipments (in category 1) and shows a higher preference for smaller shipment sizes when the value density is above the median.

Region specific dummies are included for rail in model 13 and 17, and for the transport alternative including both road, water and rail in model 30. The dummies indicate that rail has a higher probability of being chosen for transports between Sweden and the North and West of Norway (long shipments), but a smaller probability of being chosen for shipments to or from the Oslo area.

Given the three models in Table 20, we chose to compare model 30 to the results from the model estimation by Windisch (2009) and Abate et al. (2014). Model 30 is estimated using more than 25,000 registered shipments, has the highest explanatory power (adjusted ρ^2, which is a form of adjusted pseudo R^2 conventionally used for discrete choice models[13]), and parameter values closest to our expectations. We only compared coefficients for which comparison is relevant. The comparison is presented in Table 21.

This table provides some intuition regarding differences between our estimation results and results from Windisch (2009) and Abate et al. (2014). However, it must be noted that coefficients are not strictly comparable for three main reasons. First, and most importantly, the scale parameter may differ between the models. When scale parameters differ, only the relationship between two parameters can be compared. Second, coefficients may capture different effects, since the explanatory variables included in the models are not exactly the same, and since there is generally a high degree of correlation between the explanatory variables. Third, the regression analyses are not based on the same commodity groups.

Table 21. Comparison between estimation results of cost and density variables between commodity 30: consumables and results from Windisch (2009) and Abate et al (2014).

Expl. variables	**30: Consumables**		**Windisch (2009)**		**Abate et al. (2014)**	
	Relevant alt.	Coeff estimates	Relevant alt.	Coeff estimates	Relevant alt.	Coeff estimates
Cost	Rail	-0.000459**	All chains	-0.0011***	All chains	-0.0001***
	Road	-0.000802**				
	Water	-0.00279***				
Value density	Size 1	0.458***	Size 1-5	-5.79***	Size 1-2	0.122***
			Size 6-9	4.49***		
			Size 1	0.961***		
Observations		25,631	2,225,150		33,868	
Final LL-value		-29,531.929	-1,601,661		-77,652.81	
Adjusted ρ^2		0.536	0.314		0.384	
Commodity group			*All commodities*		*Metal products*	

---

[13] Adjusted rho squared is defined as $\rho^2 = 1 - \frac{\ln(L^*) - K}{\ln(L^0)}$, where L^* is the final (maximum) likelihood of the sample for the estimated model, L^0 is the null likelihood (the likelihood of an intercept model without explanatory variables) and K is the number of parameters that are estimated. A likelihood falls between zero and one, so the log of a likelihood is less than or equal to zero. If a model has a low likelihood, then the log of the likelihood will have a larger magnitude than the log of a more likely model. Thus, a small ratio of log likelihoods indicates that the model is a far better fit than the null model. The last term in the numerator is added to penalize for adding explanatory variables that do not significantly increase the final likelihood. Note that this static is meaningless in the presence of constraints, since the number of degrees of freedom will then be lower than the number of parameters.

From Table 21 we see that both Windisch and Abate et al. assume the same effect of cost on the probability of choosing alternative transport modes. The high cost sensitivity for transport including waterborne model, might be explained by a trade-off between costs and time use.

All three models show that a high value density correlates with smaller shipment sizes. Note that Windisch (2009) estimates three dummy variables for the value density, one for each of the size groups shown in Table 21. The value density dummy variable for size 1 denotes the shipments with the highest value density, above 500 SEK/kg. Abate et al. estimate the impact of a continuous value density variable.

Our estimation of commodity 30 has a higher adjusted ρ^2 than the models estimated by Windisch and Abate et al.

5.2 Extension of model 17

In addition to the models presented in Table 21, we analysed the case where time and transport cost are included as non-linear variables in the utility function, and estimated the model for commodity 17. Commodity 13 lacks observations for the alternative with small shipments transported by rail, and extending commodity 30 might lead to a slowdown of the model estimation in Biogeme, as the number of observations is rather large. Hence, none of these commodities is particularly suitable for extension. Another reason for choosing model 17 is the expectation that the real opportunities for substitution lie between road and rail transport, and less between sea transport and other modes.

In the extension of mode 17, we proceeded as follows:

1. Included a non-linear term for the time variable: We estimated the model with various non-linear transformations of the time variable in the utility function, but ended up with adding a squared term for the time variable, in addition to the linear term. This led to significant coefficients for time variables for both road and rail, with negative effects on the utility. One of the main problems with including time in the linear model specification, presented in table 5.1, was that the coefficient for time for road became positive.
2. Included a non-linear term for the cost variable: Keeping the non-linear term for time, and changing the linear cost variable to a quadratic term of transport cost, yielded a model that seemed to fit the data better than the one in table 4.1. The problem with this specification is that the value of time decreases with time, and becomes negative for slow transports. A log-transformation of the cost variable seemed to give a good representation of the correlation between the variation in the cost variable and the choice. The drawback with this log transformation is that the cost coefficients are positive.
3. Testing a nested structure of the model: We tried to estimate a nested logit model (NL) based on a) a model that included non-linear transformations of cost and time (as mentioned in this sections bullet points 1 and 2) and b) a model that included non-linear transformation of cost and time, where the effect of time differs between shipment size as well as transport modes. We nested between transport modes[14]. This gave two nests: road and rail. None of the nested logit models iterated until the optimum, and the nested logit model for model specification b) did not result in a better log likelihood value than the equivalent multinomial logit

---

[14] A nesting structure based on the four shipment sizes did not yield nest parameters different from zero.

model[15]. Both nested logit models did yield significant log sum-parameters above one, which gives nested-parameters between 0 and 1.

We were not able to estimate a reliable nested logit model with non-linearity in the utility function. Since the multinomial logit model is a restricted version of the nested logit model, this also gives reasons to doubt the estimation of the multinomial model with non-linear variables. For this reason, we chose to keep the simpler models for the discrete choice, where time and cost are linear in the utility function, instead of proceeding with extended model versions. These modelling results are presented in Table 20. Despite these estimation errors, significant and positive log sum parameters give reasons to believe that there is a nesting structure on the discrete choice presented in this documentation report. Other transformations of the variables, careful study, or improvements of the 2009 Swedish CFS dataset, might lead to other estimation results.

---

[15] A nested logit model should always have a better likelihood value than its equivalent multinomial logit model as the nested logit model has at least one more coefficient than the multinomial model, all else being equal.

6 Results from the Norwegian Logistics Model

In the following chapter, we present results from the Norwegian Logistics Model after inclusion of the estimations. We start with a description of the implementation of the estimated coefficients and their corresponding selection criteria functions in to Norwegian Logistics Model. Section 6.2 presents a calculation and discussion of cost elasticities.

6.1 Implementation in Norwegian Logistics Model

The estimated coefficients and their corresponding selection criteria functions have been implemented in the Norwegian Logistics Model for all three commodity groups (13, 17 and 30), partly according to the procedure described in section 2.5. This is done by creating a new version of the executable ChainChoi.exe. This executable uses the same input data, but applies the new selection criteria according to a logistic specification based on the estimated coefficients. The output of this executable is also slightly different; instead of one (deterministic) mode/weight combination for each freight flow, the model now outputs each potential (considered) choice alternative (i.e. each potential combination of mode choice and shipment size) as well as the predicted probabilities that these particular modes and shipment sizes are chosen jointly. Summing over the potential freight flows, multiplied by their corresponding probabilities for each mode, will therefore give the expected mode split predicted by the model. If, for each freight flow, the probabilities of all but one alternative are zero, the result of the stochastic model will be equivalent to the deterministic model.

Compared to the ChainChoi.exe executable used for the deterministic model, four additional steps have been programmed:

1. For each weight class, the number of chain types is reduced to the set used in estimation by a deterministic selection between similar chain types (for example different rail modes);
2. For each weight class and chain type combination, the deterministic part of the utility function is calculated according to the utility specifications and parameter values described in chapter 5, taking into account LoS data from the chain selected in the first step;
3. For each weight class and chain type, the corresponding probabilities are calculated according to the logistics model specification (eq. 5 from section 2.5);
4. If no choice options are available in step 3 (i.e. if the current weight class and chain type is the only element contained in C for the freight flow in question), the deterministic model is applied; the cost-minimizing alternative is chosen with a probability of 100 %.

Ideally, in addition to estimation, the stochastic model should also be calibrated to replicate actual freight flows. The reason for this is that the dataset for which the models are estimated and the dataset for which the Norwegian Logistics Model is used, differ. The estimation procedure ensures that market shares are kept constant; however, this is only a good reflection of reality if the market shares from the estimation data are comparable to the actual market shares that the model will be used to predict.

It is straightforward to calibrate the estimated model to replicate actual market shares; to increase (decrease) a predicted market share, one might add (subtract) a value to (from) the estimated alternative specific constant. However, since we do not observe the chosen combination of weight category and transport chain for all freight flows, we do not have a target to calibrate towards. This is problematic, as it also makes it difficult to determine whether the deterministic or the stochastic model gives the best results. However, at least in aggregate, the market shares of tonne-kms seem to be comparable between the two models. These are displayed in Table 22.

Table 22. Market shares of tonne-kms, for the three commodity groups 13, 17 and 30, and for the deterministic model and the stochastic model, respectively.

Commodity group:	**13: Iron and steel**		**17: Plastic and rubber**		**30: Consumables**	
Model type:*	***D***	***S***	***D***	***S***	***D***	***S***
Road	1.44 %	1.05 %	1.23 %	1.23 %	4.28 %	5.01 %
Sea	97.98 %	97.90 %	98.22 %	98.05 %	91.60 %	90.99 %
Rail	0.58 %	1.06 %	0.55 %	0.72 %	4.13 %	4.00 %

* D: Deterministic model; S: Stochastic model.

The table shows the percentage of total tonne-kms that are utilized by the three main transport modes road, sea and rail, given baseline LoS data and cost coefficients. As the table indicates, the differences between the two ChainChoi executables are almost negligibly small (in average). The table also indicates that for all three commodity groups most of the tonne-kms utilize sea as the main mode of transport.

6.2 Calculation of cost elasticities

Without better data regarding the real structure of freight flows and transport chains in Norway, it is difficult to validate the model properly. However, in this section we will compare the results to those from the deterministic model. One of the most interesting characteristics of these kind of models is how they respond to changes in the input data (simulation/scenario analyses). In this section, we compare the deterministic model to the stochastic model when it comes to demand elasticities for the three transport modes road, sea and rail, with respect to changes in costs.

There are various types of demand elasticities one might consider; we have chosen to look at how the number of tonne-kms by mode is affected by changing time and distance-based link costs. There are two reasons for this choice. First, due to limited time and budget we must restrict the comparison to a subset of the input data. We believe that time and distance-based link costs are most appropriate; they are relevant for policy analyses, as well as straightforward to implement in the modelling system. Second, a similar analysis is conducted in Abate et al. (2016), in which elasticities from a deterministic and a stochastic version of the Swedish Logistics Model are compared. We have chosen to examine the same elasticities as Abate et al., in order to present an analysis that is as comparable as possible. The scenarios we examined are displayed in Table 23 below. By an elasticity, we mean the percentage change in the output (number of tonne-kms) following a 1 % change in the input (time and distance based link costs). In Table 23, the scenarios for which we calculate elasticities are -15%, -5%, +5% and +15% changes in costs, for road, sea and rail, respectively.

Table 23. Scenarios for comparisons and calculation of elasticities for the deterministic and the stochastic NTM.

	Decrease in time and distance based link costs		Constant link costs	Increase in time and distance based costs	
Road	- 15 %	- 5 %	Base value	+ 5 %	+ 15 %
Sea	- 15 %	- 5 %	Base value	+ 5 %	+ 15 %
Rail	- 15 %	- 5 %	Base value	+ 5 %	+ 15 %

We calculate both own elasticities and cross-elasticities. Before discussing these elasticities, some points are worth noting:

- The elasticities show the percentage change in tonne-kms as a result of a 1 % change in time and distance based link costs (i.e., not the change in market shares of tonne-kms). In practice, this means that the total amount of tonne-kms can increase (decrease) due to a modal switch that increases (decreases) the total distance travelled. This is reflected in the elasticities.
- For better readability, the elasticities are displayed with the same sign as the cost change (i.e., the two first columns, -15% and -5%, show the percentage change in tonne-kms as a result of a *reduction* in costs, while the two neighbouring columns, +5% and +15%, show the percentage change in tonne-kms as a result of an *increase* in costs).
- Since the NTM operates with fixed demand matrices, the elasticities only indicate the partial effect of changing the transport structure for existing freight flows. This means that new freight flows generated as a result of the cost reduction are not taken into account. Neither are changes in freight flows resulting from a new industry structure, due to the change in transport costs. For such analyses, additional steps must be taken to predict changes in the demand matrices. In other words, the number of kms (summed over all transport modes) may vary, but the number of tonnes (summed over all transport modes) is kept constant.
- The transport chains are aggregated to the three main transport modes “Road”, “Sea” and “Rail”:
  - The transport flows that use a transport chain consisting of sea/rail, as well as road, are merged with sea/rail.
  - The transport flows that use sea *and* rail are almost exclusively transports from the southern parts of Norway to the northern parts. These transports use rail for most of the distance, and are therefore merged with “rail”.
  - The transports that use ferry from other countries (which is a separate transport mode in NTM) are merged with “Road”, as they are usually part of truck transports.
- In the tables below, we compare the conventional (deterministic) NTM with the newly developed (stochastic) NTM. It is however important to realise that some of the freight flows in the stochastic version are still determined in a deterministic manner (see step four above). This is particularly the case for commodities 13 and 17. Comparing only the stochastic flows from the stochastic NTM to their deterministic counterparts would therefore have led to larger discrepancies between the stochastic and the deterministic model.
- The fact that we only display the results of changes in time and distance specific link costs has some implications for the interpretations. This is the main element of total costs for road transport, and a large share of total costs for rail transport. However, it is a smaller share of the total costs for sea transport. The reason is that costs related to operations at ports are relatively large. Therefore, we expect that

changing the time and distance-specific link costs for sea transport will lead to a smaller change in tonne-kms for all modes.

The elasticities are presented in Table 24, Table 25 and Table 26 below, for commodity groups 13, 17 and 30, respectively.

Table 24. Elasticities calculated in deterministic and stochastic versions of the NTM for commodity group 13: Iron and steel. Own elasticities are shown in bold.

Deterministic model											
Road				**Sea**				**Rail**			
-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%
1.13	**0.45**	**-0.25**	**-1.13**	-0.45	-0.11	0.16	1.04	-0.07	-0.01	0.00	0.01
-0.02	-0.01	0.04	0.09	**0.08**	**0.04**	**0.00**	**-0.03**	0.00	0.00	0.00	0.00
0.19	0.09	0.18	0.41	-2.95	-0.46	0.18	1.23	**0.63**	**0.18**	**-0.25**	**-0.79**
Stochastic model											
Road				**Sea**				**Rail**			
-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%
2.45	**0.99**	**-1.00**	**-1.66**	-0.94	-0.44	0.11	0.49	-0.13	-0.05	0.11	0.32
-0.02	-0.01	0.07	0.10	**0.08**	**0.06**	**0.00**	**-0.01**	0.00	0.00	0.00	0.01
-1.26	-0.47	0.07	0.12	-0.63	-0.24	0.04	0.31	**0.49**	**0.08**	**-0.25**	**-0.72**

Table 25. Elasticities calculated in deterministic and stochastic versions of the NTM for commodity group 17: Plastic and rubber. Own elasticities are shown in bold.

Deterministic model											
Road				**Sea**				**Rail**			
-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%
1.72	**0.30**	**-0.39**	**-1.27**	-0.19	-0.14	0.01	0.18	-0.07	-0.04	0.01	0.06
-0.03	-0.02	0.01	0.01	**0.02**	**0.00**	**-0.01**	**-0.02**	-0.02	-0.01	0.00	0.01
-0.60	0.76	0.01	1.69	-1.78	-0.37	1.13	1.72	**1.95**	**1.32**	**-0.58**	**-1.60**
Stochastic model											
Road				**Sea**				**Rail**			
-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%
1.50	**0.30**	**-0.31**	**-1.27**	-0.21	-0.01	0.06	0.11	-0.09	-0.04	0.04	0.10
-0.02	-0.01	0.01	0.01	**0.01**	**0.00**	**0.00**	**-0.01**	-0.01	0.00	0.00	0.01
-0.58	-0.08	0.08	1.38	-0.64	-0.13	0.34	0.64	**1.11**	**0.51**	**-0.20**	**-0.99**

Table 26. Elasticities calculated in deterministic and stochastic versions of the NTM for commodity group 30: Consumables. Own elasticities are shown in bold.

Deterministic model											
Road				**Sea**				**Rail**			
-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%
1.62	**0.57**	**-0.27**	**-1.09**	-0.01	0.00	0.00	0.00	-0.22	-0.04	0.18	0.53
0.00	0.00	0.00	0.01	**0.00**	**0.00**	**0.00**	**0.00**	0.00	0.00	0.00	0.00
-1.53	-0.54	0.25	0.86	-0.01	-0.01	0.03	0.08	**0.34**	**0.10**	**-0.24**	**-0.73**
Stochastic model											
Road				**Sea**				**Rail**			
-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%	-15%	-5%	+5%	+15%
0.60	**0.21**	**-0.22**	**-0.69**	0.00	0.00	0.00	0.00	-0.07	-0.02	0.02	0.07
0.00	0.00	0.00	0.00	**0.00**	**0.00**	**0.00**	**0.00**	0.00	0.00	0.00	0.00
-0.79	-0.28	0.29	0.94	0.00	0.00	0.00	0.00	**0.12**	**0.04**	**-0.04**	**-0.13**

Examining these elasticities, some conclusions stand out, and some comments are in place:

- For all commodity groups, and for both the deterministic and the stochastic model, all own elasticities (shown in bold) have the expected sign. In other words, decreasing the cost of a certain mode increases demand, while increasing the cost decreases demand.
- For all commodity groups, and both the deterministic and the stochastic model, all cross elasticities for sea and rail scenarios have the expected sign (columns 5-12). In other words, increasing (decreasing) the cost of sea/rail leads to an increase (decrease) in tonne-kms for the other modes. This is not the case for road transport, since road also is a part of the transport chains defined as "sea" and "rail". This is to say, changing the cost for road transport will to some extent change the cost for the transport chains "sea" and "rail" as well.
- For commodity groups 17 and 30, the own elasticities in the stochastic case are consistently lower than the own elasticities in the deterministic case. For commodity group 13, there is no clear pattern. Since cost is not the only factor influencing the mode choice in the stochastic model, we would expect this model to be less responsive. However, the potential all-or-nothing assignment that characterises deterministic models might drag the results either way, depending on whether the simulated cost change is large enough to induce a change in the mode choice or not.
- Even though there seems to be some pattern supporting this, elasticities are not strictly symmetrical around zero; in some cases, they differ substantially between increasing and decreasing the costs. This is to be expected, especially taking into account that the cost changes we consider are relatively large. However, it is reasonable to believe that for marginal (very small) changes, elasticities would be more or less symmetrical.
- For virtually all commodity groups, all transport modes and both types of models, for both own elasticities and cross elasticities, the absolute value of the elasticity increases when the magnitude of the cost change increases.[16] It is difficult to come up with an explanation for such a strong relationship. Moreover, in Abate et al.

---

[16] The only exception is the cross elasticity for rail when reducing the cost of road transport for commodity group 17 in the deterministic model (see Table 25).

(2016), elasticities seem to diminish instead of increase as the change in cost increases.[17] This might be a result of the underlying structure of the transport market; however, firm conclusions would require further examination of specific freight flows.

- For all commodity groups and both model types, the elasticities for sea are the lowest in absolute value. There are three main reasons for this. First, sea clearly has the largest market shares to begin with. This means that a large absolute change in tonne-kms will lead to a smaller relative change compared to the other modes. Second, as stated earlier, changing time and distance- based link costs for sea will lead to a relatively smaller cost change than for road or rail. The reason is that other cost elements arise when commodities are transported by sea, in particular costs at ports. Such cost elements are kept constant in all scenarios. Third, a large share of the tonne-kms by sea arises as a result of long-distance transports from other countries. In reality, there are often no alternative modes of transport for these flows. This is particularly the case for commodity group 30, consumables, for which almost all sea transport consists of exports or imports.
- For all commodities and both model types, changing the costs for sea will to a large extent lead to a mode shift to/from rail. The only exception is the stochastic model for commodity 13, for which the shift is larger between sea and road than between sea and rail. Changing the costs for rail (road) will lead to a mode shift to/from road (rail). This follows intuition, as it indicates that land based transport modes (road/rail) as well as long distance transport modes (sea/rail) are closer substitutes.
- All elasticities lie within the same range as those presented in Abate et al. (2016). However, the elasticities for sea transport are significantly lower. The main reason for this is that Abate et al. only consider transport distances carried out in Sweden. In our study, we take into account the whole transport distance for all transports, also imports and exports. Since distances abroad account for a large share of the tonne-kms by sea, and these often do not have an alternative mode of transport, our elasticities for sea will naturally be lower.

---

[17] However, this phenomenon is not strictly comparable, as their cost changes are 15 and 40 percent, as opposed to 5 and 15 percent.

7 Discussion and further work

In this final chapter, we provide a brief discussion of model validation on other commodity groups and provides suggestions for further work.

7.1 Model validation on other commodity groups

Reducing the data set from the Swedish commodity flow survey to cover only transport between Norway and Sweden, leaves us with a minor number of commodities that fulfil the requirements regarding number of shipments and variation across transport modes and geographical areas. The commodities analysed are presented in Table 27, together with four other commodities, which to a lesser extent fulfil the criteria for analysis. The table includes information about the number of shipments for each transport mode and commodity in the table, and geographical dispersion (number of origin zones) in Norway and Sweden.

Table 27. Commodities for estimation. Total number of shipments and different geographical zones in Norway and Sweden, separated between transport modes. Source: Swedish CFS (2009).

Commodity		Rail		Ship		Vehicle	
Import and export	Country	Swe.	Norw.	Swe.	Norw.	Swe.	Norw.
30 Consumables	Shipments	4,043		18,432		48,883	
111 Commodities that don't belong to any other group	Shipments	2,590		2,112		2,850	
	Zones	11	400	9	350	130	250
100;109 Furniture, textiles and the like	Shipments	1,453		16,320		46,033	
	Zones	12	170	6	400	80	450
17 Plastic and rubber	Shipments	1,269		4		2,067	
	Zones	9	170	4	4	81	274
13 Iron and steel	Shipments	436		45		301	
	Zones	6	12	4	3	45	56
14 Other metals	Shipments	25		12		403	
	Zones	8	10	5	3	77	103
16 Chemical products	Shipments	24		82		894	
	Zones	5	16	13	37	82	136
27 Electric equipment	Shipments	245		32		10 919	
	Zones	17	71	12	22	157	426
32 Transport	Shipments	34		19		1 422	
	Zones	9	15	9	11	69	215

From Table 27, it is obvious why commodity 13, 17 and 30 are most suited for estimation. For other commodity groups, we either have too few observations and/or too little spread of transport modes and/or geographical areas. Still, a possible way to validate the estimated choice models for the discrete choices between shipment size categories and transport modes, is to estimate the models using other commodity groups. For example, we could estimate the model developed for commodity 13 for commodity 14, and the model for commodity 17 for commodity 16. Alternatively, we could use commodity 30 as a template to estimate the discrete choice for other commodity groups, as commodity 30 is the largest group in terms of number of shipments. Moreover, the shipment has a relatively good spread over geographical zones.

Another possibility to improve the models is to increase the number of observations and the spread of mode choice and geography. This can be done either by increasing the dataset from the commodity flow survey to cover more than transport between Norway and Sweden, or replace it completely with another data source. An alternative data source recently available is the Norwegian foreign trade statistics, that TOI has access to at shipment level, with detailed information about origin, destination, shipment size and value, and border crossing mode choice.

7.2 Model with discrete mode and continuous shipment size choice

As discussed in sections 2.1-2.4, discretising the choice of shipment size can be seen as a form of measurement error. An alternative is to estimate a simultaneous discrete-continuous structural model based on eqs. 2.1 and 2.2 (the joint choice of discrete transport mode and continuous shipment size). In this case, the error terms of the two equations of interest (ε_i and ϵ_i) are likely to be correlated because the transport planner makes the choices between transport chains and shipment sizes as parts of the same optimization process, and it is unlikely that we are able to capture all relevant variation in the choices by observed variables. This potential correlation can be controlled for by following the procedure described in Holguin-Veras (2002). We could model the discrete choice component (eq. 2.1, section 2.3) as the structural equation of interest, replacing actual shipment size with a prediction from a shipment size auxiliary regression (eq. 2.2, section 2.3). Using this prediction, we can calculate a variable $V_i = ABS(M_i - \gamma^e)$ which gives the absolute difference between the average observed shipment size M for a certain mode i and the estimated shipment size γ^e. We hypothesise that at its average observed shipment size, the capacity of the mode and the shipment match very well (also assuming that most shipments are not consolidated with others). When the shipment deviates more from this average (either smaller or larger), the probability of choosing that mode for this shipment will decrease. We thus expect a negative estimated coefficient for this variable.[18]

---

[18] In the project in Sweden, reported in Abate et al. (2014), not only the Holguin-Veras discrete-continuous model is estimated, but also a specification following Dubin and McFadden (1984). This specification is less interesting when the focus is on transport chain or mode choice, as is the case with the National Logistics Models, and therefore not used in the current paper.

References

Abate, M., Vierth, I. and de Jong, G.C. (2014). Joint econometric models of freight transport chain and shipment size choice. Report prepared for Trafikverket, VTI, Stockholm.

Abate, M., Vierth, I., Karlsson, R., de Jong, G. & Baak, J. (2016). Estimation and implementation of joint econometric models of freight transport chain and shipment size choice. Stockholm: CTS working paper 2016:1.

Ben-Akiva. M., de Jong. M. (2013). The Aggregate-Disaggregate-Aggregate (ADA) Logistics Model System, in Moshe Ben-Akiva, Hilde Meersman, Eddy Van de Voorde (ed.) Freight Transport Modelling, pp.69 - 90

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models , Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Caspersen, E. & Hovi, I.B (2014). Vulnerability and emergency response in freight transport. TØI report 1324/2014. Oslo.

Grønland, S.E. (2015). Cost models for freight transport and logistics – base year 2012. TØI report 1435/2015. Oslo.

Grønland, S.E. (2012). Cost models for transportation and logistics. TØI report 1127/2011. Oslo.

Johnson, D. and de Jong, G.C., 2011: Shippers' response to transport cost and time and model specification in freight mode and shipment size choice. Proceedings of the 2nd International Choice Modeling Conference ICMC 2011, University of Leeds, United Kingdom.

Jong, G.C. de and Ben-Akiva, M., 2007: A micro-simulation model of shipment size and transport chain choice. *Transportation Research Part B* 41(9), 950–965 (2007).

Jong, G.C. de, Ben-Akiva, M., Baak, J. and Grønland, S. (2013) Method Report – Logistics Model in the Norwegian National Logistics Model System (Version 3).

Jong, G.C. de and Johnson, D., 2009: Discrete mode and discrete or continuous shipment size choice in freight transport in Sweden. Paper presented at the European Transport Conference - 2009, Noordwijkerhout (NL).

Significance (2008) *Method Report – Logistics Model in the Norwegian National Logistics Model System.* Deliverable 6A for the Working group for Transport analysis in the Norwegian national transport plan.

Significance and SITMA, 2013: Method report – logistics model in the Norwegian National Logistics Model system (version 3). Report for the Work group for transport analysis in the Norwegian national transport plan. Project 12028, Significance, The Hague.

SIKA Statistik, 2006. Varuflödesundersökningen 2004/2005. 2006:12.

Trafikanalys, 2010: Varuflödesundersökningen 2009. Sveriges officiella statistik.

Windisch, E., de Jong, G.C and van Nes, R., 2010: A disaggregate freight transport model of transport chain choice and shipment size choice. Paper presented at ETC 2010, Glasgow.

Institute of Transport Economics (TØI)

Norwegian Centre for Transport Research

Established in 1964, the Institute of Transport Economics is an interdisciplinary, applied research centre with approximately 70 professionals. Its mission is to develop and disseminate transportation knowledge that has scientific quality and practical application.

A private, non-profit foundation, TØI receives basic funding from the Research Council of Norway. However, the greater part of its revenue is generated through contract research. An important part of its activity is international research cooperation, mostly in the form of projects under the Framework Programmes of the European Commission.

TØI participates in the Oslo Centre for Interdisciplinary Environmental and Social Research (CIENS) located near the University of Oslo. See www.ciens.no

TØI covers all modes of transport and virtually all topics in transportation, including road safety, public transport, climate change and the environment, travel behaviour, tourism, land use and urban planning, decision-making processes, freight and travel demand, as well as general transport economics.

Claiming copyright to its products, TØI acts independently of its clients in matters of scientific approach, professional judgment and evaluation. TØI reports are generally downloadable for free at www.toi.no.

Visiting and postal address:
Institute of Transport Economics
Gaustadalléen 21
NO-0349 Oslo

+ 47 22 57 38 00
toi@toi.no
www.toi.no