Centre for Transport Studies
STOCKHOLM

Estimation and implementation of joint econometric models of freight transport chain and shipment size choice

Megersa Abate – VTI
Inge Vierth – VTI
Rune Karlsson – VTI
Gerard de Jong – Significance
Jaap Baak – Significance

CTS Working Paper 2016:1

Abstract

As part of the further development of the Swedish national freight model system (SAMGODS), we developed a stochastic logistics model in the form of a disaggregate random utility-based model of transport chain and shipment size choice, estimated on the Swedish Commodity Flow Survey (CFS) 2004-2005. Moving from the current deterministic logistics model within the SAMGODS model to a stochastic one, is important because it bases the model on a stronger empirical foundation. The deterministic model was not estimated on observed choice outcomes, but just postulates that the least cost solution will be chosen.

We estimated logit models which explain the joint choice of shipment size (in discrete categories) and transport chain separately for sixteen different commodity types. A transport chain (e.g. truck-vessel-truck) is a sequence of modes used to transport a shipment between the locations of production and consumption. Transport cost, travel time and value density are some of the main determinants included in the models. It is important to note that by their very nature these probabilistic models 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 in the random utility model, the response functions (now expressed in the form of probabilities) become smooth instead of lumped at 0 and 1 as in a deterministic model. This in turn will address the problem of "overshooting" that is prevalent in a deterministic model when testing different scenarios or policies.

Centre for Transport Studies
SE-100 44 Stockholm
Sweden
www.cts.kth.se

vti

KTH VETENSKAP OCH KONST

For two of the commodity types (metal products and chemical products) for which we estimated a transport chain and shipment size choice model, we also implemented the model in the SAMGODS framework. The implementation takes place at the level of the annual firm-to-firm flows by commodity type between producing and consuming firms that are generated by the first steps of the SAMGODS model (PC flows between zones that have been allocated to individual firms at both ends). For every firm-to-firm flow, shipment size and transport chain choice probabilities are calculated and added over the firm-to-firm flows of the PC relation (sample enumeration, as used in several disaggregate transport models). From this, the aggregate OD matrices by mode can be derived straightforwardly, as well as results in terms of tonne-kilometres by mode. It was not possible to empirically model transshipment location choices, because they are not stated in the CFS. Therefore, the determination of the optimal transshipment points for each available chain type from the set of available locations is still done deterministically.

The implemented models were applied to produce elasticities of demand expressed in tonne-kilometres for various changes in cost and time for road, rail and sea transport. These elasticities are compared to those for the same commodity types in the deterministic model and to the available literature. The elasticities clearly differ between the two models, they are usually smaller (in absolute values) in the stochastic model, as expected.

In the paper, we report the basic differences between a stochastic and a deterministic logistics model, the estimation results for the sixteen commodities, the way the stochastic model was implemented within the SAMGODS model, the elasticities that we obtained for the implemented stochastic model and the comparison with elasticities from the deterministic model and the literature.

Keywords: Freight, Choice model, SAMGODS

Estimation and implementation of joint econometric models of freight transport chain and shipment size choice

Megersa Abate[1] Inge Vierth[1] Rune Karlsson[1]

Gerard de Jong[2] Jaap Baak[2]

1 Introduction

As part of further development of the Swedish national freight model system (SAMGODS) model, in this project we setup a stochastic logistic model. To move from the current deterministic logistic model within the SAMGODS model to a stochastic one, first and foremost, it is important to base the underlying behavior of transport agents on a stronger empirical foundation.[3] While there are several factors which determine firms' choices of shipment size and transport chain, cost is the only variable considered in the deterministic version of the SAMGODS model. Such an approach has rather weak empirical foundation and needs to be improved by thoroughly analyzing the main determinants of these choices.

We used the 2004/5 Swedish Commodity Flow Survey to estimate Multinomial Logit Models (MNL) models which explain the joint choice of shipment size and transport chain. Analyzing the shipper/transport agents' decisions using such a disaggregate and revealed preference data set has improved model prediction and allowed the stochastic model to mimic realistic freight transport decisions.

The MNL models provide coefficient estimates for the determinants of transport chain and shipment size choices. Transport cost, travel time and value density are some of the main determinants included in the MNL models estimated in this project. It is important to note that by their very nature these are probabilistic models because they include a stochastic component to

[1] Swedish National Road and Transport Research Institute (VTI)

[2] Significance

[3] The development of a stochastic model was planned from the beginning (see, SIKA 2004) but postponed several times.

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 in the utility model, the response functions (now expressed in the form of probabilities) become smooth instead of lumped at 0 and 1 as in a deterministic model. This in turn will address the problem of "overshooting" [4] that is prevalent in a deterministic model when testing different scenarios or policies.

The choice models applied in this project were estimated for various commodity categories separately, but not for all commodity groups that are in the current SAMGODS. It is important to note that for some commodity groups (e.g. bulk commodities which are very strongly connected to one mode of transport) a deterministic model may still be good enough. Here, due consideration needs to be given how to move from the extended NSTR classification to the NST 2007 classification for commodities. (The NST 2007 classification is used in transport statistics since 2007.)

In summary, to establish a version of SAMGODS that is based on random utility modelling, the following steps were undertaken:

1. Estimation of joint econometric models of shipment size choice and transport chain choices for selected commodity groups.
2. Implementation of the utility functions and their coefficients to determine shipment size and transport chain choice probabilities and applying the model at the level of firm-to-firm flows using sample enumeration. It was not possible to empirically model transhipment location choices, because they are not stated in the Commodity Flow Survey (CFS). Therefore, the split between the determination of the optimal transhipment points and the choice of transport chain was kept separate (as in the deterministic model). The determination of the optimal transhipment locations for each available chain type from the

[4] "Overshooting" happens when the relevant part of the logistics costs function is rather flat and a small change in logistics costs can lead to a shift to a completely different optimum shipment size and transport chain (Abate et al. 2014). On the other hand there could also be "sticky" choices in a deterministic (all-or-nothing) model when one alternative is clearly cheaper than the other alternatives. Improving the other alternatives will then not lead to any change in market shares until one of these other alternatives becomes the cheapest and then the deterministic choice is suddenly completely altered.

set of available locations is still done deterministically. The random utility model in the new SAMGODs will refer to shipment size and transport chain choice.

The rest of the report is organized as follows: Section 2 presents the discrete choice model set up and results from estimation; Section 3 describes the stochastic model setup based on the inputs from Section 2; Section 4 compares model outputs from the stochastic and deterministic models; finally, Section 5 draws conclusions from this work and points to future research needs.

2 Econometric model, Data and Results

2.1 Empirical framework

Econometric studies of freight mode choice involve joint models because mode choice entails simultaneous decisions on how much to ship (see, for example, Abate and de Jong, 2013; Johnson and de Jong, 2011; Holguin-Veras, 2002; Abdelwahab 1998; Abdelwahab and Sargious, 1992; Inaba and Wallace, 1989; McFadden et al., 1986). This simultaneity in decisions requires the use of econometric techniques such as discrete-continuous (DC) models. An alternative is a discrete-discrete (DD) model by classifying shipment sizes into a number of size classes (as in Johnson and de Jong and (2011) and Windisch et al. (2010)). In addition to recognizing this simultaneous decision process, these studies show that various haul, carrier, and commodity characteristics affect the decisions regarding the optimal shipment size choice and choice of transport mode.

During the first phase of the stochastic freight model development, three econometric models were tested for suitability (see Abate et al 2014 for details) and the DC model which treats transport mode chain choice as a discrete and shipment sizes as continuous was found to be theoretically sound. However, given the size of the CFS data and the number of commodity groups involved, a pragmatic alternative is a DD model where both transport mode chain and shipment size choices are treated as discrete alternatives. What follows presents a short description of the DD model.

A joint model with discrete mode and discrete shipment size choice is specified as:

$$U_i = \beta X_i + \varepsilon_i \tag{1}$$

Where U_i is the utility derived from choosing a discrete combination of transport chain and a shipment size category i, X_i is a vector of independent variables explaining mode choice and shipment size choice, β is a vector of parameters to be estimated and ε_i is an error term. The model setup allows for simultaneous consideration of transport chain and shipment size decisions. The main variables included in X_i are transport cost, transport time, infrastructure access indicators, value density, domestic/international shipment indicators. We estimate Eq. 1 using a multinomial Logit model (MNL).[5]

2.2 Data

The main data source for this paper is the 2004/2005 Swedish Commodity Flow Survey (CFS). The data has 2,986,259 records. Each record is a shipment to/from a company in Sweden, with information on origin, destination, modes, weight and value of the shipment, sector of the sending firm, commodity type, access to rail tracks and quays, etc. In the CFS a shipment is defined as a unique delivery of goods with the same commodity code to/from the local unit or to/from a particular recipient/supplier (SIKA, 2004). From this we selected a file of around 2,897,010 outgoing shipments (domestic transport and export, no import) for which we have complete information on all the endogenous and exogenous variables.

Although the CFS data is extensive, it does not contain information on important variables such as transport costs and transport time. Given the importance of these variables in mode/shipment size choice analysis, the logistic module of the SAMGODS model was used to generate transport cost and time variables for each shipment in the CFS based on a number of the transport mode chain and shipment size combinations (see below for definition of these combinations). These variables were generated both for the chosen mode-shipment alternatives in the CFS and for potential non-chosen alternatives tailored to each shipper based on the transport network of the origin and destination of their shipment.

Table 1 below presents descriptive statistics for the main variables of interest for selected commodity groups in the CFS.[6] The mean values of transport cost, shipment distance and transport

[5] There could be correlations between alternatives, especially given that there are alternatives that have a transport chain (or a shipment size) in common. The MNL model assumes choice alternatives are independent, and therefore could suffer from the 'red bus – blue bus' problem (that is that similar alternatives should have higher cross-elasticities, but do not have these in MNL). A relatively straightforward solution in such cases is the nested logit model. Windisch et al. (2010) tested various nested logit models on the CFS 2004-2005 (but not by commodity) and found that a nesting structure with transport chain choice above shipment size choice worked best (this means that there is more substitution between shipment sizes than between modes). More complicated nesting structures can be tried in mixed logit and multivariate probit models, but these model types have very long run times, especially on large data sets as we have here.

[6] For model estimation, we found that it is more instructive to analyse selected commodities than all commodities identified in the CFS. The main reason for this the fact that trucking is the most dominant transport chain for some commodity groups (for example, the share of trucking is more than 98 per cent for ten commodity groups). There is

time are from the SAMGODS model. The remaining variables are from the 2004/5 CFS. In the whole CFS, 2 per cent of firms had access to rail at the origin of their shipment and 0.4 per cent of them had access to quay. Paper pulp and Iron ore shippers have by far the highest access to rail. Paper pulp shippers also report the highest share of access to quay at origin of a shipment.

2.3 Main results

Table 2 presents results from the MNL models for the commodity groups presented on Table 1. The choice alternatives in each model are a discrete combination of a transport mode chain and shipment size. By and large, the results reported in Table 2 are plausible and are in line with expectations. As expected, for most commodities transport cost has a negative effect on the utility of a choice alternative, implying that higher delivery costs make a chain less attractive. We have used a single cost coefficient for all alternatives, building on the idea that 1 SEK is 1 SEK, whatever the alternative it is spent on. Other forms than linear could be tried for the cost specification (such as logarithmic, spline or a combination of linear and logarithmic), but for a comparison with the deterministic model, it is best to use a linear cost specification, since the deterministic model also uses linear costs. While this effect is statistically significant, the parameter values are small, which can imply that cost has a rather limited influence on a choice alternative. It is important to note, however, that the unit of measurement and dimensions of change all contribute to this low level estimates.

The variable for inventory costs during truck transport (transport time times value of the shipment) has the expected (negative) sign and is highly significant for most commodity groups. This variable captures time costs related to the capital cost of the inventory in transit and maybe also those related to deterioration and safety stock considerations. The time-dependent link-based transport costs (labour and vehicle costs) have already been taken into account in the transport costs. Estimation of separate transport time times value coefficients for road, rail and vessel transport did not lead to significant coefficients. This variable, representing capital costs on the inventory in transit, can be expected to be most relevant for truck transport and turns out to be statistically significant for this mode, but the point estimates are rather low. Including the size of firms did not lead to intuitive results.

---

little to learn about the determinants of mode choice decisions of shippers when there is such overwhelming dominance of one mode of transport. For the remaining 16 commodity groups presented on Table 1 there is relatively less dominance of trucking.

The access to rail/quay dummy variables were included in the utility functions of choice alternatives where rail/quay was used as the first or second mode in the chosen transportation chain. The interpretation of the parameter values is that shippers located in the proximity of rail track or quay yard are more likely to choose chains that start with a rail/quay leg (or use these modes on the second leg of the chain). The two dummies are, however, not significant for most commodity groups.

For most commodity groups, we find a significant positive effect for the value density variable. This implies that high value products correlate with smaller shipment sizes, which might also imply frequent shipments. We also find that international shipments tend to be shipped using chains that use rail, ferry or vessel. The transportation chain-specific constants mostly have negative signs and are significant for metal products. This is expected given that trucking, the reference chain type, is preferred to the other modes for its flexibility and ease of access (which are not measured in the CFS). For some commodity groups, the effect of the alternative specific constants is positive.

3 Stochastic model set-up

The stochastic logistics model has been estimated on shipments from the CFS 2004-2005. In application we do not use the CFS records, but we apply the estimated transport chain and shipment size models to the annual firm-to-firm (f2f) flows that are also used in the current SAMGODS model (these f2f flows can thus remain the same for the stochastic model)[7]. For every f2f flow within a commodity group, the stochastic logistics model now predicts the choice of transport chain and shipment size and it does so by producing choice probabilities for every available alternative.

So far, the stochastic logistics model has been implemented (programmed) for the commodities 17 (metal products) and 23 (chemical products) so that a comparison can be made between the deterministic and the stochastic logistics model in terms of their outcomes and sensitivities

During application of the stochastic logistic model the following steps are performed:

a) *Determine the longlist of transport chains using the existing BUILDCHAIN program.* This step fully corresponds to the corresponding step in the deterministic model. Transport

[7] It would be helpful for building the stochastic model if the CFS would also ask for the annual volume or annual transport frequency of the goods for which information on a specific shipment is collected.

chains with optimal transshipment locations are determined for each of the chain types distinguished within the deterministic model. For these chains, transport distance and time are calculated. The BUILDCHAIN program reads unimodal Level of Service matrices for all possible chain leg modes. It then builds optimal chains using a one-to-many algorithm that follows a stepwise approach in adding extra legs to chains and determining the optimal transfer locations. This algorithm is explained in detail in the method report for the Swedish National freight Model System (Significance, 2015).

b) *Reduce the number of chain types to the more limited set (shortlist) distinguished in the stochastic model by a deterministic choice amongst similar chain types.* Within the deterministic model several rail modes (kombi train, feeder train, wagonload train, system train) and sea modes (direct sea, feeder vessel, long-haul vessel) are available. On the other hand, within the stochastic model only one rail and one sea mode are distinguished. To select the rail and sea modes to be used in the stochastic model, as well as to determine the vehicle types to be used on each leg, the deterministic model is applied. This has to be done for all of the available weight class choice option separately. After step (b) the best chains and vehicle types are available for the chain types and weight classes available in the stochastic model:

Chain types:

Truck

Vessel

Rail

Truck-Vessel

Rail-Vessel

Truck-Truck-Truck

Truck-Rail-Truck

Truck-Ferry-Truck

Truck-Vessel-Truck

Truck-Air-Truck

Truck-Ferry-Rail-Truck

Truck-Rail-Ferry-truck

Truck-Vessel-Rail-Truck

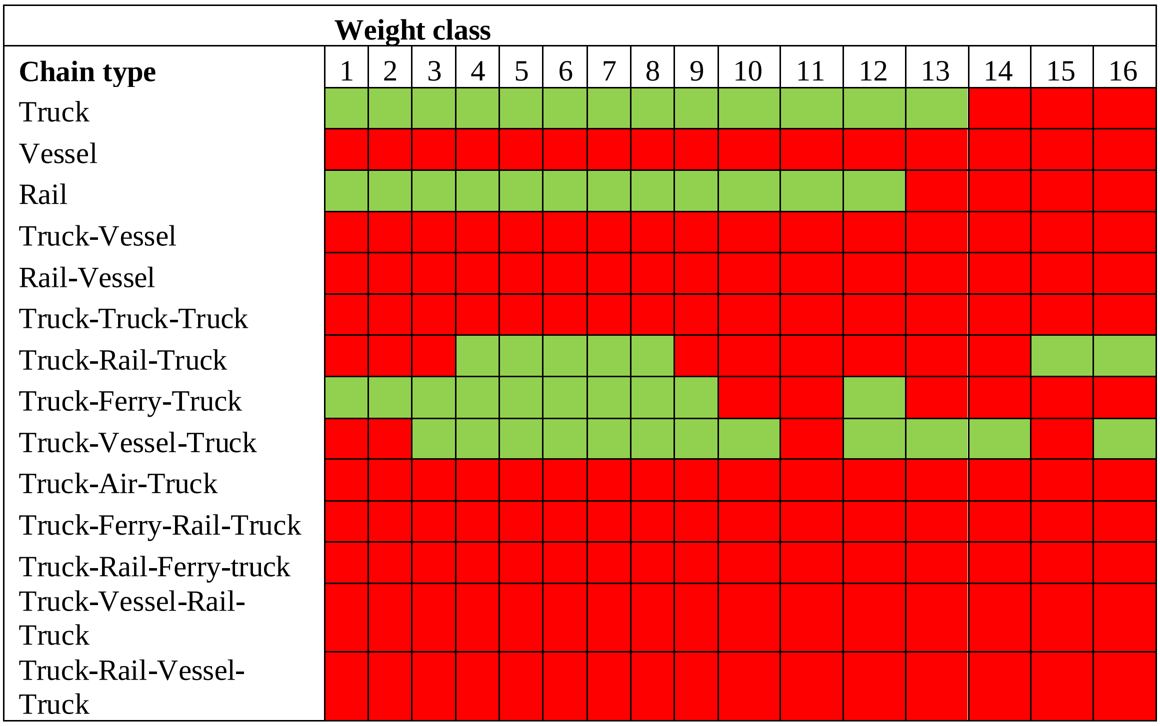
Truck-Rail-Vessel-Truck

16 weight classes:

0	-	50	kg
50	-	200	kg
200	-	800	kg
0.8	-	3.0	tonnes
3.0	-	7.5	tonnes
7.5	-	12.5	tonnes
12.5	-	20	tonnes
20	-	30	tonnes
30	-	35	tonnes
35	-	40	tonnes
40	-	45	tonnes
45	-	100	tonnes
100	-	200	tonnes
200	-	400	tonnes
400	-	800	tonnes
> 800			tonnes

However, not all the above choice options will be available for each commodity. As an example, Figure 1 shows the combinations of chain type and weight class that are available in the stochastic model for commodity 17 (based on the actual frequencies in CFS 2004-2005).

Figure 1: Available combinations of chain type and weight class (red=unavailable, green=available) for commodity 17 Metal products, based on the frequencies observed in the CFS 2004-2005



Calculate the utilities for each of the choice options in the stochastic model. In step (b) the number of available chain types has been reduced to at most 13, the number of chain types distinguished within the stochastic model. Within the third step the utility functions are calculated for each of the available choice options (combinations of transport chain and shipment size) given above. The estimated coefficients are multiplied with the relevant chain parameters obtained from the chains determined in step (b). Within CHAINCHOI there is no information available on the value of goods or the value density on specific firm-to-firm relations. Therefore the average commodity value of these variables by commodity type is used in application of the model. The dummy coefficient for direct rail access is always applied to chains consisting of a single rail leg and never for the other chains. Quay access is not used in the implemented models for metal and chemical products.

c) *Calculation of the choice probabilities.* When the utilities have been calculated for all available chain types and weight classes, the probability for each choice option can be calculated as:

$$P(\{c,w\}) = \exp(U_{\{c,w\}})/\Sigma_{\{c',w'\}}\exp(U_{\{c',w'\}})$$

where {c, s} denotes the choice option with chain type *c* and weight class *w*. The numerator is the exponentiated utility function of the alternative whose probability is being calculated, whereas the denominator is the sum over all available alternatives of their exponentiated utilities.

d) *Aggregation of flows.* Similar to the deterministic model, all firm to firm flows are aggregated to obtain OD-flows. However, instead of the single best chain generated by the deterministic model, we now aggregate over all choice options and weight each choice option with the probability calculated in step (d).

All of the steps (b) to (e) are executed in the CHAINCHOI program.

4 Deterministic vs. Stochastic, a comparison using two commodities groups

4.1 Method

The stochastic approach applied in this paper is intended to be a substitute or complement to the deterministic model, which currently constitutes the very heart of the logistics model in the Samgods model system. Both the deterministic and the stochastic model have been implemented into an executable, ChainChoi.exe, for metal products (commodity 17) and chemical products (commodity 23). By switching these executables, we may conveniently switch between the deterministic and the stochastic model when running Samgods. Both models operate on the same set of input data when it comes to demand matrices and costs for 2006. The stochastic logistics model makes also use of the CFS 2004/5. The framework of this project did not allow the use of the CFS from 2001 and 2009 or updated LOS-data and base matrices to study how sensitive estimation results are to different input data (different future values for variables like value density and value of goods for future years can be assumed in different scenarios; in this paper, we have used current values only).

Two other ongoing projects are also using the Swedish CFS to estimate stochastic logistics models and to implement these in a transport forecasting model. In Norway, the Institute of Transport Economics (TØI), working in cooperation with Significance, have estimated transport chain and shipment size models on observations from the CFS 2009 that relate to Norway. The estimated models have also been implemented in a stochastic version of the national model that can be compared to the deterministic one. For the European Commission DGMOVE, the CFS 2009 is used together with the French shipper-carrier survey ECHO to estimate logistics models that will be implemented at the European scale.

The Samgods system can hence be run for each of these models and comparisons of the outcome can be done. The topic of the current section is to empirically investigate and analyze differences in results. All results in this section have been obtained using Samgods version 1.0 (April 2015). It should be noted that the ChainChoi.exe program carries out computations without considering the Rail Capacity Management (RCM) procedure. A special program has been developed to handle the RCM case, ChainChoi4RCM.exe in the deterministic model. However, no corresponding stochastic procedure has been developed yet. Therefore the comparisons between the deterministic and the stochastic case are restricted to the standard logistic model (without RCM).

There are two ways to present the results from the stochastic logistics model. One is to present the probabilities to choose certain transport chain types (chain frequencies and weight classes) and the other to multiply the probabilities with path flows in order to present flows on the network. So far the first method has been applied; the probabilities for the different chain types are presented in addition to the quantities that are included in the deterministic model. For commodities 17 and 23 the number of relations are about 14 times higher when different probabilities are included. The second method allows in principle output in the same format as in the deterministic model. However, in order to implement the method additional work is needed. We recommend to adjust the output data to the CUBE-format (and not to adjust the CUBE format).

The results presented here (see i.e. Table 3) are derived from the direct output from the logistics model. These are less precise than those from corresponding assigned quantities, and introduces an extra uncertainty in the results, in particular when it comes to computed tonkm within Swedish territory.

4.2 Calibration procedure for the stochastic model

The stochastic logistics model described above has been estimated on the CFS 2004-2005. The model includes alternative-specific constants for all transport chain alternatives (minus one). This means that the model will reproduce the market shares (in terms of the number of shipments) for the chains as they are in the estimation data (which is based on the CFS, but also depends on the question whether we have level-of-service data for a particular chain and PC relation). This is not necessarily a good reflection of the actual importance of the various modes for the commodity involved. We also have observed aggregate data on the tonne-kilometers by mode (e.g. from traffic counts). For metal products and chemical products these numbers for the year 2006 are in the columns labelled 'statistics' in Table 3.

When we compare the tonne-km by mode (by OD-leg, so also access/egress tonne-km are counted) from the uncalibrated stochastics model to these observations, we see that it overestimates the road and the sea tonne-km for both products. For metal products there is some underestimation of rail, and for chemical products the stochastic model predicts a very limited (less than one million ton-km) use of rail transport. This is in line with the CFS, but not with the calibration data (where rail has a market share of more than 10% for chemical products). The deterministic logistics model (without the rail capacity module) on the other hand overestimates the observed rail tonne-km.

To calibrate the stochastic logistics model, we use the observed tonne-km shares as targets and add to each transport chain alternative constant in the utility functions of the stochastic model:

$\text{Ln}\ (O_j/M_j)$

In which:

O_j: observed share of mode j

M_j: Modelled share of mode j

This makes under-predicted modes more attractive and over-predicted ones less attractive. To reach the observed targets, this procedure needs to be repeated several (probably many) times; it is an iterative calibration procedure. For the comparison of elasticities in this report we performed a couple of iterations with the stochastic model for both metal products and chemical products, which brought us much closer to the observed targets, but still not very near.

4.3 Comparison of model prediction on selected output measures

The starting point in the comparisons is the base scenario. We have chosen to use the uncalibrated default base scenario in Samgods 1.0, Base2006. This was originally calibrated for the RCM module, so results from the standard logistics model (without RCM) deviates significantly from official statistics, as can be seen from Table 3. For example, the total rail tonne-km is much larger in model output than in the transport statistics. Such deviations in the base scenario will have consequences in computed elasticities.

In the first step we checked the outcome of the model runs against the statistics. Table 3 below shows that both the deterministic and the stochastic model overestimate the tonne-km in Sweden a lot. This needs to be checked ideally with help of the visualization of statistics and model predictions (see above).

Another observation that can be made is that the deterministic model calculates relatively high shares for rail while the stochastic model calculates relatively high shares for road and sea. Both the overestimation of the total tonne-km and the deviation from the modal split in the statistics will have consequences for the calculation of the elasticities.

4.4 Comparison of elasticities

Scenarios

Of major interest is to compare the model's responses to small (or larger) perturbations in input data, i.e. elasticities. The Samgods model comprises large sets of both input and output data. Only a few elasticities have been possible to investigate here (and one has to take into account that the total demand per commodity is constant and be aware of the problems described above). Our choice has been to study, on the input side, the link costs that comprise the distance- and time based costs for all vehicle types within road, rail and sea and on the output side, tonne-km in Sweden. In Table 4 we summarize the scenarios investigated.

Results for metal products

In Table 5, results for change in tonne-km in Sweden[8] are shown for the different scenarios, computed with the deterministic and stochastic model. The following are some of the interesting results that jump out:

[8] Tonne-km in Sweden is the sum of the domestic transports and the domestic parts of international transports that are carried out in Sweden.

- all own-price elasticities have the expected sign
- the own elasticities for changes in road and rail cost are in all cases smaller in the stochastic model than in the deterministic model. Especially for road cost changes, the stochastic model elasticities are more plausible (e.g. they do not become as strong as -2.87 as in the deterministic model). For changes in the sea transport cost, some elasticities are stronger in the deterministic model and some in the stochastic model. The elasticities can differ substantially between cost increases and decreases (in a logit model elasticities for increases and decreases do not have to be the same, this depends on where the starting point is located on the S-shaped logit curve).
- Nearly all cross price elasticities have also the opposite sign of the direct elasticity, which is what one should expect from a model in which the modes would be mutually exclusive ('competing') alternatives. However, both the deterministic and the stochastic logistics model have transport chains in which several modes are combined (e.g. with rail as main haul mode and truck for access and egress). As a result, increasing the cost of rail transport could lead not only to an increased share of the truck only chain (competition), but also to a reduced truck use in the truck-rail-truck chain (complementarity)[9]. This usually refers to rather short road access and egress distances, but still it reduces the elasticities (in absolute values) and can even lead to cross elasticities with the same sign as the own-price elasticities.
- it is reasonable to believe that the inclusion of other factors than costs in the stochastic model and the move away from the all-or-nothing choice in the deterministic model reduces the modal shifts (that are calculated for the deterministic model)
- most of the shifts (in both models, especially in the stochastic model) are from/to the land based modes to/from sea, which is not necessarily expected
- transfers to/from rail are very low in the stochastic model. This could imply that current rail shippers are captive to the mode to some extent (note that metal products is characterized by the dominance of one big shipper). On the other hand, it could also imply that other modes are competitively priced to rail, implying that larger price incentive or availability of infrastructure is needed to attract more shippers to rail.

[9] Furthermore there can also be changes in shipment size in both models as a result of cost changes.

Results for chemical products

In Table 6, results for change in tonne-km in Sweden are shown for the different scenarios, computed with the deterministic and stochastic model. The following are some of the interesting results that jump out:

- in all cases, the own-price elasticities have the expected sign
- the own elasticities for changes in road, rail and sea transport cost are smaller in the stochastic model than in the deterministic model. For all these three cost changes, the elasticities of the stochastic model seem more plausible (the deterministic model has elasticities here that go beyond -2). Again, there are substantial differences between cost increases and decreases.
- in most cases the cross price elasticities have also the opposite sign as the own elasticity. For the stochastic model, this is always the case, but for the deterministic model, there are stronger complementarities between modes.
- it is reasonable to believe that the inclusion of other factors than costs in the stochastic model reduces the modal shifts (that are calculated for the deterministic model)
- large differences in modal split in the base (see Table 3) lead to very different elasticities
- transfers from road to rail (in the stochastic model) are higher for chemical products than for metal products. Also the own elasticity of rail costs is stronger for chemical products than for metal products. This is all probably due to the lower share of rail transport for chemical product shippers compared to metal product shippers. Given this low share of rail in chemical product shipments, any price incentive will attract shippers to shift to rail.
- For chemical products, sea transport has a higher share than for metal products. This is reflected in the elasticities of the stochastic model which yields stronger sea cost elasticities in the model for metal products than for chemical products.

Overall results

Elasticities differ according to commodities, regions (modal split etc.), distance class, modelling approaches and measures (ton, tonne-km, vehicle-km), see e.g. de Jong et al. (2010). This source does not contain recommendations per commodity type. For the all commodities the recommended road tonne-km price elasticity on the number of tonne-km by road through mode choice in de Jong et al. (2010) is -0.4 and the lower bound provided is -1.3. Some of the road costs elasticities of the deterministic model for metal and chemical products are clearly beyond this lower bound. The own elasticities, measured in tonnes, calculated with help of a weighted logit mode-choice model for the Öresund region (Rich, Holmblad & Hansen (2009) are in about the same range as the own elasticities measured in tonne-km calculated in this paper.

5 Conclusions and ideas for further development

As part of further development of the SAMGODS model, in this project we have setup a stochastic logistic model for two commodity groups, metal products and chemical products. Although the stochastic model is implemented for the two commodities, we have estimated multinomial logit models for 14 commodities for which a stochastic model could be implemented in the future. We compared transport cost and time elasticities for tonne-km between the stochastic and deterministic models for the two commodities, which has not been done before for such models. These elasticities differ between the two models, they are usually smaller in the stochastic model. In future endeavors, the difference between the two models could be further studied by looking at elasticities on other output measures such as vehicle kilometer etc.

Finally, to have a full-fledged stochastic model the following further steps are absolutely needed (point 1- 3 below) or would by useful to develop:

1. The CUBE programme that is used to run the SAMGODS model needs to be adapted to handle multiple choice alternatives with a probability in the infrastructure network. The model results in this report were derived from the logistics model by itself.

2. Identification for which commodities the stochastic logistics model should be used and for which commodities the deterministic model should be used. Consideration needs to be given how to move from the commodity classification NSTR to the commodity classification NST 2007.
3. Estimation of remaining stochastic logistic models

4. Using updated/better data to study how sensitive estimation results are to different CFS, LOS-data from Samgods and base matrices

5. Analyses if/how the stochastic logistics model(s) can be combined with the rail capacity management tool (RCM)

6. Update all stochastic logistics models (see point 2. and 3. above) to Samgods Version 1.1 (updated vehicle types, costs and base matrix 2012) The current stochastic model is based on version 1.0 of the deterministic model. Currently version 1.1 of the deterministic model is in an advanced stage of development. This version includes updated cost parameters, but also the introduction of new model modes like inland waterways and 74 tonne lorries. Once version 1.1 has been completed, further improvements are foreseen that will result in version 1.2 of the deterministic model. These changes to the deterministic model must also be included in the final version of the stochastic model.

7. Test and validate elasticities in total logistic model (commodity wise for stochastic models and deterministic model) for base year 2012

8. A test and validation of the resulting OD flows by mode against observed aggregate data for a (new) base year 2012 is recommended, since a model estimated on the CFS will not necessarily match with other data, such as traffic counts. Preferably this will be done on the basis of assignment results (for which the adaptation under 1 is required).

9. Update all stochastic logistics models to Samgods Version 2 (based on CFS 2016)
10. Test of other formulations of the cost functions (e.g. log, spline) This could reduce potential scale problems. For comparison with deterministic model it's best to use linear cost in the stochastic model.
11. Test of other choice sets (to find out if another choice set sample would give different estimates (This could be investigated by re-sampling methods such as the Jack-knife)
12. Test if differences in preferences between different f2f-relations can be integrated into the logistics model (e.g. value density)
13. Test how sensitive the estimation (and prediction) is to the alternatives that are included in the choice-set is for each f2f-relationship (the choice set generation).
14. Test how much accuracy would be lost by allowing only one transhipment choice per type chain in the deterministic module and in the stochastic model
15. Analysis how to calibrate whole model (cost parameters etc.) for both deterministic model and stochastic model simultaneously
16. Reduce run time and storage space in order to manage complexity and data volumes

Acknowledgements

The authors would like to thank Christian Hansen Overgård and Jonas Westin for useful comments on an ealier version of the paper and the Swedish Transport Administration for funding the project.

References

Abate, M., Vierth, I., & de Jong, G. (2014). *Joint econometric models of freight transport chain and shipment size choice.* Stockholm: CTS working paper 2014:9.

Abate, M. and de Jong, G.C., 2014. The optimal shipment size and truck size choice- the allocation of trucks across hauls. *Transportation Research Part A*, 59(1) 262–277

Abdelwahab, W., & Sargious, M., 1992. Modelling the demand for freight transport: a new approach. *Journal of Transport Economics and Policy*, 49-70.

de Jong, G., Schroten, A., van Essen, H., Otten, M., & Bucci, P. (2010). Price sensitivity of European road freight transport - towards a better understanding of existing results - A report for Transport & Environment. Delft: Significance & Delft.

Holguín-Veras, J., 2002: Revealed Preference Analysis of Commercial Vehicle Choice Process. *Journal of Transportation Engineering,* 128 (4), 336--346.

Inaba, F.S. & Wallace, N.E., 1989: Spatial price competition and the demand for freight transportation. *The Review of Economics and Statistics*, 71 (4), 614--625.

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, 4 - 6 July.

McFadden, D., Winston, C., and Boersch-Supan, A., 1986: Joint estimation of freight transportation decisions under non-random sampling. In: A. Daugherty, ed. Analytical studies in transport economics. Cambridge University Press, 137--157.

Rich, J., Holmblad, P., & Hansen, C. (2009). A weighted logit freight mode-choice model. *Transportation Research Part E (Vol 45)* , 1006–1019.

Significance (2015) Method Report - Logistics Model in the Swedish National Freight Model System, Deliverable for Trafikverket, Report D6B, Project 15017, Significance The Hague.

SIKA ,2004: The Swedish National Freight Model: A critical review and an outline of the way ahead, Samplan 2004:1, SIKA, Stockholm

Vierth, I., Jonsson, L., Karlsson, R., & Abate, M. (2014). *Konkurrensytta land - sjö för svenska godstransporter.* VTI (VTI Rapport 822/2014).

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

Table 1 Descriptive Statistics

Commodity group*	Rail Access (%)	Quay Access (%)	Shipment weight (KG)	Shipment value (SEK)	Value density** (SEK/KG)	Transport costs (SEK)	Transpor time (Ho
Wood (6/7)	15	0.05	14,176	53,150	2,599	7,618	7.1
Textile (09)	0.4	0.006	77.32	12,040	786	4,200	5.9
Iron ore (15)	46	0	4,158,336	2,328,523	13.2	162,468	10.9
Nonferrous ore and waste (16)	32	0.13	119,762	931,699	1,203.6	1.63e+09	3.4
Metal products (17)	57	0.5	6,556	31,943	24	3,684	3.5
Earth, gravel (19/20)	1	0.12	88,461.6	37,745	17.4	13,302	3.1
Coal chemicals (22)	0.1	0.04	1,732.3	1,124,820	14,728	6,423.9	7
Chemical Products (23)	0.03	0.03	4,023	42,907	288	6,783	10.37
Paper pulp (24)	66	0.42	112,297	448,287	8.4	29,636	25
Transport Equip (25)	2	0.004	827.8	77,913	1,094	8,347	6.3
Metal manufactures (26)	5	0.01	2,291	56,254	431.6	3,803	4.8
Glass (27)	1	0.02	1,680	27,209	139.8	4,111.9	5.4
Leather textile (29)**	2	0.004	488.8	13,978.5	2,416		
Machinery (32)	4	0.003	265.7	25,725.6	8,030	10,423.8	3.1
Paper board (33)	6	0.02	6,170	43,117	424.9	6,228	4.7
Wrapping material (34)	50	0.004	28,007	51,538.6	4.4		
All Commodities	2	0.4	26,011	37,122	1,231		

*SAMGODS commodity classification number in parenthesis. **Note that the mean of the value density variable is not calculated by dividing the mean values of sl sample. It is calculated as the mean of the value density for each shipment in the CFS. The two values could be close to each other if both variables are greater than however, the weight and value variables are recorded as having values less than one in the CFS, which explains the difference between the two statistics.

Table 2 Multinomial Logit model results

Variable	Relevant alternative	Parameter Estimates						
		Textile (09)**	Iron ore (15)	Nonferrous ore and waste (16)	Coal chemicals (22)	Paper pulp (24)	Transport Equip (25)	Meta (26)
Cost (SEK per shipment)	All chains	-0.000852 (-7.77)	0.000599 (0.67)	-4.44e-006 (-3.68)	-0.00015 (-9.34)	-1.36e-005 (-5.48)	-3.03e-006 (-10.14)	-0.00 (-17.
Transport time (in hours) times value of goods (in SEK)	Truck	-3.24e-008 (-2.08)	-8.31e-006 (-0.05)	-1.07e-007 (-2.33)	-9.09e-009 (-2.89)	-1.71e-006 (-9.25)	-6.78e-008 (-7.58)	-5.80 (-2.4
Dummy variable for access to rail track	Rail	-0.0479 (-0.02)		5.70 (16.96)		0.640 (2.51)	-0.0313 (-0.06)	-0.44 (-1.2
Dummy variable for access to quay	Ferry/vessel	-0.165 (-0.29)		1.93 (1.97)	-0.282 (-0.47)		1.36 (3.53)	-0.00 (-0.0
Value density (SEK/KG)	All modes: smallest 2 shipment sizes	0.0182 (36.33)	-0.05 (1.08)	0.000456 (0.97)	0.000315 (9.41)	0.001 (2.96)	0.0156 (46.73)	0.01 (26.3
Dummy variable for international shipment	Rail, Ferry, Vessel	0.881 (8.17)		3.89 (4.56)	1.21 (4.36)	1.14 (3.99)	4.84 (53.39)	0.18 (1.60
Dummy variable for Air	Constant				0.135 (0.00)		-7.78 (-25.64)	0.00 (0.00
Dummy variable for rail	Constant					-2.27 (-7.08)		
Dummy variable for Truck-Rail-Truck	Constant	-10.6 (-11.91)		-8.06 (-26.50)		-6.16 (-5.99)	-7.99 (-53.24)	-8.24 (-29.
Dummy variable for ferry	Constant	-2.49 (-18.27)		-5.56 (-7.20)	-2.08 (-9.33)	-2.52 (-8.69)	-6.25 (-74.17)	-0.84 (-7.1
Dummy variable for Rail-Vessel	Constant					-4.17 (-7.34)		
Dummy variable for Truck-Vessel-Truck	Constant					-7.67 (-14.02)	-6.10 (-69.80)	-0.01 (-0.0
Dummy variable for truck	Constant	Fixed						
Number of observations Final log-likelihood Rho-square		22623 -24350.2 0.637	59 -13.797 0.792	555 -1336.3 0.193	925 -2064.1 0.208	632 -1628.3 0.14	29616 -39717.7 0.641	3696 -670 0.47

** SAMGODS commodity classification number in parenthesis.

Table 2 continued…

Variable	Relevant alternative	Parameter Estimates				
		Leather textile (29)**	Machinery (32)	Wood (6/7)	Earth, gravel (19/20)	Metal (17)
Cost (SEK per shipment)	All chains	-0.000661 (-13.74)	-0.000160 (-73.41)	-3.01e-005 (-10.74)	-5.75e-006 (-4.48)	-1.96 (-5.2
Transport time (in hours) times value of goods (in SEK)	Truck	-8.43e-008 (-2.15)	5.67e-00 (0.49)	-1.78e-007 (-5.08)	-5.39e-006 (-8.53)	-3.78 (-14.
Dummy variable for access to rail track	Rail	-0.0165 (-0.05)	-0.148 (-0.43)	5.37 (10.85)	-0.0496 (-0.06)	0.70 (4.42
Dummy variable for access to quay	Ferry/vessel	-0.0165 (-0.02)	-0.0125 (-0.03)	0.653 (3.83)	3.09 (6.90)	
Value density (SEK/KG)	All modes: smallest 2 shipment sizes	0.035 (38.75)	0.0156 (18.99)	0.0368 (26.05)	0.0600 (7.10)	0.13 (149.
Dummy variable for international shipment	Rail, Ferry, Vessel			5.69 (22.17)	5.50 (14.85)	3.09 (33.3
Dummy variable for Truck-Air-Truck			0.0071 (0.20)			
Dummy variable for rail	Constant	-0.520 (-3.38)		-10.7 (-18.33)	-6.33 (-7.95)	
Dummy variable for Truck-Rail-Truck	Constant	-1.61 (-16.48)	-4.45 (-12.42)			-3.97 (-122
Dummy variable for Truck-Ferry-Truck	Constant	-0.0751 (-0.53)	-2.68 (-7.72)		-6-71(-21.48))	-4.28 (-95.
Dummy variable for Vessel	Constant	-4.53 (-31.30)		-4.34 (-43.83)		
Dummy variable for Truck –vessel-Truck	Constant		-3.04 (-116.5)	-3.73 (-23.33)	-4.86 (-16.91)	-5.74 (-63.
Dummy variable for Truck-Rail-Vessel-Truck			-3.83 (-131.53)			
Dummy variable for truck	Constant	Fixed				
Number of observations Final log-likelihood Rho-square		55357 -71392.4 0.625	91329 -121097.6 0.642	16765 -39952.108 0.324	2597 -7058.9 0.158	33908 -81898 0.383

** SAMGODS commodity classification number in parenthesis.

Table 3 Million tonne-km for metal products and chemical products within the borders of Sweden according to Trafikanalys transport statistics 2006, * *deterministic model and stochastic model*

Million tonne-km	**Metal products**			**Chemical products**		
	Statistics	Deterministic model	Stochastic Model	Statistics	Deterministic model	Stochastic model
Road	1,217	2,195	4,790	1,608	1,883	2,794
Rail	4,972	6,908	5,301	482	2,013	558
Sea	801	2,509	1,945	1,803	1,843	2,150
Total	**6, 990**	**11,612**	**12,036**	**3, 893**	**5,738**	**5,501**

*See Table 5 in (Vierth, Jonsson, Karlsson, & Abate, 2014) , 1/3 of the international road transports performed inside and outside Sweden are included.

Table 4 Scenarios for comparisons between deterministic and stochastic model

	Decrease in distance- and time based link costs		**Constant link costs**	**Increase in distance- and time based link costs**	
Road	-45%	-15%	base	+15%	+45%
Rail	-45%	-15%	base	+15%	+45%
Sea	-45%	-15%	base	+15%	+45%

Table 5 Elasticities calculated in deterministic and stochastic model for all transports of metal products on Swedish territory

Deterministic model											
ROAD				RAIL				SEA			
-45%	-15%	+15%	+45%	-45%	-15%	+15%	+45%	-45%	-15%	+15%	+45%
-2,87	**-2,82**	**-1,10**	**-0,94**	0,81	0,79	0,84	0,53	0,33	0,73	0,15	0,13
0,78	0,63	0,49	0,41	**-0,80**	**-1,03**	**-0,78**	**-0,69**	0,38	0,21	0,25	0,28
0,21	0,83	-0,13	-0,18	1,04	1,58	0,97	1,31	**-1,84**	**-2,06**	**-0,91**	**-0,80**
Stochastic model											
ROAD				RAIL				SEA			
-45%	-15%	+15%	+45%	-45%	-15%	+15%	+45%	-45%	-15%	+15%	+45%
-0,49	**-0,32**	**-0,11**	**-0,04**	0,02	0,01	0,03	0,10	0,07	0,10	0,27	0,18
0,21	0,09	-0,05	-0,07	**-0,02**	**-0,02**	**-0,03**	**-0,12**	0,03	0,00	-0,08	-0,02
1,11	1,75	0,78	0,48	0,01	0,01	0,00	0,03	**-0,53**	**-0,69**	**-1,62**	**-1,03**

Table 6 Elasticities calculated in deterministic and stochastic model for all transports of chemical products on Swedish territory

Deterministic model											
ROAD				RAIL				SEA			
-45%	-15%	+15%	+45%	-45%	-15%	+15%	+45%	-45%	-15%	+15%	+45%
-2,14	**-1,58**	**-2,22**	**-1,01**	0,68	0,61	0,10	0,25	0,57	1,40	-0,35	0,24
1,37	1,39	0,73	0,66	**-1,83**	**-2,18**	**-1,25**	**-1,41**	1,18	0,73	1,31	0,48
0,14	0,00	0,55	0,02	1,20	1,69	1,29	1,39	**-2,04**	**-2,00**	**-1,56**	**-0,70**
Stochastic model											
ROAD				RAIL				SEA			
-45%	-15%	+15%	+45%	-45%	-15%	+15%	+45%	-45%	+45%	-15%	+15%
-0,52	**-0,32**	**-0,19**	**-0,12**	0,02	0,04	0,10	0,08	0,07	0,06	0,07	0,04
0,97	0,69	0,54	0,24	**-0,29**	**-0,56**	**-0,51**	**-0,50**	0,01	0,00	0,01	0,01
0,19	0,03	0,09	0,06	0,02	0,04	0,00	0,00	**-0,14**	**-0,30**	**-0,28**	**-0,07**