Recommendation for a new commodity classification for the national freight model Samgods[1]

Inge Vierth - VTI
Samuel Lindgren - VTI
Gerard de Jong – Significance
Jaap Baak – Significance
Inger Beate Hovi – TOI
Moa Berglund – WSP
Henrik Edwards – SWECO

CTS Working Paper 2017:11

The objective of this report is to recommend a new commodity classification for the next version of the Swedish national freight model system Samgods. The recommendation is based on i) a comparison of commodity classifications in transport models in other countries, ii) an evaluation of classifications in from the viewpoint of modelling transport demand, iii) how well the classification captures behavioral differences among firms in the freight market and iv) statistical considerations. We recommend the classification to be based on the divisional level of the NST 2007 and to include commodity groups 1-14. We recommend to split up group 1 into one category containing round wood and another containing the rest of the items. We also think it is useful to add a commodity group for air freight by combining fractions from other commodities. In total, our recommendation consists of 16 groups.

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

[1] The authors thank Henrik Edwards, Petter Wikström and other seminar participants from Trafikverket for valuable comments. We are also thankful to Henrik Petterson at Trafikanalys for helpful comments and for providing us with information about the commodity classifications in freight transport statistics. Funding from Trafikverket is gratefully acknowledged.

Content

1 INTRODUCTION

In 2007, Eurostat replaced the NST/R[2] commodity classification for freight transports by the NST 2007[3] classification (see Appendix 2). The difference between the two is that NST/R is based on the physical characteristics of the goods while NST 2007 considers the economic activity from which the goods originate. Each of its items is connected to an item of the European Union product and activity classifications CPA[4] and NACE.[5,6] The highest level of classification in NST/R contains 10 divisions and the highest level in NST 2007 contains 20 divisions. Countries can apply different sub-divisions.

The change of commodity classification has implications for the national and international freight transport models in Europe as these are typically specified per commodity. Within the model systems the same commodity classification is used for describing transport demand and the choice of logistic and transport solutions.[7]

The objective of this project is to recommend a new commodity classification for the next version of the Swedish national freight model system Samgods. The project is funded by the Swedish National Transport Administration (Trafikverket) who is responsible for the development of the Samgods model. The project builds on earlier work in which the correspondence between existing industry and commodity classifications was analyzed (see section 3 and Appendix 1). The main outcome from that analysis was that a new commodity classification based on NST 2007 should be considered for Samgods. In this project, we therefore pay particular attention to what a classification derived from the NST 2007 framework should look like.

Our recommendation takes several aspects into account. In chapter 2, we compile the commodity classifications used in different national and international transport models and identify similarities and differences between them. We also discuss what Sweden can learn from other countries. In chapter 3, we evaluate different alternatives for commodity classifications from the viewpoint of modelling transport demand in the Swedish and Norwegian national freight transport model system. In chapter 4, we analyze behavioral differences among shippers in the Swedish freight market and assess how well the NST 2007 classification captures these differences. In chapter 5, we review additional statistical considerations for the development of the new commodity classification. The recommendation for a new commodity classification in chapter 6 is based on the outcomes in chapter 2-5.

---

[2] NST/R = Standard Goods Classification for Transport Statistics/Revised 1967
[3] NST = Standard goods classification for transport statistics
[4] CPA = Classification of products by activity
[5] NACE = Statistical classification of economic activities
[6] CPA and NACE are consistent with their counterparts at the UN level, CPC and ISIC.
[7] Since 2007 it is for example difficult or not possible at all to validate commodity specific model results at against official statistics due to the difference in commodity classification.

2 COMMODITY CLASSIFICATIONS IN DIFFERENT MODELS

The commodity classification applied in various national freight transport model systems are presented in section 2.1 and the classification in model systems that comprise several countries in section 2.2. Possible implications for the Samgods model are discussed in section 2.3.

2.1 National level

Sweden

There are 35 commodities based on NST/R in the Samgods model system (Trafikverket, 2016). In the NST/R24 classification, commodities are represented by 24 commodity groups with subdivisions, making up a total of 30 commodity groups. For the Samgods model, four commodities are further divided due to their varying logistic properties such as value and shipment size. For example, the group Paper and pulp is split into Paper pulp and waste paper (24) and Paper, paperboard and manufactures thereof (33). Further, a commodity group for goods transported by air freight (35) is created by allocating fractions of certain commodities to this group. Commodities 8, 30 and 34 are not used in the Samgods model version 1.1[8]. This means that the Samgods model operates with 32 commodities in total.

All commodities are associated with an aggregate commodity type: dry bulk, liquid bulk or general cargo. Table 1 presents the 35 commodity groups along with the NST/R code and aggregate commodity types that are used to specify transfer costs.

[8] For number 8, the reason is that it was not clear during the generation of transport demand matrices which products to include since it is formulated as a residual commodity ("Other wood and cork"). The data necessary for the construction of transport demand matrices are unavailable for commodity groups 30 and 34 (Trafikverket, 2016).

Table 1. Commodity classification in Samgods model system version 1.1

Samgods no.	Commodity	NST/R code	Aggrgate commodity
1	Cereals	011-019	Dry bulk
2	Potatoes, other vegetables, fresh or frozen, fresh fruit	020, 031-039	Dry bulk
3	Live animals	001	Dry bulk
4	Sugar beet	060	Dry bulk
5	Timber for paper industry (pulpwood)	051	Dry bulk
6	Wood roughly squared or sawn lengthwise, sliced or peeled	052, 056	Dry bulk
7	Wood chips and wood waste	057	Dry bulk
8 (not used)	Other wood or cork	-	Dry bulk
9	Textiles, textile articles and manmade fibres, other raw animal and vegetable materials	041-049, 091-099	General cargo
10	Foodstuff and animal fodder	111-179	General cargo
11	Oil seeds and oleaginous fruits and fats	181-182	Liquid bulk
12	Solid mineral fuels	211-233	Liquid bulk
13	Crude petroleum	310	Liquid bulk
14	Petroleum products	321-349	Liquid bulk
15	Iron ore, iron and steel waste and blast-furnace dust	410, 462-467	Dry bulk
16	Non-ferrous ores and waste	451-459	Dry bulk
17	Metal products	512-568	General cargo
18	Cement, lime, manufactured building materials	641-692, 992	Dry bulk
19	Earth, sand and gravel	611-615	Dry bulk
20	Other crude and manufactured minerals	621-639	Dry bulk
21	Natural and chemical fertilizers	711-729	Dry bulk
22	Coal chemicals	831-839	Liquid bulk
23	Chemicals other than coal chemicals and tar	811-820, 891-896	Dry bulk
24	Paper pulp and waste paper	841-842	Dry bulk
25	Transport equipment, whether or not assembled, and parts thereof	910	General cargo
26	Manufactures of metal	941-949	General cargo
27	Glass, glassware, ceramic products	951-952	General cargo
28	Paper, paperboard; not manufactures	972	Dry bulk
29	Leather textile, clothing, other manufactured articles than paper, paperboard and manufactures there	961-971, 975-979, 993-999	General cargo
30 (not used)	Mixed and part loads, miscellaneous articles	-	General cargo
31	Timber for sawmill	055	Dry bulk
32	Machinery, apparatus, engines, whether or not assembled, and parts thereof	920-939	General cargo
33	Paper, paperboard and manufactures thereof	973-974	General cargo
34 (not used)	Wrapping material, used	991	Dry bulk
35	Air freight		General cargo

Norway

The logistics model of the Norwegian national freight model system has been developed by Significance, TOI and Sitma (de Jong et al, 2008), and has a similar functionality as the Swedish model. However, the commodity classification is adapted for the Norwegian manufacturing industry, and the classification is based on the different commodities' needs for transport requirements. The classification is also taking into account the demand model's (Spatial Computable General Equilibrium model (SCGE) Pingo) need for consistency with the logistics model and the national account statistics. The commodity classification is based on the NACE classification (and the SITC-classification for foreign trade) and not the NST 2007 classification. The reason for this is that the commodity flow survey for Norway includes detailed information about delivering firms' industry (mapped at 5-digit NACE), but no information about the shipped commodity. The reason for not asking about commodity is that it reduces the reporting burden for the respondent, but still it is possible to obtain information about all shipments the company has had one year. At such detailed level the CPA describes the main activity of the firm. Conversions keys between the commodities in the freight model and the NST 2007 have been developed.

The 39 different commodities in the Norwegian model are presented in Table 2 (Hovi et al, 2015). All 39 commodities are used in the latest version of the model system. The six aggregated groups (see Table 2) are used in the presentation of results from the model. At the moment, TØI are working out new commodity flow matrices based on the CFS-2014 for Norway for domestic deliveries and the foreign trade statistics for import and export. More details are presented in chapter 3.

The goods flows for commodity 37, Recycling, cover the flows from the terminals to the plants, and not the collection of goods that are recycled from the households.

Table 2. Commodity classification in Norwegian model

	Commodity classification 2012/2013	Aggregated groups
1	Agricultural products	Dry bulk
2	Fruit, vegetables and flowers	Other thermo
3	General cargo, living animals	General cargo
4	Thermo input	Other thermo
5	Fresh fish	Fish
6	Frozen fish	Fish
7	Thermo consumption	Other thermo
8	Consumption food	General cargo
9	Beverages	General cargo
10	Animal foodstuff	Dry bulk
11	Organic inputs	Industrial goods
12	Other inputs	Industrial goods
13	Iron and steel	Industrial goods
14	Other metals	Industrial goods
15	Metal goods	Industrial goods
16	Chemical products	Liquid bulk
17	Plastic and rubber	Industrial goods
18	Timber and products from forestry industry	Timber
19	Wood products	General cargo
20	Pulp and chips	Industrial goods
21	Paper intermediates	Industrial goods
22	Paper products and printed matters	General cargo
23	Coal, ore and scrap	Dry bulk
24	Stone, sand, gravel and earth	Dry bulk
25	Minerals	Dry bulk
26	Machinery and tools	Industrial goods
27	Electronic equipment	Industrial goods
28	General cargo, building materials	General cargo
29	Cement, plaster and cretaceous	Dry bulk
30	General cargo, consumption	General cargo
31	General cargo, high value	General cargo
32	Vehicles	Industrial goods
33	Crude oil	Liquid bulk
34	Petroleum gas	Liquid bulk
35	Refined petroleum products	Liquid bulk
36	Bitumen	Liquid bulk
37	Waste and recycling	Dry bulk
38	Other fish (conserved)	Fish
39	Fertilizers	Dry bulk

Denmark

The Danish national freight model system (and the connected model for the Fehmarn Belt crossing) was developed by DTU and Significance (Significance 2012, 2014). It uses firms (in the setting of firm-to-firm flows in a logistics model, as for Samgods) with a distinction by NACE-sector and production and consumption in 23 commodity classes. The disaggregate mode choice model as it was estimated (Significance, 2012) only distinguishes (e.g. for the cost coefficients) between:

1. Food and agricultural products
2. Manufactured goods
3. Other non-dangerous products
4. Dangerous goods.

Finland

The freight model FRISBEE is primarily based on official transport statistics from Eurostat and the Finnish road transport survey. Information from sources describing infrastructure, traffic flows, goods flows, and the national accounts is also used (Transport Analysis 2011a). The model uses 13 commodity classes based on NACE/SITC: [9]

1. Food products and live animals
2. Beverages and tobacco
3. Raw materials
4. Coal, coke and briquettes
5. Animal and vegetable oils and fats
6. Chemicals and chemical products
7. Paper and paperboard and articles thereof
8. Metal and metal products
9. Manufactured goods
10. Machinery and equipment
11. Other manufactured goods
12. Office-, electrical- and telecommunications apparatus
13. Petroleum

The Netherlands

The Dutch national freight transport model system BasGoed was developed for the Ministry of Transport in a series of projects starting in 2009. In the estimation of the distribution and modal split sub-models (Significance et al., 2010), it used aggregate transport statistics for the year 2004. The commodity classification in this data set and these sub-models is NST/R at the 1-digit level with separate sub-models for:

1. Agricultural products
2. Food products
3. Solid mineral fuels
4. Petroleum
5. Ores
6. Metal and metal products
7. Raw minerals

[9] Björn Silfverberg, WSP Finland, April 26 2017.

8. Fertilisers
9. Chemical products
10. Other products.

The generation/attraction component of the BasGoed model was adopted from the trade model within the older SMILE+ model (Bovenkerk, 2015) that could distinguish more detailed commodities. However, since the distribution and modal split model distinguish between the ten categories above, BasGoed operates at this level and the transport demand outputs are also for the ten NST/R/1 categories (the network assignment does not use a commodity classification).

The Dutch Ministry of Transport plans to contract out the development of a new version of BasGoed in 2017. This development will include estimation on new data. The commodity classification in these new data is NST 2007. The new BasGoed model, which is supposed to be completed in 2018, will then use 20 NST 2007 classes or a subset of these.

Flanders

In Belgium, the strategic freight transport model for Flanders for the Flemish Government was redeveloped a few years ago, (version 4), using various data sources for the year 2010 (Borremans et al., 2015). Most of the data use the NST 2007 commodity classification (or classifications that can be easily linked to NST 2007), and so does the model itself:

1. Agricultural products
2. Coal, crude petroleum and gas
3. Ores
4. Food, beverages and tobacco
5. Textiles
6. Wood and wood products
7. Refined petroleum and cokes
8. Chemical products
9. Non-metal mineral products
10. Metals
11. Machines and equipment
12. Transport equipment
13. Furniture and other manufactured goods
14. Waste
15. (Mail and parcels)
16. (Material for transport of goods)
17. Removals
18. (Grouped goods)
19. Unknown
20. (Other)

Some of the data sources however were for NST/R and a conversion table (from the University of St. Gallen in Switzerland) was used to get NST 2007. This table is not a simple aggregation table, but has NST fractions for each NST/R category.

The data (aggregate transport statistics by mode) had no observations for the NST classes 15 (mail and parcels), 18 (grouped goods) and 20 (other). The NST 2007 classes 12 (transport equipment), 16 (material for transport of goods) and 19 (unknown) were merged into a single new class 12 because for the rail sector these distinctions could not be made. For NST 2007 classes 5 (textiles) and 17 (removals) there were not enough rail and inland waterway observations to estimate a modal split model, so it is assumed that for these commodities that road is the only possible mode. Containers can be used in all commodities 1-14 and 17, though the container share varies between these classes. Aggregate modal split (and vehicle type choice) models were estimated for NST 2007 classes 1-4, 6-14. The transport demand outputs of the version 4 model are also for these NST 2007 classes (plus 5 and 17, see above).

Germany

A disaggregate modal split model for the German Bundesverkehrswegeplan (BVWP) 2015 was developed for the Federal Ministry of Transport (BVU and TNS Infratest, 2014). The same study was carried out to provide freight values of transport time and reliability (the data used include a new SP survey among shippers). It uses a commodity classification of its own in 10 classes:

1. Maritime combined transport
2. Continental combined transport
3. Shipments of 100 tonnes and more
4. Agricultural and food products
5. Stone and earth
6. Crude petroleum products
7. Chemical products and fertilisers
8. Metals and metal products
9. Vehicle and machines
10. Other intermediate and final products

Transport statistics in NST 2007 are also used, and allocated to the commodity classes as above. This is done by means of the following conversion table.

Table 3. Commodity segmentation used in the Stated Preference survey and models versus NST2007

Segment no.	Segment Name	NST2007 number (subgroups)
1	Maritime combined transport	All container goods
2	Continental combined transport	All combined transport with rail as main mode
3	Shipments of 100 tonnes and more	21 (coal), 22 (lignite), 31 (ores), 71 (cokes)
4	Agricultural and food products	10 (agri), 40 (food products)
5	Stone and earth	33 (stone and earth), 90 (other minerals), 140 (recycling and waste products)
6	Crude petroleum products	23 (crude petroleum), 72 (oil products)
7	Chemical products and fertilisers	32 (fertilisers), 80 (chemical products)
8	Metals and metal products	100 (iron and steel)
9	Vehicle and machines	110 (machines), 120 (vehicles)
10	Other intermediate and final products	50 (textiles), 60 (wood and paper), 130 (furniture), 150-190 (other products)

United Kingdom

The EUNET model (Jin et al., 2005) was originally developed for the Trans Pennine region in North-England as a demonstration project for the European Commission's Fourth Framework Programme. Later, it was extended to two other regions in the UK and to the whole of Great Britain. It has been used to provide national base year matrices for freight transport (WSP, 2012). EUNET uses 31 product groups (sectors each producing one good or service) and aggregations of these (largely consistent with NST/R).

2.2 International level

Freight model for Oresund region

The freight model for analysing choice of mode and crossing in the Oresund region (Rich et al., 2009) uses 13 commodity groups that are based on 52 NST/R-groups at the two-digit-level.

1. Agriculture products (NST/R-groups 00,01,02,03,06)
2. Food and feed (NST/R-groups 11,12,13,14,16,17,18)
3. Wood, Cork, Textile fibres etc. (NST/R-groups 04,05,09)
4. Non-liquid fossils (NST/R-groups 21,22,23)
5. Oil products (dangerous freight) (NST/R-groups 31,32,33,34)
6. Chemical products (dangerous freight) (NST/R-groups 81,82,83,89)
7. Ore products (NST/R-groups 41,45,46)
8. Metallurgic products (NST/R-groups 51,52,53,54,55,56)
9. Paper mass (NST/R-group 84)

10. Stone, sand, concrete, and fertilisers (NST/R-groups 61,62,63,64,65,69,71,72)
11. Machines (NST/R-groups 91,92,93)
12. Manufactured goods (NST/R-groups 94,95,96,97)
13. General cargo (NST/R-group 99)

The final choice has been partly inspired by the Samgods model system (SAMPLAN, 2001). Compared to Samgods, raw wood and wood products are joined in the same group. Also, fertilisers have been moved from chemical products to stone, sand, and concrete. Thus, chemical products are dealt with separately. Machines and manufactured goods have been decomposed into machines, manufactured goods, and general cargo.

Europe

The new transport model for the European Commission, Transtools 3, includes a freight transport model (Fjendbo Jensen et al., 2016). This freight model includes a transport chain choice model that was estimated on disaggregate data (the Swedish CFS and the French ECHO). In application the model uses PC data for Europe for 2010 that was provided by the ETISplus project and that used NST/R. Therefore, Transtools 3 also uses the NST/R commodity classification, both for the trade forecasting model and the transport chain choice. The level used is NST/R/1 (10 groups, see above in the description of the Basgoed model for the Netherlands). In the transport chain model, separate models were estimated for:

- Solid bulk goods
- Liquid bulk goods
- General cargo and containerised goods.
- Within these models however, a distinction is made in the 10 NST/R/1 classes (commodity-specific dummy variables).

2.3 Implications for commodity classification in Samgods

Table 4 summarizes the commodity classifications applied in the different model systems. So far, the NST 2007 classification has only been used in the German and Flemish model systems. The Netherlands and Sweden have plans/consider to use the NST 2007 classification in the next version of their national freight transport model system.

The overall picture is that the number of commodities is lower when the NST 2007 classification is applied. In general, fewer commodities imply a lower level of detail when describing transport demand and cost functions. However, sensitivity analyses (Vierth, Karlsson and Westin, 2016) with twelve instead of 32 commodities in the Samgods-model system[10], which only allows consolidation within commodities, indicate that the accuracy of the model results regarding mode choice and choice of vessel size is improved when the model is based on fewer commodities. The merge of the commodities leads to 6,9% more sea tonne-km, 4,9% more road, 4,3% more rail tonne-km and total tonne-km increase by 3,8%. These figures are more in line with the observed values in Sweden.

[10] Air freight was not included.

Table 4. Commodity classifications in freight models

	Classification based on NST/R	Classification based on NST 2007	Classification based on NACE/SITC	Number of com-modities
Sweden	x (existing model)	x (planned model)		35 (32/33)
Norway			x	39
Denmark			x	23 (4)
Finland			x	13
Flanders	x	x		14
Germany		x		10
Netherlands	x (existing model)	x (planned model)		10
United Kingdom	x			31
Öresund	x			13
Europe (Transtools 3)	x			10

More models adopting the NST 2007 classification strengthen the case for using that categorization (or some form of it) in the Samgods model. A common classification would make it easier to compare results between freight models/countries and integrate the models.

In the Danish freight model, dangerous goods are used as a commodity group. We do not think that it is necessary to specify this commodity because NST 2007 groups 2 (crude petroleum), 7 (petroleum products) and 8 (chemicals) already cover most of the dangerous goods that are transported in Sweden. In addition, there is an element of ambiguity about which goods are considered dangerous. This depends on the volume of the product and which other commodities it is transported together with.

3 COMMODITIES IN DEMAND-MATRICES

This chapter will evaluate different alternatives for commodity classifications from the viewpoint of modelling transport demand in the Swedish and Norwegian national freight transport model system. It pays particular attention to the suitability of the NST 2007 classification. The aim is to answer the following research questions:

- What are the advantages and disadvantages of using NST 2007 from the viewpoint of the base (and forecast) matrix project?
- Do data sources cover all commodities in the proposed classification (NST 2007 or any aggregation)?
- Do the data sources for the base matrices allow translation to the proposed classification (NST 2007 or any aggregation)?
- Do estimated economic growth rates for industry aggregates allow translation to the proposed classification (NST 2007 or any aggregation)?
- Does the proposed classification (NST 2007 or any aggregation) increase or reduce uncertainties, does it lead to more or fewer calculation steps in the PWC matrix generation, and does it increase or reduce the need for input data?

3.1 Introduction

The Samgods model's demand matrices are called *PWC Matrices* since they describe Production, Wholesale[11] and Consumption per zone.[12] The matrices describe the estimated transport demand through the elements *(r, s)* which represent yearly goods flow from zone *r* to zone *s*. Zones within Sweden correspond to the 290 municipalities and abroad 174 larger regions are defined. Goods flows include domestic transport, Swedish import and export, and transit, i.e. transports between zones abroad that travel on Swedish infrastructure.

The PWC matrices are the most extensive data set in the input data needed to run the Samgods model. Since they are commodity-specific, the generation of PWC matrices is highly dependent on the choice of commodity classification. As described below, the generation of the matrices is based on statistics and forecasts of production, consumption and trade. These statistics and forecasts mainly follow the official industry sector classifications. Some industry sectors only produce and/or consume one type of commodity, but most industry sectors handle several commodity types, and most commodity types are being handled by several industry sectors. Thus, whichever commodity classification is chosen

[11] Wholesale is included in the matrices and estimated separately from production and consumption. However, it is not specified in the resulting matrices whether individual shipments are sent from a producer or a wholesale unit, or received by a consumer or wholesale unit. In other words, production and wholesale have been added together, as well as consumption and wholesale, in the final matrices. The reason is that the Swedish Transport Administration during the generation of the matrices decided that it this information was not necessary for the purposes of the Samgods model.

[12] Commodity- and year (base year and forecast)-specific matrices are derived according to methods developed during 2013-2016. These methods are briefly described in a section below.

for Samgods and the PWC matrices it has to be connected to the industry classifications in an exact way.

3.2 Background

Swedish freight model

The method development for the PWC matrix generation started from a blank page regarding the commodity classification (the options were later narrowed down to the only alternative to use the current classification, since the other parts of the model were not ready for a revised classification). While inventing available data sources for the matrix generation, the correspondences between existing industry and commodity classifications were described in a short PM (see Appendix 1). The main conclusion of the PM was that the NST 2007 classification could be connected to available data sources in a similar way to the NST/R classification[13]. Therefore, it was suggested that a new commodity classification, based on NST 2007, should be considered for Samgods (since the NST/R classification has been abandoned for official transport statistics).

Furthermore, classifications in available data sources were briefly described and how they could be connected to the Samgods commodity classification (assuming it was based on either NST/R or NST 2007). It was also concluded that the correspondence *between* NST/R and NST 2007 is not straightforward. The reason is that "NST 2007 is based on the production process where the goods are coming from, while NST/R is based on the physical characteristics of the goods".

The methods for generation of base year (2012) and forecast (2040) PWC matrices are described by WSP, Sweco & KTH (2015), WSP & Sweco (2016) and WSP (2015). These reports give detailed information on the estimations that are made based on available statistics and forecasts. Except for the report on the base year matrices (WSP, Sweco & KTH, 2015), which includes the same findings as the PM in Appendix 1, the commodity classification is not discussed but assumed to be given by the current Samgods model version. However, the method descriptions give insights to the implications of the choice of commodity classification, which are described in the next section.

Norwegian freight model

The PWC matrices in the Norwegian freight model represent yearly goods flows between zone pairs (Hovi et al. 2015). Zones within Norway correspond to the 430 municipalities, while the six largest cities in Norway are divided into five to twelve city zones. The continental shelf is represented with six zones. For imports and export within Europe, country is the main zone unit, while for overseas transport, continent is the main zone unit. Sweden, however, is represented with 13 zones in the Norwegian model and some of Norway's other main trade partners within Europe are represented with more than one zone per country. In total, the model encompasses 554 zones, 76 of which are foreign. In the

[13] The current Samgods commodity classification is based on the official NST/R commodity classification.

Norwegian model, goods flows include domestic transport, transport to and from the continental shelf, imports, exports, and transit, i.e. transports between foreign zones, for which Norwegian infrastructure is used.

The commodity classification in the Norwegian national freight model is adapted for the Norwegian manufacturing industry, and the classification is based on the different commodities' needs for transport quality. The classification also takes into account the SCGE-model's (Pingo) (Hovi et al. 2017) need for consistency with both the freight model and the national account statistics. The commodity classification is based on the NACE classification (SITC for foreign trade) and not on the NST 2007. The reason is that the main input data for the PWC matrices consist of the commodity flow survey (CFS) for Norway and the foreign trade statistics (Hovi et al., 2015), which are based on the NACE[14] and SITC nomenclatures. Since the CFS covers only domestic deliveries from manufacturing industries and wholesale trade (Statistics Norway 2012, 2016), the data are supplemented with data from the primary industry, mining and quarrying, basic data from port statistics and lorry statistics, and selected information from the business sector.

In earlier versions of the model, PWC matrices for the base year were developed mainly from economic statistics that derived margins for production and consumption and estimated delivery patterns based on gravity models. However, this framework resulted in large inconsistences regarding the number of tons loaded and unloaded at different terminals. Since data from the first commodity flow survey for Norway became available in 2011, the CFS has been the main data source for the commodity flow matrices, supplemented with the above-mentioned data sources. In the present version of the matrices, the inverse factors of the sample probability (the blow-up ratios) in the CFS from 2008 have been used as calibration factors, in order to achieve consistency between the freight model and the statistics regarding national transport performance and tons loaded and unloaded at ports and railway terminals. However, using blow-up ratios will lead to uncertainties at the detailed geographical level and for particular commodities. Therefore, the methodology was changed in the CFS for 2014. Instead of applying blow-up ratios to extract national volumes from the sample, Statistics Norway now relates the sample to the unit registry. Information about turnover and goods flows of firms in the survey is used to estimate relationships between the turnover and the quantity of goods. These relationships are further used to impute goods flows for firms in the unit registry that are not included in the sample. Information about delivery pattern is based on a "nearest neighbor" principle, meaning that firms for which goods flows are imputed are assumed to have the same delivery pattern as the nearest firm within the same three-digit NACE code.

[14] The CFS in Norway does not obtain information about the commodity that is transported, but the detailed industry of the delivering firm. The commodities in the freight model is therefore derived from the main commodity that the firm is offering (4-digit CPA). This applies to domestic deliveries.

Commodity flow matrices for imports and exports are based on the foreign trade statistics. Information about place of origin for export and destination for import, is based on enterprise numbers in the customs clearance statements (in the TVINN registry[15]), which, coupled to the enterprise registry, give information on the firm's location. Statistics Norway has attempted to identify the location of the origin (based on information about county for production) or destination of shipments, to avoid that the address of the corporate headquarters is used as the sender or receiver. Extensive tests and corrections are needed to identify whether the goods go from/to the correct (or most likely) location. In the latest data set, information is available at the shipment level, and includes the postal code of origin and destination. Because the specified postal code might be the location where the ownership of the shipment is transferred (which depends on the transport agreement, Incoterms, and might be domestic, in the country of origin/destination, or in a third country), this information might not be fully usable.

The method for generating base matrices for future years is described in Hovi et al. (2017). The SCGE model Pingo is used for regionalizing national growth rates for GDP and the split on different commodities between 89 domestic and 7 foreign regions, and divided into 19 different commodity-delivering industries. The growth rates are connected to the PWC matrices in the freight model to achieve base matrices for future years. In the present forecasts, base matrices are worked out for the following future years: 2022, 2030, 2040 and 2050.

Experiences from the commodity classification in the Norwegian model

The PWC matrices for Norway are divided into 39 different commodity groups, adapted for the Norwegian industrial structure and the different commodities needs for transport quality. An increased number of commodities increase the need for data quality, but also allows identification of missing commodity flows in the data. One challenge with the detailed commodity grouping is to maintain the confidentiality of the companies in the PWC matrices. To get access to basic data from Statistics Norway, a confidentiality agreement must be signed. The confidentiality agreement express that data must not be published if the tasks from a single firm can be identified. This is an increasing problem with increased number of zones and/or commodities in the model, and set requirements for the user of the model. Results are therefore never reported for a single commodity. If results are reported for selected commodities, an aggregation into seven groups are defined.

An advantage of running the freight model with more commodities is that this can reduce the problems caused by the all or nothing assignment, given that the unit costs between the additional commodities differ. For example, different vehicle types (thermo truck, tank truck or articulated trailer) can be related to different commodities or there can be differences in average load weight (due to different density, tons per m^3). These differences affect the utilization of the vehicle and the costs compared to other vehicles and modes.

[15] TollVesenets INformasjonssystem med Næringslivet

3.3 Analysis

In this section, the various steps and features of the matrix generation methods in Sweden are analyzed in order to answer the research questions stated above.

Correspondence tables

As part of the PWC matrix generation during 2013-2016, the correspondences between industry sectors and commodities described above were derived from data provided by Statistics Sweden (SCB). Two data sets were used:

1. The Industrial goods statistics (IVP); a yearly survey that has been conducted since 1996, with the purpose to describe the Swedish production of commodities at a detailed level.
2. The Intermediate consumption statistics (INFI) describe the industry's consumption of input materials, per industry sector and commodity type.

All data were classified according to the Samgods commodity classification (32 commodity types currently in use) as well as the NST 2007 classification (at the most detailed level with 81 available subgroups). Except from the commodity classifications, the data also included the industry sector code for the producing (1) and consuming (2) industry sectors respectively. All data concern monetary values.

The two data sets were used to derive correspondence tables for production and consumption respectively, i.e. to construct tables linking industry sectors to commodity types (Samgods or NST 2007).[16] The results showed that in the table connecting *production* per industry sector to commodities, values for 5-digit industry sectors had to be split up 44 % more often using the Samgods classification compared to the NST 2007 classification. The corresponding number for the *consumption* table was 25 % more often for the Samgods classification compared to NST 2007.[17]

This is an indication that any aggregation of the NST 2007 classification will result in correspondences that are one-to-one to a higher extent than the current Samgods classification. Since every split in the tables decrease reliability and precision in the matrix generation, this will contribute to higher quality of resulting matrices.

[16] These correspondence tables could be useful for other model purposes as well. However, the data sets used to derive them are confidential due to their high level of detail. Possibly, aggregated versions of the keys could be published and/or used in other projects, but this would require permission from the responsible actors (Statistics Sweden and the Swedish Transport Administration).

[17] When an industry sector code had links to two commodities, it was counted as one split. If there were three links, it was counted as two splits, etc. Since there are 32 available Samgods commodities and 81 available NST 2007 commodities, the number of splits were normalized by dividing it with then number of commodity classes for each classification respectively. The reason for that is that the NST 2007 classification is assumed to be aggregated in some way to be useful and an aggregation will most probably reduce the number of splits.

Matrix generation

Below, a brief description of the methods for matrix generation (base year and forecast) is given, together with discussions on how the methods will be affected by a revised commodity classification.

Base year matrices are generated by following these steps (WSP & Sweco, 2016). For each step, the effects of changing commodity classification are assessed.

1. Estimation of national totals of production, consumption, imports, exports and wholesale per commodity for the base year. Estimations are made in economic value based on detailed versions of official statistics.
    - **Effects**: The detailed statistics have already been delivered classified by Samgods as well as NST 2007 before. Since microdata are originally specified by the detailed classification Combined Nomenclature at an 8-digit level (CN8), it is sufficient that the chosen commodity classification can be aggregated from CN8, which is the case for both NST/R and NST 2007 (see Eurostat's [2017] Metadata Server RAMON for more information regarding this). Additional statistics that are used for these estimations use yet another, detailed commodity classification. Previously, these have been manually connected to the Samgods classifications and this can be done to any new classification again without much effort.

2. The total levels of production, consumption and wholesale are distributed over zones (municipalities) using primarily employment statistics per industry sectors and municipality. Import and export are specified per country in the data from step 1, but for some countries, values are split into regions according to the distribution in previous base matrices (2006).
    - **Effects**: The employment statistics used for the municipality distribution is given on detailed industry sector classification (5-digit SNI2007). Thus, the distribution procedure uses the industry sector codes given in the statistics in step 1, and is not dependent on the commodity classification. For the distribution on foreign zones (only for cases where countries are divided into two or more zones), fractions from previous matrices (2006) have been used; these are commodity-specific and thus need to be re-estimated using the new commodity classification.

3. Conversion of economic values to tons per commodity, using average commodity values (commodity specific SEK/ton values are estimated based on foreign trade statistics and commodity flow survey (CFS) data)
    - **Effects**: Commodity-specific SEK/ton values have to be derived according to the new commodity classification. The values derived from foreign trade statistics are based on the detailed CN8 classification described above, thus any classification that can be aggregated from CN8, such as NST 2007, can be used without

adding any calculation steps compared to the present model. Other commodity values are based on CFS data. In order to re-estimate these, CFS microdata has to be possible to aggregate to the new commodity classification. The CFS 2016 uses a commodity classification that is connected to NST 2007, but not at the most detailed level for all groups[18], see Table 12.

4. The resulting levels constitute row and columns sum constraints for the matrices. Matrix elements are predicted using models estimated on CFS data.
    - **Effects**: Prediction models have to be re-estimated for the new commodity classification. Since the models use CFS data as the base for estimation, (again) CFS data needs to be available grouped according to the new classification.

5. The predicted á priori matrices are adjusted to fit the Samgods models specifications in three regards: (i) transport demand in each relation is split into "firm-to-firm" demand, (ii) large observed flows, primarily railway flows, (if not estimated properly) are added to the matrices and compensated by lowering other levels for the same commodity, (iii) transit flows are added.
    - **Effects**: Firm-to-firm splits (i) are done using industry sector data and a correspondence table (see section above). Thus, a new table is needed, but it will be possible to derive from the previous steps. If adjustment to observed flows (ii) is to be made in future matrix generation, these data need to be transformed to the new classification manually. Since these flows are relatively few, this should not cause any problems. The same holds for transit flows (iii).

For the forecast matrices, the following steps are added:

6. Results from step 3 above, that have been aggregated to the national level, are multiplied by growth rates from official economic forecasts, specified at an aggregated industry sector level. Levels are also divided by forecasted growth rates for commodity values. This results in national totals in ton per commodity, for the forecast year
    - **Effects**: Since economic growth rates are specified at an aggregated industry sector level, this step uses the industry sector classification in the data and therefore the commodity classification does not matter. Growth rates for commodity values are forecasted by commodity type, but if these are available for the base year (together with time series data from the same source); they can be forecasted using the present model.

[18] Fredrik Söderbaum, Transport Analysis, April 6, 2017.

7. Distribution of results from step 6 (production, consumption and wholesale) to municipalities using employment forecasts per industry sectors and municipality.
    - **Effects**: None, since the step is not dependent on commodity classification (as for step 2).

8. Distribution of results from step 6 (imports and exports) to zones abroad using foreign trade forecasts per industry sector aggregates. For zones that are smaller than one country, country-specific estimates are split up according to the fractions from the base year.
    - **Effects**: Same procedure as for step 6, the conversion to commodities is made using data from the base year, which means that the commodity classification does not add any steps to the method for the forecast. For split-up to zones smaller than one country, see discussion for step 2.

9. The same adjustments as in step 5 are made, but based on forecasted data instead of observed data.
    - **Effects**: Base year data is forecasted using results from steps 6-8, industry sector classified data/forecasts and forecasted GDP growth for the respective countries, which are not dependent on commodity classification.

Coverage of all commodities in classification

The current Samgods commodity classification does not include commodities such as

- Household, municipal and other waste or secondary raw materials (except when they are produced and sold by the considered industry sectors[19] and thus included in the statistics), NST 2007 commodity 14
- Mails and parcels, NST 2007 commodity 15
- Equipment and material utilized in the transport of goods (containers, pallets, etc.), NST 2007 commodity 16
- Goods moved in the course of household and office removals; baggage and articles accompanying travelers; motor vehicles being moved for repair; other non-market goods, NST 2007 commodity 17

The reason is that they are either not included in the NST/R classification, or they are produced (or generated) outside the considered industry sectors. Therefore, they are not regarded in the PWC matrix generation method. However, these four groups of goods correspond to classes 14-17 in the NST 2007 classification. If they are to be included in the new commodity classification, the matrix generation method needs to be extended to additional data sources and estimations.

[19] Only commodities produced by the agriculture, forestry, mining and quarrying or manufacturing sectors are included in the matrices

Since some of these commodity types are not always traded the same way as other commodities, some model development probably has to take place to be able to estimate e.g. recycling flows in the matrices. Currently, recycling activities and e.g. heating plants (that use waste as an input) are not included in the considered industry sectors, which they will have to be in case these products are included in the classification.

Split of commodity groups in classification

The current Samgods commodity classification distinguish between commodities with different trade patterns. For instance, products of agriculture and hunting and forestry products are traded in more or less separate systems. The product groups originate from different industries, namely the agricultural sector and the forestry sector. These industries are geographically connected to certain land-uses and the geographical distribution of the agricultural sector differs from that of the forestry sector. In the other end of the trade relation, the consumption of the respective sub-groups mainly takes place in different industries as well, namely the paper/wood industry and the foodstuff industry. Estimating the two subgroups in aggregation thus could result in inaccurate P-C relations geographically; e.g. a production unit in a strictly agricultural area could be incorrectly connected to a paper industry plant. This is a reason for keeping these sub-groups separated also in the next classification in Samgods.

3.4 Implications for commodity classification in Samgods

The findings from the preceding section can be summarized as:

1) The new commodity classification should be possible to aggregate directly from the detailed CN8 classification (which is the case for e.g. NST 2007 at any detail level).
2) It should be possible to classify CFS data according to the new commodity classification (which is the case for e.g. NST 2007 at a specific detail level; it is recommended to coordinate the choice of new classification directly with the producers of the CFS).
3) Any detail level of NST 2007 will probably result in sparser correspondence tables connecting industry sectors to commodity types than the current Samgods commodity classification, which will add accuracy to the forecast matrices.

If the two first requirements are fulfilled, the method for generation of matrices will be functional without adding any calculation steps to existing models (but some sub-models need to be re-estimated according the new classification). However, if the proposed commodity classification will include products that are not included in the current classification, such as waste and recycled raw materials, the method has to be expanded to include data sources and calculation steps accounting for the new commodity types. The answers to the questions posed in the beginning of this chapter thus are:

- **What are the advantages and disadvantages of using NST 2007 from the viewpoint of the base (and forecast) matrix project**:
    - No disadvantages have been found, except that some models and distributions need to be re-estimated (which will be the case for all alternatives as long as the current classification is not kept) and that any new products need to be covered by new data sources and estimated with corresponding methods. An advantage is that NST 2007 is closer related to the industry sector classification than NST/R, which gives matrices that are more accurate for the forecast year.

- **Do data sources cover all commodities in the proposed classification (NST 2007 or any aggregation)?**
    - Since the NST 2007 classification covers more commodity types (such as waste, mail and parcels) than the NST/R classification, new data sources need to be added to the matrix generation method if these commodities are to be included in the new classification. No analysis of the availability of this kind of data has been conducted.

- **Do data sources for base matrices allow translation to the proposed classification (NST 2007 or any aggregation)?**
    - Yes, at least as well as the current Samgods commodity classification.

- **Do estimated economic growth rates for industry aggregates allow translation to the proposed classification (NST 2007 or any aggregation)?**
    - Yes, at least as well as the current Samgods commodity classification.

- **Does the proposed classification (NST 2007 or any aggregation) increase or reduce uncertainties, does it lead to more or fewer calculation steps in the PWC matrix generation, and does it increase or reduce the need for input data?**
    - It probably reduces uncertainties to some extent. If new commodity types are added to the Samgods model such as waste or mail the required amount of input data will increase and calculation steps will probably be added to models. Other than that, the number of calculations steps and input data requirements will remain unchanged.

4 COMMODITY CLASSIFICATION IN LOGISTICS MODEL

4.1 Introduction

There is growing evidence that firms differ in their sensitivity to transport cost when they choose transport mode and shipment size. The cost sensitivity differs both depending on observed attributes such as shipment characteristics (de Jong et al. 2010; Johnson and de Jong, 2011) and unobserved attributes like shippers' attitudes (Arunotayanun and Polak, 2011).

Differences in cost sensitivity, i.e. taste heterogeneity[20], is typically handled in freight transport models by dividing the freight market according to the commodity of the shipment (although other firm and shipment variables are sometimes also used). In Samgods, both the logistics costs and availability of transport modes are set to vary by commodity.

The development of a new commodity classification in Samgods therefore calls for an analysis of how well this classification captures the tastes of decision-makers in the Swedish freight market. The new commodity classification in Samgods should ideally result in segments of shippers who have very similar tastes, as this would strengthen the behavioural foundation of the model.[21]

The purpose of this chapter is to analyse observed and unobserved differences in tastes of shippers in the Swedish freight market. More specifically, we will first investigate the extent to which shippers differ in how much transport cost influence their choice of transport chain and shipment size. We will then analyse how well the NST 2007 classification captures differences in cost sensitivity among shippers. We will also provide a short description of the distribution of shipment characteristics for different commodity groups.

4.2 Short Review of Empirical Evidence of Taste Heterogeneity

Empirical evidence of taste heterogeneity in the Swedish freight market is documented in Johnson and de Jong (2011) who use the CFS 2001 to analyse differences in the influences of transport cost and time. They use controls[22] for firm size (in terms of the number of employees), value density of the shipment, whether the consignment commodity is a metal product or chemical product as well as alternative-specific constants that capture the difference in the shippers' utility of the modal alternatives that is not accounted for by the other variables. In all their specifications, the results are essentially the same; there is a difference

[20] Taste heterogeneity generally refers to the differences across individuals in their preferences (Greene and Hensher, 2007). In the context of decision-making in the freight market, heterogeneity essentially means that the value or importance that freight agents put on a parameter (such as cost and time) varies between them.

[21] A commodity classification in Samgods that result in groups of shippers with homogenous cost sensitivity would also make the development of a stochastic logistics module easier. It would mean higher precision for the generic cost coefficients of each commodity group.

[22] The "controls" or "control variables" are included to account for differences in taste heterogeneity among shippers that can be explained by these variables. Any remaining differences in sensitivity to cost or time found in the model cannot be attributed to these variables.

in the influence of transport cost across firms but no statistically significant difference in the influence of transport time other than among shippers using air transports.

Abate et al. (2014) use the Swedish CFS from 2004/05 to investigate taste heterogeneity in the choice of transport chain and shipment size. The authors focus on shipments of metal products and include controls for transport time, access to rail and value density as well as alternative-specific constants. The evidence of taste heterogeneity is mixed: the authors document heterogeneity in the influence of cost when considering domestic and international shipments separately but not when the full sample is analysed.

de Jong et al (2004) present the results from an interview study where respondents from Swedish and Norwegian firms are segmented by commodity group. The results show variation in the respondents' preferences for transport chains, time, cost and reliability both between and within commodity groups.

Abate and de Jong (2014) analyse the choice of vehicles using a Danish heavy trucks trip diary for 2006 and 2007. They investigate taste heterogeneity in the influence of operating cost and a cargo-vehicle-fit variable (measured as the difference between a vehicle's weight capacity and the weight of the cargo). Using controls for the age and the weight class of the vehicles, they find no evidence of differences in the influence of operating cost among firms. They do document taste heterogeneity when it comes to the weight put on the cargo-vehicle-fit.

Arunotayanun and Polak (2011) use a dataset based on a survey of shippers in Indonesia in 1998/1999. They identify behaviourally homogenous segments based on the value of the shipment, whether the shipment is containerized or not and the frequency of delivery. They show that commodity type alone is not a strong explanatory variable of the underlying heterogeneity.

The overall results from the literature indicate the presence of taste heterogeneity in the Swedish freight market and elsewhere. But it is less clear how the tastes of freight agents vary with commodity groups. This topic will be investigated in the following sections.

4.3 Data and sample selection for the analysis based on CFS 2009

The key data source for our analysis is the Commodity Flow Survey (CFS) 2009. The data set contains records of about 3,5 million individual shipments to or from a company in Sweden, with detailed information on shipment and company characteristics. The shipments are described by their origin, destination, value, weight, cargo type, commodity class, and the mode(s) used in the transport chains. The survey distinguishes between road, rail, maritime and air transport and includes an additional category for unknown modes. Company characteristics include employment size, sector affiliation and geographical location.

The second source we use contains estimations of transport cost and time for all the shipments in the CFS 2009. For each shipment, the data set contains

information on cost and time for combinations of 16 shipment weight classes and transport chains. In other words, the values are generated both for the chosen and non-chosen weight and mode combinations.

The estimations were carried out in the logistics module of the Samgods model. The Samgods network was used to calculate level-of-service (LOS)- matrices, which provide vehicle transport time, distance and network-related infrastructure cost between origin and destination by mode. This information was combined with vehicle cost parameters from Trafikverket (2017) to calculate transport costs, which consist of distance-based linked cost, time-based link cost (including the time cost of the vehicle but excluding the capital cost of goods during transit), cost for loading at the sender and unloading at the receiver and transhipment cost (see de Jong and Baak [2015] for additional information on the cost functions).

We merge the cost and time estimates with their corresponding shipments in the CFS. Matching the shipments with the choice alternatives produces a large number of records. To reduce the computational burden of our models we draw a random sample leaving us with 152,958 records (from 2,705 unique shipments). Matching the observed choice for a shipment in the CFS to an alternative in the data set derived from Samgods is non-trivial because there is not a one-to-one matching of transport chains between the data sets. The chains in the CFS are specified in terms of combinations of the main transport modes, while the Samgods model uses a more fine-grained definition of the chains. This means that a single observed choice for a shipment in the CFS can be linked to several chains in the Samgods data set (for instance, an observed choice of truck in the CFS may correspond to both light and heavy lorry).

To handle this, we aggregate chains in both data sets into four groups; road, rail, waterborne and air. Following Johnson and de Jong (2011), all chains including an air transport mode are classified as air transport, chains including water transport modes (but not air transport) are labelled as water transport and chains including rail transport (but not air or water transport) are categorized as rail. Chains including only road transports are coded as road.[23] While this makes it easier to link transport chains between the data sets, it produces duplicate alternatives for the shipments. Since we cannot allow duplicate alternatives in our models we randomly delete duplicate alternatives from the data. The final choice set consist of the four transport modes in combinations with the 16 groups of shipment sizes. Table 5 and 6 below show summary statistics and combinations of modes and shipment sizes for the sample.

[23] Abate et al. (2016) aggregates into 14 transport chain categories. This improves the matching of chains between data sets, but the larger choice set increases the computational burden of the models.

Table 5: Summary statistics

	Mean	SD
Transport chain variables		
Cost (10 000s SEK)	148	3,148
Time (100s hours)	0.078	0.127
Shipment variables		
Value (10 000's SEK)	3.228	2,966
Weight (10's KG)	1,895	7,352
Value density (value/weight)	0.463	5.914
International	0.119	0.323
Truck load (<40 tonnes)	0.631	0.483
Liquid bulk	0.054	0.227
Solid Bulk	0.008	0.088
Containers (>20ft)	0.004	0.064
Containers (<20ft)	0.001	0.033
Palletized goods	0.291	0.454
Self-propelled units	0.061	0.240
Other mobile units	0.001	0.033
Unknown unit	0.005	0.072
Other cargo units	0.574	0.495

Note: All variables except the transport chain variables, shipment value, weight and density are dummy variables. The mean of each variable shows the share in the respective population for which the variable takes the value 1.

Table 6: Combination of mode and shipment size choices

	Road	Rail	Water	Air	Total
0-0,05t	708	11	125	16	860
0,05-0,2t	121	6	4	1	132
0,2-0,8t	120	12	6	2	140
0,8-3t	129	7	11	0	147
3-7,5t	112	1	4	0	117
7,5-12,5t	158	2	1	0	161
12,5-20t	178	3	5	0	186
20-30t	198	0	3	0	201
30-35t	186	0	6	0	192
35-40t	274	0	4	0	278
40-45t	161	0	3	0	164
45-100t	94	0	9	0	103
100-200t	10	0	0	0	10
200-400t	6	0	0	0	6
400-800t	6	0	0	0	6
+800t	1	0	1	0	2
Total	2462	42	182	19	2705
Observations	2705				

4.4 Econometric specification

In all our specifications, we start from a joint model of discrete mode and discrete shipment size choice. Our baseline model is the multinomial logit model (MNL), in which the utility that decision-maker (i.e. shipper) i derives from choosing a combination of transport mode and shipment size, denoted j, is given by:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \qquad \text{i = 1,2, ..., N} \qquad \text{j= 1,2, ..., J}$$

$$V_{ij} = \beta Cost_{ij} + \delta\, Time_{ij} + \sum_{m=2}^{4} \alpha_m d_m \qquad (1)$$

where d_m is a dummy variable for transport mode m. The coefficients β and δ give the utility weights of cost and time respectively. The size of the coefficients is not informative of their impact on the probability of an alternative being chosen, but their signs are. A negative cost coefficient means that an increase in the cost of one alternative makes it less likely that the alternative is chosen and more likely that the other alternatives are chosen (Cameron and Trivedi, 2010). The baseline model serves as a benchmark that we can compare the other models to. To investigate heterogeneity in the reaction of shippers to cost we interact the transport costs with a set of variables describing the commodity group:

$$V_{ij} = \beta_1 Cost_{ij} + \sum_{c=2}^{C} \beta_c Cost_{ij} * d_c + \delta\, Time_{ij} + \sum_{m=2}^{4} \alpha_m d_m + \boldsymbol{\lambda}'_m \boldsymbol{X}_i \qquad (2)$$

where d_c is a set of dummy variables for each available NST 2007 group and X is a vector of controls. The coefficient β_1 gives the utility weight associated with transport cost for the reference commodity group. The coefficients β_c give the additional utility weight associated with commodity group c relative to the reference category. In other words, the utility weight of transport cost for group c is given by $\beta_1 + \beta_c$.

We run this model with and without controlling for shipment value. We expect that the reason why cost sensitivity may vary depending on commodity type is partly due to differences in the value of the consignment. Evidence of heterogeneity in cost sensitivity among shippers when controlling for shipment value would indicate that this factor alone does not explain behavioural differences.

The MNL model can only capture heterogeneity within some limits and tastes that vary with unobserved variables or simply randomly cannot be accounted for in the model. We therefore extend our analysis by estimating a latent class model, where the probability of shipper i choosing alternative j is given by:

$$P_{ij} = \sum_{m=1}^{M} S_{im} \frac{\exp\left(b'_m I_{ij}\right)}{\sum_{l=1}^{J} \exp(b'_m I_{il})}$$

$$S_{im} = \frac{\exp\left(\gamma'_m Z_i\right)}{\sum_{m=1}^{M} \exp(\gamma'_m Z_i)}$$

where the variables in I_{ij} are level-of-service variables (cost, time, indicators for transport mode and truck load). The coefficients b_m are fixed within each of the m classes/segments but can differ across segments, thus incorporating taste heterogeneity. The variables in Z are segment variables which are used to predict the class membership of each observation. We include firm variables (employment size) and shipment variables (cargo type, value, international origin or destination) in this vector. The results from the latent class model are subsequently used to calculate cost coefficients for individual shippers in the sample:

$$b_i = \sum_{m=1}^{M} b_m G_{im}$$

where b_m is the segment-specific cost coefficient and G_{im} is the posterior class membership probability of shipper i that follows from the latent class model (Train, 2003, ch. 11).[24] We plot the distribution of the cost coefficients for the different NST 2007 classes to investigate how cost sensitivity varies between and within groups.

4.5 Results

This section starts with a presentation of descriptive results from the CFS 2009, where we relate NST2007 groups to shipment characteristics. In the subsequent section, we present the estimation results from the MNL and latent class models.

Descriptive results

Tables 7-9 show the distribution of shipment value, weight and value density (value/weight) by NST2007 group. The first column shows the NST2007 class, the second gives the average value in each group, columns 3-5 provides the value at the 25th, 50th (median) and 75th percentile, and the last column shows the number of observations in each commodity group. All figures are based on shipments in the CFS 2009 which have been assigned to a NST2007 group based on their CFS commodity classification and a conversion key provided by Transport Analysis (2011b).[25] The key can also be found in Appendix 3.

NST2007 commodities 1-14 and 20 are represented in the survey. We focus our analysis on a selection of commodities that will have an impact on our recommendation in chapter 6.

[24] The posterior class membership probability depends both on the probability of a shipper belonging to a specific class and the probability of a shipper choosing a particular transport chain and shipment size solution. In other words, it is calculated based on the observed choice and choice situation faced by each shipper.

[25] Unless specified, we use the shipments for which the variables are specified by the respondents (rather than being imputed by Transport Analysis). Less than one percent of the shipments have their weight imputed while almost half of all shipments have their value imputed. This problem is almost exclusively limited to shipments of round wood and liquid refined petroleum products.

Table 7. Shipment weight (kg) by NST2007 class

NST 2007	Mean	P25	P50	P75	Obs.
1	29,835	11,900	28,660	37,900	1,773,748
2	3.1e+08	4.9e+07	8.5e+07	2.1e+08	58
3	186,146	6,000	23,000	23,000	5,661
4	5,179	57	263	1,163	190,405
5	8.1	.71	1.4	2.8	522,546
6	8,478	4	24	2,503	78,361
7	201,903	2,519	9,804	53,289	121,644
8	1,353	1.3	7.3	40	144,736
9	12,972	36	208	12,000	20,413
10	4,946	11	103	1,682	73,828
11	137	.67	2.1	10	232,734
12	2,701	15	907	1,480	70,537
13	132	.83	1.4	3.1	181,155
14	42,209	24,820	30,000	33,100	263
20	257	1.5	7.7	36	167,046

Table 8. Shipment value (SEK) by NST2007 class

	Mean	P25	P50	P75	Obs.
1	56,069	3,833	5,598	13,250	236,205
2	2.8e+08	1.1e+08	2.7e+08	4.1e+08	42
3	127,325	1,760	1,840	2,450	5,658
4	68,759	1,421	5,062	18,951	189,364
5	1,718	319	577	1,171	520,558
6	49,811	231	1,555	26,728	77,876
7	1,343,609	15,284	90,000	573,075	22,907
8	46,776	839	3,761	16,954	140,024
9	39,203	1,470	5,980	30,723	21,418
10	64,511	960	5,376	32,882	72,583
11	20,377	577	1,736	5,084	232,923
12	210,305	2,720	57,114	118,844	70,163
13	3,619	110	297	567	178,647
14	132,409	17,944	38,850	112,632	250
20	19,119	219	891	4,169	166,361

Table 9. Shipment value density (value/weight) by NST2007 class

	Mean	P25	P50	P75	Obs.
1	13	.32	.32	2.8	235,947
2	33	3.4	3.4	3.4	42
3	15	.08	.08	1.7	5,648
4	234	7.7	14	40	185,983
5	626	237	498	991	520,455
6	904	8.2	20	123	77,358
7	9.7	4	4.6	5.2	22,870
8	6,680	63	418	2,472	138,416
9	251	2.6	28	70	20,014
10	1,777	14	44	141	71,931
11	9,649	220	700	2,000	228,532
12	219	70	94	170	69,455
13	248	36	141	298	176,059
14	37	.97	1.4	2.5	236
20	1,553	23	112	384	163,527

Note: P25 shows the value of the shipment in the 25th percentile, P50 the 50th (median) and P75 the 75th. Source: CFS 2009.

Table 10. Descriptive statistics for sub-groups in NST2007 group 1.

	Mean	P25	P50	P75	Obs.
1.1. Cereals					
Value (SEK)	144,883	24,568	55,774	136,341	26,960
Weight (kg)	99,285	18,750	38,240	95,136	26,960
Value density	5.6	1.1	1.3	1.6	26,960
1.2. Potatoes					
Value (SEK)	35,067	4,800	10,285	33,522	176
Weight (kg)	8,045	601	1,363	6,240	176
Value density	58	5.7	8	8.7	176
1.3. Round wood					
Value (SEK)	210,531	27,500	80,471	112,749	228
Weight (kg)	167,769	21,606	31,833	37,000	228
Value density	20	.76	3.2	3.8	228
Value (SEK)	*10,491*	*5,613*	*9,496*	*11,877*	*1,396,214*
Weight (kg)	*32,183*	*18,650*	*33,060*	*38,910*	*1,396,214*
Value density	*.33*	*.27*	*.28*	*.33*	*1,396,214*
1.4. Raw milk					
Value (SEK)	170,745	5,464	58,084	232,789	44,848
Weight (kg)	59,815	1,775	19,391	80,020	44,848
Value density	3	2.9	3.1	3.1	44,848
1.5. Live animals					
Value (SEK)	4,603	1,042	2,590	4,992	1,224
Weight (kg)	14	1	2	6	1,224
Value density	1,143	416	808	1,394	1,224
Value (SEK)	*25,004*	*4,041*	*11,345*	*27,572*	*142,875*
Weight (kg)	*1,445*	*223*	*536*	*1,378*	*142,875*
Value density	*28*	*15*	*23*	*23*	*142,875*
1.7. Fresh fish					
Value (SEK)	56,254	3,604	10,813	37,000	734
Weight (kg)	1,819	42	150	627	734
Value density	301	41	65	95	734
1.8. Other forestry products					
Value (SEK)	77,262	19,980	73,671	100,901	198
Weight (kg)	18,909	2,177	24,238	31,910	198
Value density	16	3	4.7	8.1	198
1.9. Other substances of vegetable or animal origin					
Value (SEK)	9,660	3,686	5,152	6,861	161,579
Weight (kg)	14,919	11,103	15,510	19,520	161,579
Value density	6.6	.32	.32	.32	161,579
Total group 1					
Value (SEK)	56,118	3,834	5,598	13,251	235,947
Weight (kg)	33,120	10,747	15,985	22,662	235,947
Value density	13	.32	.32	2.8	235,947
Value (SEK)	*17,713*	*5,074*	*9,225*	*12,279*	*1,774,016*
Weight (kg)	*29,831*	*11,893*	*28,660*	*37,900*	*1,774,016*
Value density	*3.4*	*.28*	*.31*	*.45*	*1,774,016*

Note: Summary statistics for NST2007 group 1. Figures in italics are based on both reported values and imputed values. P25 shows the value of the shipment in the 25th percentile, P50 the 50th (median) and P75 the 75th. Source: CFS 2009

Commodity group 11 and 12 exhibit different shipment characteristics. Group 11 (machinery and equipment not elsewhere classified) have much lower shipment weight and value, both on average and across the distribution of shipments, compared to group 12 (transport equipment). There is also a difference in value density between the groups.

In Table 10 above, we look closer at the sub-groups of NST2007 group 1 (products of agriculture, hunting, and forestry; fish and other fishing products). Because the figures for a large share of observations in this group are based on imputations, we show summary statistics based on both reported and imputed figures (in italics) and on reported figures only. In terms of average figures, most sub-groups are not very different from the group means (bottom row).

In Table 11, we look closer at NST 2007 group 6 (Wood and products of wood and cork (except furniture); articles of straw and plaiting materials; pulp, paper and paper products; printed matter and recorded media). Average shipment value is similar for paper, paper products and products of wood and cork. Shipments of pulp tend to have higher value while printed and recorded media have the lowest shipment value on average. Because the average weight of the shipment follows the same pattern, value density is not very different between groups 6.1-6.3. The values between the 25th and the 75th percentile (i.e. half of all the observations) fall between 3 and 29. Printed and recorded media on the other hand, stands out as having a considerably higher value density.

Table 11. Descriptive statistics for sub-groups in NST2007 group 6.

	Mean	P25	P50	P75	Obs.
6.1 Products of wood and cork (except furniture)					
Value (SEK)	81,073	9,876	50,000	94,103	16,883
Weight (kg)	19,016	687	11,355	26,600	16,883
Value density	80	3	4.4	16	16,883
6.2 Paper, paper products					
Value (SEK)	53,067	74	577	12,996	39,337
Weight (kg)	6,591	5	21	800	39,337
Value density	157	8.4	19	29	39,337
6.3 Pulp					
Value (SEK)	376,418	120,000	150,249	252,968	866
Weight (kg)	89,920	28,000	36,000	57,100	866
Value density	65	3.7	4.3	4.8	866
6.4 Printed and recorded media					
Value (SEK)	4,631	294	747	2,335	20,736
Weight (kg)	150	1.5	4	13	20,736
Value density	3,010	135	181	216	20,736
Total group 6					
Value (SEK)	49,924	230	1,530	26,912	77,358
Weight (kg)	8,506	4	24	2,498	77,358
Value density	904	8.2	20	123	77,358

Note: Summary statistics for NST2007 group 6. P25 shows the value of the shipment in the 25th percentile, P50 the 50th (median) and P75 the 75th. Source: CFS 2009.

In summary, the descriptive results presented show that in terms of average shipment value, weight and value density, there are noticeable difference between groups 11 and 12 but no large differences within groups 1 and 6.

Model estimation results

This section presents the result from the multinomial logit model and the latent class model. Table 12 shows the results from the multinomial logit model. It includes the results from the baseline model in equation (1) and two additional columns that show the results from equation (2) (with and without controls for shipment value). The cost variable is interacted with indicators for 13 out of the 20 NST 2007 commodity classes.

In column 2, the cost coefficients measure for each commodity type the additional weight that shippers assign to cost, relative to the reference category (arbitrarily set to NST group 1, agriculture and forestry products). Almost all cost coefficients are statistically significant from zero, which indicates that cost sensitivity indeed differ depending on the commodity group. To assess whether the differences can be explained by shipment value, we include this variable in the model. This ensures that differences in cost sensitivity due to different shipment value are accounted for. The results are shown in column 3. Although this changes the size and sign of some of the coefficients there is still evidence of taste heterogeneity.

Table 12: MNL model estimation results

	(1)		(2)		(3)	
Cost	0.0000092	(0.000017)	-0.35***	(0.026)	-0.41***	(0.030)
Cost, NST 3			-0.27	(0.91)	-0.23	(0.94)
Cost, NST 4			-1.44***	(0.26)	-2.29***	(0.33)
Cost, NST 5			0.35***	(0.026)	0.41***	(0.030)
Cost, NST 6			0.0015	(0.14)	-0.051	(0.17)
Cost, NST 7			0.19**	(0.060)	-0.11	(0.11)
Cost, NST 8			-6.73***	(1.43)	-8.49***	(1.66)
Cost, NST 9			-0.44	(0.37)	-1.08*	(0.50)
Cost, NST 10			-1.17**	(0.38)	-1.81***	(0.46)
Cost, NST 11			0.35***	(0.026)	0.41***	(0.030)
Cost, NST 12			-0.76*	(0.31)	-3.33***	(0.68)
Cost, NST 13			-14.3***	(2.46)	-14.9***	(2.51)
Cost, NST 20			-6.58***	(0.94)	-7.09***	(0.98)
Time	3.10***	(0.25)	2.57***	(0.24)	2.44***	(0.24)
Road Constant	4.74***	(0.23)	3.41***	(0.23)	2.48***	(0.29)
Rail Constant	0.67*	(0.28)	-0.43	(0.28)	-0.86*	(0.34)
Water Constant	1.82***	(0.25)	0.83***	(0.25)	0.12	(0.30)
Controls for shipment value	No		No		Yes	
Observations	152,958		152,958		152,958	
Log-likelihood	-8381.1		-7576.4		-7386.8	

Note: The cost variable is scaled by 10,000 and the time variable is scaled by 100. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. No available data for NST group 2 and 14-19 in the sample.

Some additional results from Table 12 are worth noting. The coefficients on the alternative-specific constants sea and road are positive and statistically significant. This means that shipper prefer these options over the outside alternative (air), after correcting for the influences of the other included variables. The coefficient on transport time is positive, which goes against prior expectations. One explanation is that because the cost variable already includes time-based cost, the time variable picks up other effects of longer transport time. For instance, alternatives with longer transport time may also exhibit desirable characteristics, such as higher reliability and transport service.

Table 13. Estimation results from the latent class model

	Segment 1	Segment 2	Segment 3	Segment 4	Segment
			Level-of-service variables		
Cost	-0.000036	0.00000072	-0.0000042	-2.00	-0.064
Time	9.37	10.1	0.052	-5.39	13.0
Truck Load	20.9	14.9	19.6	3.45	2.22
Road Constant	0.14	2.22	3.87	3.26	4.29
Rail Constant	-4.26	-1.85	0.65	1.50	-2.91
Water Constant	2.01	-2.68	-1.30	-2.24	-6.21
			Segment variables		
Solid Bulk	0.18	-1.51	-2.66	0.93	
Containers (>20ft)	0.79	-2.08	-2.11	-1.84	
Containers (<20ft)	-10.5	-1.58	0.076	-0.093	
Palletized goods	0.26	-0.039	0.54	0.91	
Self-propelled units	2.57	0.52	1.67	1.93	
Other mobile units	7.00	1.27	4.39	5.39	
Unknown unit	-0.40	-1.99	0.29	0.22	
No. of employees	0.00092	0.0010	0.0012	0.0013	
Shipment Value	-0.0029	-0.023	-0.16	-0.026	
International	5.14	1.42	-1.14	-1.39	
Constant	-4.05	-1.47	0.97	-0.41	
Observations					152,958
Log-likelihood					-2962.6
AIC					5925.3

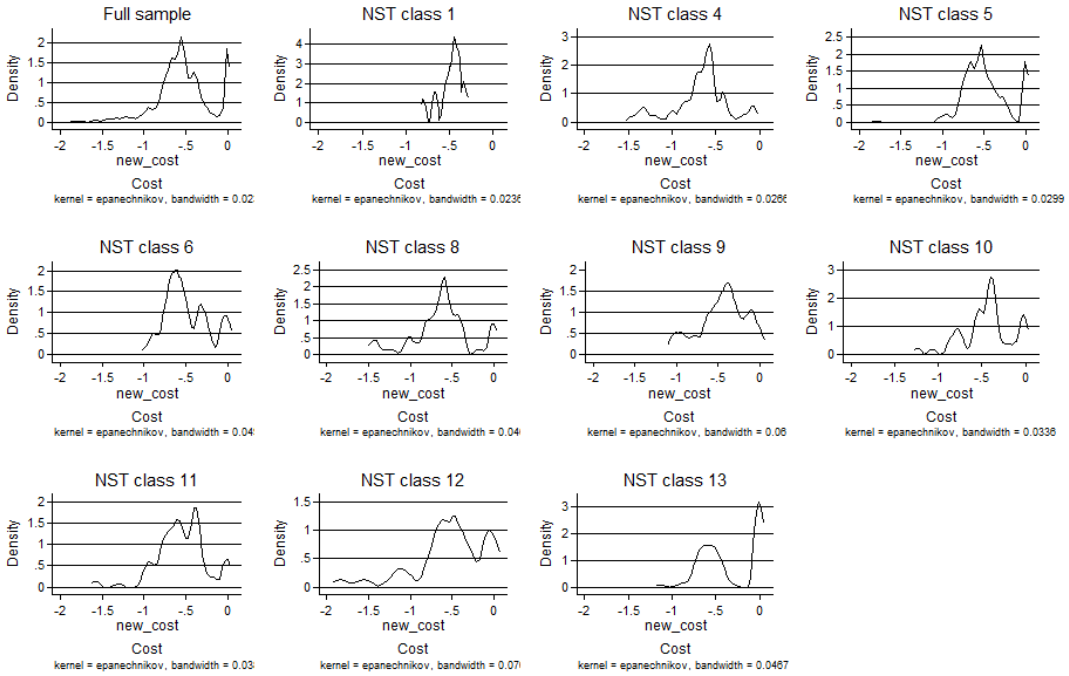
Next, we turn to the results from the latent class model. We have estimated several models with a varying number of classes. Both the AIC and the BIC are minimized with five latent classes. The results from this model are shown in Table 13. It consists of two sets of variables: level-of-service variables whose coefficients show the utility weight associated with each variable, which are

assumed to be the same for each class but may differ across classes. Second, segment variables that explains the class membership of the shippers and whose coefficients are informative of the underlying heterogeneity across shippers.

We use the information in Table 14 to calculate shipper-level cost coefficients and plotting the distribution of coefficients for the available NST 2007 groups as well as for the full sample. This allows us to investigate the variation of shippers' reaction to cost within each segment. This is a key question for the classification of commodity groups in Samgods; shippers within a commodity segments should be relatively homogenous, as this would strengthen the behavioural foundation of the model.

The results are shown in Figure 1. The density in the top-left corner is based on the full sample whereas the others are based on the available NST 2007 groups. The distribution of cost sensitivity in the full sample seems dispersed; there is a large mass of shippers with very low cost sensitivity (which are likely to care more about the service quality) and a sizeable group of shippers who care greatly about the cost.

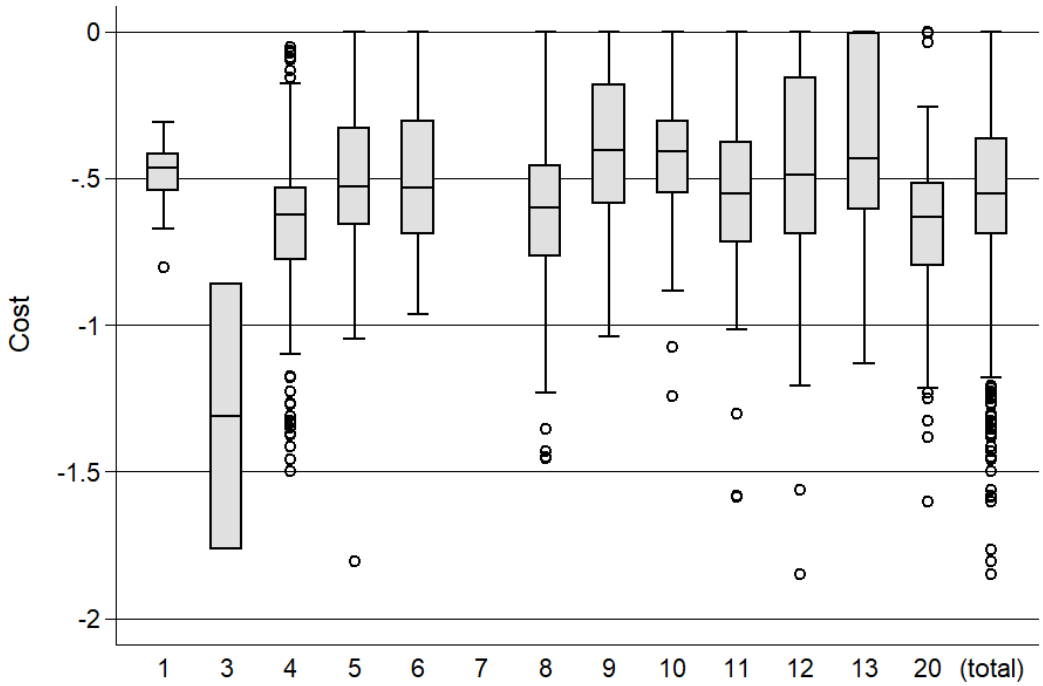
Figure 1: Distribution of shipper-level cost parameters by NST 2007 groups



By visual inspection it appears that many of the NST groups exhibit a more homogenous distribution of cost-sensitivity. This is confirmed in Figure 2, which show box-plots for the different NST groups and the full sample (far right) in Figure 2. The spread between the 25th and 75th percentile (given by the height of the box) is for many groups similar to the one in the population. But the spread of the upper and lower adjacent values (given by the length of the line) is in most cases much lower for the individual commodity groups. In addition, the occurrence of outliers (dots) is higher for the population than for the commodity groups. By these standards, the NST classification produces groups with

relatively more homogenous distributions of cost-sensitivity compared to the full population.

Figure 2: Box-plots of cost coefficients



The robustness of the latent class model, including the estimates of shipper-level cost parameters, depends on whether it is correctly specified and accurately estimated. Following Pacifico and Yoo (2011), we investigate how well the latent class model does in differentiating between several classes of preferences by calculating the average (over shippers) of the highest posterior class probability. A high average posterior probability means that the model does very well in distinguishing among different underlying taste patterns for the observed choice behaviour. As shown in Table 15, our model has an average of 0.57, which suggest it does okay at differentiating among patterns.

Table 15. Summary statistics for class probability

	Mean	St. dev.	Min	Max
Class probability	0.570	0.169	0.261	1

4.6 Implications for commodity classification in Samgods

This analysis has investigated taste heterogeneity in the reaction to transport cost among shippers in Sweden and examined shipment characteristics across commodity groups. It offers four main conclusions.

- First, previous studies provide substantial evidence of the existence of taste heterogeneity in the Swedish freight transport market. In other words, the value shippers assign to transport cost differ between them.

- Second, the empirical analysis in this chapter showed the distribution of shipment characteristics (value, weight and value density) over the NST2007 groups. The relevant lessons from this exercise are that in terms of average shipment value, weight and value density, there are noticeable difference *between* commodity groups 11 and 12 but no large differences *within* groups 1 and 6.

- Third, the empirical analysis documented differences in cost sensitivity *between* shippers with different commodity groups. This holds true for various model specifications. This confirms the notion that commodity class is a key variable for segmenting the freight market. Not accounting for commodity class would risk grouping together companies with very different tastes and preferences that influence the transport chain and shipment size choice.

- Fourth, the analysis showed that many NST groups exhibit a more homogenous distribution of cost-sensitivity compared to the full population of shippers. In these cases, grouping shippers together according to the NST 2007 classification reduced the taste heterogeneity among shippers. This result speaks in favour of recommending the NST 2007 as the new commodity classification in Samgods (although it does not necessarily follow that the NST 2007 classification reduces heterogeneity more than any other classification would).

5 STATISTICAL CONSIDERATIONS

In this chapter, we review additional considerations for the development of the new commodity classification.

5.1 Correspondence with data sources

There should be a high degree of correspondence between the commodity classification of Samgods and the transport data sources. Here, we consider the classification used in the Commodity Flow Surveys, the Road Freight Survey as well as the maritime, air and rail freight statistics.

Table 11. Overview of commodity classification in freight transport data

Data source	Commodity classification
Rail	NST2007 and dangerous goods
Road (Swedish trucks)	NST2007 and dangerous goods
Road (Foreign trucks)	NST2007
Maritime	NST2007
Air	Freight and mail
CFS 2016	NST2007 (1-14, 16, 19, 20)

Source: Transport Analysis (2016d).

The statistics from these sources are generally *published* at the divisional level (i.e. two-digit) of the NST 2007. The exceptions are the CFS 2009 (which uses the current Samgods classification) and the air transport statistics (for which no commodity classification exists). The preliminary commodity classification for the CFS 2016 covers NST 2007 groups 1-14, 16, 19 and 20.

In most cases, the data from each source are *collected* at a more detailed level (three-digit NST2007) and aggregated before publication. Table 12 shows the commodity codes used for the collection of the CFS 2016 and the corresponding NST2007 divisions. Taken together, our conclusion is that a Samgods commodity classification based on NST2007 will have a high degree of compatibility with the freight transport data sources.

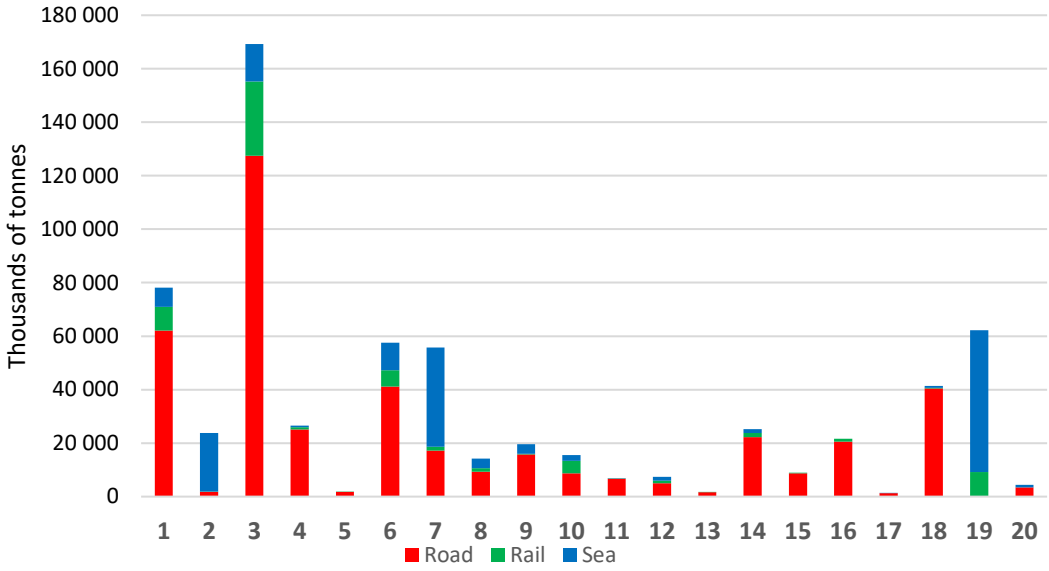
Table 12. Commodity classification in CFS 2016 and NST2007

Commodity	Description	CFS 2016 code	NST 2007 code
Products of agriculture, forestry and fishing	Cereal	10	1
	Potatoes	11	1
	Round wood	12	1
	Raw milk	13	1
	Live animals	14	1
	Fresh fish and products thereof	15	1
	Other forestry products than round wood	16	1
	Other raw materials of animal origin	17	1
	Sugar beet	18	1
Crude petroleum, natural gas and coal	Coal and lignite	20	2
	Crude petroleum	21	2
	Natural gas	22	2
Metal ores and other mining and quarrying products (excl. earth, sand and gravel)	Iron ores	30	3
	Non-ferrous metal ores	31	3
	Chemical and natural fertilizer minerals	32	3
Food products, beverages and tobacco	Meat, meat products, raw hides and skins	40	4
	Fish and fish products, processed and preserved	41	4
	Fruit and vegetables, processed and preserved	42	4
	Animal and vegetable oils and fats	43	4
	Dairy products, beverages and ice cream	44	4
	Grain mill goods, starches, starch goods, prepared animal feeds	45	4
	Other food products and tobacco products	46	4
Wood and products of wood and cork	Manufactured products of wood	50	6
	Wood chips and waste woods	51	6
	Other products of wood	52	6
Paper and paper pulp	Paper and paper products	53	6
	Paper pulp	54	6
Solid and liquid fuels	Coke oven products	60	7
	Liquid refined petroleum products	61	7
	Gaseous, liquefied or compressed, petroleum products	62	7
	Sold refined petroleum products	63	7
	Ethanol	64	8
	FAME	65	7
Chemical products (excl. final goods like pharmaceuticals)	Chemicals, chemical products, man-made fibres, basic plastics and synthetic rubber in primary form	70	8
Earth, stone and construction materials	Earth, stone, gravel and sand	80	3
	Cement, lime and plaster	81	9
	Peat	82	3
	Other construction materials	83	9
Basic metals, fabricated metal products, except machinery and equipment	Basic iron and steel	90	10
	Non-ferrous metals and products thereof	91	10
	Structural metal products, tubes, pipes, hollow profiles and related fittings	92	10
Valuable goods	Textiles and textile products, leather and leather products	100	5
	Printed matter and recorded media	101	6
	Pharmaceuticals and parachemicals	102	8
	Rubber or plastic products	103	8
	Glass and glass products, ceramic and porcelain products	104	9
	Boilers, hardware, weapons and other fabricated metal products	105	10
	Transport equipment	106	12
	Agricultural and forestry machinery	107	11
	Other machines and apparatus; office machinery and computers, television, radio, electronic components; medical, precision and optical instruments	108	11
	Furnitures and other manufactured goods	109	13
Household waste, other waste and secondary raw materials	Household waste	111	14
	Other waste and secondary raw materials	112	14
Equipment, materials utilized in the transport of goods	Containers and swap bodies in service, empty	114	16
	Pallets and other packaging in service, empty	115	16
Other goods, n.e.c	Other goods not elsewhere classified	117	20
	Unidentified goods	119	19

5.2 Freight distribution between commodities

Another aspect is how the transport volume is distributed over the commodities in the new classification. Figure 3 shows the modal shares in tonnes for each of the 20 NST 2007 groups. In terms of absolute volumes, commodity group 3 (Metal ores and other mining and quarrying products; peat; uranium and thorium ores) and group 1 (Products of agriculture, hunting, and forestry; fish and other fishing products) dominate.

Figure 3. Tonnes modal share in Sweden year 2015, by NST 2007 group



Note: Road transport consists of national and international road transports with Swedish registered lorries. Sea transport is international goods loaded and unloaded in Swedish ports and national goods loaded in Swedish ports. NST 2007 classification is found in Appendix 2. Source: Transport Analysis (2016a, table 7a and table 13; 2016b, table d11; 2016c, table 4a, table 4b and table 6a).

Commodity groups 5 (Textiles and textile products; leather and leather products), 13 (Furniture; other manufactured goods), 17 (Goods moved in the course of household and office removals; baggage and articles accompanying travellers; motor vehicles being moved for repair; other non-market goods) and 20 (Other goods) account for very low volumes. Road is the dominating transport mode for most commodity groups. Note that the modal shares in Figure 3 are presented in tonnes, the figures in terms of tonne-kilometres differ from the one above.

5.3 Use of transport modes and containers by NST2007 classification

We also investigate the use of transport modes and containers by NST2007 group. The figures are based on shipments in the CFS 2009 which have been assigned to a NST2007 group based on their CFS commodity classification and a conversion key provided in Appendix 3. Commodities 1-14 and 20 are represented in the survey.

Table 17 shows the percentage of shipments within NST2007 groups that is utilizing containers. In group 14 (wastes), containers are used for the majority of shipments while for several groups (1, 2, 4, 5, 7, 11, 12, 13 and 20) containers are

seldom used. In the other groups, containers are utilized every tenth to twentieth shipment.

Table 17. Percentage container shipments within each NST2007 group

NST2007	1	2	3	4	5	6	7	8
Container shipments	0.035%	0%	3.4%	0.64%	0.11%	5.9%	0.13%	4.9%
NST2007	**9**	**10**	**11**	**12**	**13**	**14**	**20**	**Tot**
Container shipments	7.7%	8.2%	0.83%	0.64%	0.58%	55%	0.38%	0.74%

Source: CFS 2009.

Table 18 shows the use of transport modes by commodity class. The share in each cell shows the number of shipments utilizing a specific mode relative to the number of shipments in that NST2007 group. A transport mode is considered used by a shipper if it is part of the chosen transport chain. Because several modes can be used for each shipment, the percentages within each group sum to over 100.

The use of air transport is of particular relevance for our recommendation of a commodity classification, as it might be meaningful to construct a specific commodity group for air freight. According to Table 18, air transport tends to attract products from group 11 (machinery and equipment), 12 (transport equipment), 10 (basic metals) and, in particular, group 20 (goods not classified elsewhere).

Table 18. Use of transport modes by NST2007

NST2007	Road	Rail	Sea	Air	No. of shipments
1	100%	0.0%	2.2%	0.0%	1,774,016
2	3.4%	3.4%	97%	0.0%	58
3	98%	1.3%	2.7%	0.2%	5,671
4	91%	11%	1.1%	0.0%	194,201
5	100%	0.0%	29%	0.5%	522,652
6	98%	3.1%	12%	0.7%	78,891
7	99%	1.1%	1.0%	0.0%	121,698
8	100%	4.1%	4.4%	0.8%	147,772
9	99%	0.7%	6.3%	0.7%	21,835
10	91%	9.2%	9.6%	1.9%	74,513
11	99%	0.1%	2.3%	3.8%	240,281
12	99%	12%	19%	2.1%	71,435
13	100%	3.2%	48%	0.1%	183,791
14	89%	4%	9%	0%	277
20	100%	8.1%	12%	14%	169,969
Total	99%	1.8%	9.6%	1.1%	3,607,060

Note: The figures show for each NST2007 group, the share of shipments using each transport mode (as direct transport or at least once in a transport chain). Because several modes can be used for each shipment, the percentages within each group sum to over 100. Source: CFS 2009.

5.4 Implications for commodity classification in Samgods

The findings from the preceding sections can be summarized as:

- A commodity classification based on the NST 2007 will have a high degree of compatibility with the freight transport data sources.
- There are differences in the distribution of the tonnage across the NST2007 groups. Commodity commodity group 3 and group 1 dominate, while several groups (5,13, 17, 20) account for very low volumes.
- Container transports are frequently used in group 14 (wastes) while for several groups (1, 2, 4, 5, 7, 11, 12, 13 and 20) containers are seldom used. In the rest of the groups (3, 6, 8, 9 and 10) containers are utilized every tenth to twentieth shipment.
- Air transport tends to attract products from group 11 (machinery and equipment), 12 (transport equipment) and 10 (basic metals), which all exhibit relatively high value density.

6 RECOMMENDATION

Table 19 below shows our suggested classification, which is based on the results from chapter 2-5. The components do not necessarily have equal weight for our recommendation.

We recommend the classification to be based on the divisional level (i.e. the two-digit code) of the NST2007 and to include commodity groups 1-14. We recommend group 1 to be split up into one category containing round wood and another category consisting of the rest of the items. Further, we recommend to add a commodity group for air freight by combining fractions from other commodities. In total, our recommended classification consists of 16 groups. In the following sections, we provide motivation for our recommendation.

Table 19. Suggested classification

NST division	Description	No.
01*	Products of agriculture, hunting, and forestry; fish and other fishing products (excl. round wood)	1
02	Coal and lignite; crude petroleum and natural gas	2
03	Metal ores and other mining and quarrying products; peat; uranium and thorium.	3
04	Food products, beverages and tobacco	4
05	Textiles and textile products; leather and leather products	5
06	Wood and products of wood and cork (except furniture); articles of straw and plaiting materials; pulp, paper and paper products; printed matter and recorded media	6
07	Coke and refined petroleum products	7
08	Chemicals, chemical products, and man-made fibers; rubber and plastic products; nuclear fuel	8
09	Other non-metallic mineral products	9
10	Basic metals; fabricated metal products, except machinery and equipment	10
11	Machinery and equipment n.e.c.; office machinery and computers; electrical machinery and apparatus n.e.c.; radio, television and communication equipment and apparatus; medical, precision and optical instruments; watches and clocks	11
12	Transport equipment	12
13	Furniture; other manufactured goods n.e.c.	13
14	Secondary raw materials; municipal wastes and other wastes	14
01**	Round wood	15
- ***	Air freight	16

* corresponds to NST division 01, excluding round wood. ** corresponds to NST division 01, including round wood only. *** consists of fractions of other commodity groups.

6.1 NST2007 classification

We propose the classification to be based on the NST 2007 nomenclature. This is motivated by several reasons. First, the NST 2007 is suitable when constructing the demand matrices in the Samgods model. As shown in chapter 3, the new classification should be possible to aggregate directly from the detailed CN8 classification and it should be possible to categorize CFS data according to the new classification. NST2007 meets both these conditions. The NST2007 is also

likely to result in sparser correspondence tables connecting industry sectors to commodity types, which will add accuracy to the forecast matrices.

Second, basing the classification on NST2007 provides comparability and compatibility with freight transport statistics. As shown in section 5.1, the main freight data sources are based on the NST2007 nomenclature.

Third, using the NST2007 in Samgods provides comparability and compatibility with other national freight models and any output thereof. As shown in chapter 2, the NST 2007 classification is used in the German and Flemish model system and there are plans to use it in the next version of the Dutch national freight transport model system. Although the freight models in Norway, Denmark and Finland use the NACE/SITC classification we do not recommend it for Samgods. The reason is that the Commodity Flow Survey 2016 is based on the NST2007 (see Table 12).

Finally, the NST2007 is likely to provide an appropriate segmentation of firms in the freight market. The results from chapter 4 suggest that grouping shippers according to the NST2007 classification reduces the taste heterogeneity among shippers.

6.2 Inclusion of commodities

Commodity group 14 (Secondary raw materials; municipal wastes and other wastes) is not in the current classification in Samgods. As pointed out in section 3.4, including this group in the new commodity classification means that the matrix generation method needs to be extended to additional data sources and estimations and that model development probably also has to be done. Recycling activities and e.g. heating plants (that use waste as an input) will probably have to be included in the considered industry sectors.

Our recommendation to include this group in the new classification is motivated by the potential future increase in transports of this group, considering the trends towards sustainable production and logistics. Regarding data sources, information on group 14 is covered by national freight transport statistics and by the CFS. There are also waste statistics from the Swedish Environmental Protection Agency that could be utilized.

6.3 Exclusion of commodities

We recommend to exclude the NST2007-commodities 15-20, partly because they are not connected to the European product and activity classifications CPA (Classification of products by activity) and NACE (Statistical classification of economic activities). Commodities 15-20 account for about a fifth of the amount transported in Sweden, with grouped goods and unidentifiable goods making up the majority (see Figure 3). These groups also cover a considerable share of the container transports.

Our recommendation to exclude these groups is also motivated from the point-of-view of demand matrix generation. These groups are not in the current Samgods classification and including them would require an extension of the matrix generation method to additional data sources and calculation steps. It is also motivated by the fact that containers are still utilized in the groups of the recommended commodity classification. As showed in chapter 5, container transports take place for groups 6, 8, 9, 10 and, especially, 14.

Regarding commodity 16 (containers and swap bodies in service, empty) it is important to remember that Samgods treats containers as a specific type of cargo unit. This makes it possible to model how competitive container transports are/will be compared to "conventional transports". The modelling of firms' choice between predefined container-chains and non-container-chains typically implies that a share of the different commodities is transported by container on vessels, combi-trains or trucks. It is important to validate the model outcomes measured in loaded/unloaded tonnes in containers, container tonne-km and vehicle-km of the vehicles/vessels that transport the containers. Empty containers are not modelled in the existing Samgods version.

6.4 Inclusion of a commodity group for air freight

A commodity group for air freight is used in the current Samgods model. The benefit of having this category is that it can make the model produce more reasonable levels of air transport. Without it, air transports risk being chosen in the model far too seldom compared to the actual levels. Considering this, we recommend, at least in the short-term, to include an air freight commodity in the Samgods model and construct it by allocating shares from other commodity groups.

Section 5.3 provides an overview of the occurrences of air transports for each NST2007 group in the CFS 2009. We recommend constructing the air freight commodity group based on a similar approach using figures from the CFS 2016, as this data set will be up-to-date and have a closer connection to the NST2007 than what the previous commodity flow surveys have.

6.5 Splitting of commodities

NST2007 groups could be split if there are decisive differences between the sub-groups. The main candidates are group 1 (products of agriculture, hunting, and forestry; fish and other fishing products) and 6 (Wood and products of wood and cork (except furniture); articles of straw and plaiting materials; pulp, paper and paper products; printed matter and recorded media). They account for a substantial share of the amount of tonnes transported and consist of products that a priori are dissimilar in several regards.

Another reason for splitting groups is the occurrence of diverging trends for products within the same commodity group. If the demand for the products are guided by different underlying variable, their growth rates may differ. However, if the economic forecasts will be made the same way in the future as they currently are, there will be separate growth rates available for the various sub-

groups. (The growth rate for an aggregated group will be a weighted average of its sub-groups). Separate growth rates can be used because the input data are equally detailed. Diverging trends is therefore unlikely to be a problem and is not a reason good enough for splitting the groups.

As for group 1, the analysis of the CFS 2009 showed that the mean shipment value, weight and density in most sub-groups are not very different from the group average (see section 4.6). This is a reason for keeping the group as it is. However, as noted in section 3.3, the trade of products of agriculture and hunting on the one hand and forestry products on the other, takes place in more or less separate systems. [26] The product groups originate from different industries, namely the agricultural sector and the forestry sector. These industries are geographically connected to certain land-uses and the geographical distribution of the agricultural sector differs from that of the forestry sector. In the other end of the trade relation, the consumption of the respective sub-groups takes place in different industries as well, namely the paper/wood industry and the foodstuff industry (mainly). Estimating the two subgroups in aggregation thus could result in inaccurate P-C relations geographically; e.g. a production unit in a strictly agricultural area could be incorrectly connected to a paper industry plant. Further, the transport system of the forestry products is quite unique, with specialized vehicles and single-purpose roads, which is another argument for separating the two sub-groups of group 1. Considering this, we recommend to create an own category for round wood.

As for group 6 (Wood and products of wood and cork (except furniture); articles of straw and plaiting materials; pulp, paper and paper products; printed matter and recorded media), section 4.6 showed that average shipment value is similar for paper, paper products and products of wood and cork. Shipments of pulp tend to have higher value while printed and recorded media have the lowest value. Because the average shipment weight follows the same pattern, value density is not very different between the groups, except for printed and recorded media. Altogether, we do not deem the shipment characteristics in this group to be different enough to motivate a split, considering the additional efforts in matrix generation it would entail.

As mentioned in chapter 2 we do not think that it is necessary to specify an own commodity group for dangerous goods because NST 2007 groups 2 (crude petroleum), 7 (petroleum products) and 8 (chemicals) already cover most of the dangerous goods that are transported in Sweden.

The Norwegian approach to connect different commodity types to different vehicle types (thermo truck, tank truck or articulated trailer) is not relevant as the Samgods model does not differentiate between these vehicle types.

[26] Products of fishing is not included in the PWC matrices due to the very specific trade patterns. Fish products that are transported from the harbor and onwards can be seen as foodstuff.

6.6 Merging of commodities

NST2007 groups could be merged if they are similar in important regards. Two candidates are group 11 (Machinery and equipment not elsewhere classified) and 12 (Transport equipment). As shown in section 4.6 however, these groups exhibit differences in important shipment characteristics. Group 11 have much lower shipment weight and value, both on average and across the distribution of shipments, compared to group 12. There is also a difference in value density between the groups, although it is less pronounced. In other words, we see no apparent similarities that could motivate a merge of these groups.

6.7 Implications for the set-up in Samgods model

The commodities in the current classification in Samgods are linked to three aggregate classes (dry bulk, liquid bulk and general cargo, see Table 1). These aggregated classes provide different levels of loading and unloading cost and time in the current model system. We recommend to skip these aggregated commodity classes and calculate transhipment costs and times for the 16 commodity classes.

The commodities in the current classification are also linked to 12 STAN-commodities. We do not consider it necessary to aggregate the proposed commodities into 12 STAN-commodities. Aggregation is more meaningful when the number of commodities (to be aggregated) is large, as with the current 32-35 commodities. The proposed classification contains only 16 commodities and aggregation is therefore of limited use.

REFERENCES

Arunotayanun, K., Polak, J. (2011) Taste heterogeneity and market segmentation in freight shippers' mode choice behaviour. Transportation Research Part E, 47, p. 138-148.

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

Abate, M., Vierth, I., de Jong, G. (2014) Joint econometric models of freight transport chain and shipment size choice. CTS Working Paper 2014:9

Borremans, D. et al. (2015) Opbouw strategisch vrachtmodel Vlaanderen versie 4.1.1, http://www.verkeerscentrum.be/extern/VlaamseVerkeersmodellen/StrategischVrachtmodelVlaanderen/01.1_20150331_Rapport_Opbouw_Strategisch_vrachtmodel_Vlaanderen_versie4.1.1.pdf, Antwerp 2015.

Bovenkerk, M. (2005) SMILE+: the new and improved Dutch national freight model system, Paper presented at the European Transport Conference, Strasbourg .

BVU and TNS Infratest (2014) Entwicklung eines Modells zur Berechnung von modalen Verlagerungen im Güterverkehr für die Ableitung konsistenter Bewertungsansätze für die Bundesverkehrs-wegeplanung, vorläufiger Endbericht, BVU/TNS Infratest, Freiburg/München.

Cameron, C., Trivedi, P. (2010) Microeconometrics using Stata. Stata Press.

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. In E. van de Voorde, & Th. Vanelslander (Eds.), Applied transport economics, a management and policy perspective. Antwerpen: De Boeck.

de Jong, G., Ben-Akiva, M., Baak, J. (2008) Method Report - Logistics Model in the Norwegian Freight Model System. Deliverable 6A. , Significance, Den Haag.

de Jong, G., Baak, J. (2015) Method Report - Logistics Model in the Swedish National Freight Model System. Deliverable 6B. Significance, Den Haag.

de Jong, G., Ben-Akiva, B., Bexelius, S., Rahman, A., van de Voort, M., Florian, M., Baker, M., Gibbs, P., Hester, U., Lingbrant, L., Lundqvist, C. (2004) The Specification of Logistics in the Norwegian and Swedish National Freight Model Systems. Model scope, structure and implementation plan. SIKA Project 04074.

Eurostat (2017) RAMON, Eurostat's Metadata Server, viewed March-April 2017

Fjendbo Jensen, A., M. Thorhauge, G.C. de Jong, J. Rich, T. Dekker, D. Johnson, M. Ojeda Cabral, J. Bates and O. Anker Nielsen (2016) A model for freight transport chain choice in Europe, paper presented at the HEART 2016 conference, Delft.

Greene, W., Hensher, D. (2007) Heteroscedastic control for random coefficients and error components in mixed logit. Transportation Research Part E 47, p. 610-623.

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

Hovi, I. B., Caspersen, E., Grue, B. (2015). Commodity flow matrices for Norway, TØI rapport 1399/2015 Transportøkonomisk institutt.

Hovi, I. B., Hansen, W., Gjerde Johansen, B., Jordbakke, G.N., Madslien, A. (2017) Forecasts for Norwegian freight transport 2016-2050, TØI-report 1555/2017 (in Norwegian)

Jin, Y, I. Williams and M. Shahkarami (2005) Integrated regional economic and freight logistics modeling: results from a model for the Trans-Pennine Corridor, UK, paper presented at the European Transport Conference 2005, Strasbourg.

Johnson D., de Jong, G., (2011) Heterogeneous response to transport cost and time and model specification in freight mode and shipment size choice. Paper presented at the International Choice Modelling Conference 2011

SAMPLAN (2001) The Swedish Model System for Goods Transport – Samgods, SAMPLAN report 2001:1

Keane, M, Wasi, N. (2013) Comparing Alternative Models of Heterogenetiy in Consumer Choice Behaviour. Journal of Applied Econometrics 28, p. 1018-1045.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In Frontiers in Econometrics, Zarembka P (ed.). Academic Press: New York; 105–142.

Pacifico, D., Yoo, H. (2013): A Stata module for estimating latent class conditional logit models via the Expectation-Maximization algorithm. Stata Journal 13 (3), pp. 625-639.

Rich, J., PM Holmblad, C O Hansen (2009) A weighted logit freight mode-choice model, Transportation Research Part E: Logistics and Transportation Review, Volume 45, Issue 6, Pages 1006–1019

Significance, NEA and Demis (2010) Schatting BasGoed, Rapportage DP1, Report for DVS-Rijkswaterstaat, Significance Den Haag.

Significance (2012) Logistic Model Denmark, Report for DTU, Significance, Den Haag.

Significance (2014) Prototype Fehmarn Belt crossing model, Report for DTU, Significance, Den Haag.

Statistics Norway (2012) The Norwegian commodity flow survey 2008. Documentation and Methodology, SSB Notater 60/2012, https://www.ssb.no/a/publikasjoner/pdf/notat_201260/notat_201260.pdf (in Norwegian)

Statistics Norway (2016) The Norwegian commodity flow survey 2014, https://www.ssb.no/transport-og-reiseliv/statistikker/vsu/aar/2016-05-26#content (in Norwegian)

Trafikverket (2016). Representation of the Swedish transport and logistics system in Samgods v. 1.1. 2016-11-25 .

Trafikverket (2017) Kostnader Samgods ASEK. Retrieved from http://www.trafikverket.se/tjanster/system-och-verktyg/Prognos--och-analysverktyg/Samgods/

Train, K. (2003) Discrete Choice Methods with Simulation. Cambridge University Press, MA.

Transport Analysis (2011): Transportmodeller i ett internationellt perspektiv 2011. Report 2011:6.

Transport Analysis (2011b): Metodrapport Varuflödesundersökningen 2009. PM 2011:13.

Transport Analysis (2016a): Swedish national and international road goods transport 2015. Statistics 2016:27.

Transport Analysis (2016b): Rail traffic 2015. Statistics 2016:18.

Transport Analysis (2016c): Shipping goods 2015. Statistics 2016:17.

Transport Analysis (2016d): Statistik och kunskapsunderlag om godstransporter. PM 2016:11

Vierth, Karlsson & Westin, (2016) *Validering av sjötransporter i Samgodsmodellen Version 1.1.* VTI (VTI-Notat 30/2016) .

WSP (2012) Base Year Freight Matrices, Final Report for ITEA Division of DfT, WSP, Cambridge.

WSP, (2013) PM: New correspondence keys for associating industries with commodities, delivery to Trafikverket 2013-06-14.

WSP, Sweco & KTH (2015) PWC Matrices: new method and updated Base Matrices, Final Report, delivery to Trafikverket 2015-05-18

WSP (2015) Nya varuvärden 2040 – data, metod och resultat, delivery to Trafikverket 2015-11-09 (in Swedish)

WSP & Sweco (2016) PWC-matriser år 2040, delivery to Trafikverket 2016-03-23 (in Swedish)

APPENDIX 1

PM: New correspondence keys for associating industries with commodities

Purpose of the key

To create the PWC matrices, which are giving the demand of freight transport between Swedish municipalities (and regions abroad); several different statistical data sets are used. The PWC matrices are defined for a number of separate commodity groups. Part of the statistics needed for the calculations are available for some kind of classification of commodities. But the major part instead describes different aspects of industries, such as employment, production value, trade and export, etc. To be able to use this information to analyse trade with different commodities, each industry has to be associated with a number of commodity types. More specifically, the production value per industry in each region has to be distributed over a set of commodity types.

In order to do this, a key is needed, that for each industry provides a distribution of the production value over the specified commodity types, in a way that the sum of all fractions is 100 % for every industry (this applies to the production of goods and not services, electricity or other products that are not transported).

The key should be defined on a detailed level, so that e.g. differences in the composition of products derived from agricultural units in northern and southern Sweden are taken into account. This way, regional differences are included in the calculations.

The key will be applied on data defined on the industry classification SNI 2007 and give output divided into the commodity groups of the Samgods model.

Existing key

In ”Swedish Base Matrices Report” (Edwards et al, 2008-03-11) , the construction of the currently used key is described. This key has been used to create the existing base matrices. The key is applied on data on SNI 92 2-digit level and allocates the values into the 34 Samgods commodities (these can be aggregated into the 12 STAN commodity groups).

The 34 Samgods commodities are based on the European commodity classification NST/R (Standard Goods Classification for Transport Statistics/Revised, 1967). The existing key has two parts, where the first step is a key associating SNI 92 with the Combined Nomenclature (CN, in Swedish KN) which classifies commodities on a detailed level (using 8-digit codes), and the second step is a key associating CN with NST/R[27].

The first sub-key (SNI-CN) has been calculated using the foreign trade statistics 1998-2004, which is reported with both industry codes (SNI) and commodity codes (CN) for all import to and export from Sweden. An average of import and export flows has been used and applied to all flows including the domestic trade.

The second sub-key (CN-NST/R) was derived in the 90’s and was updated to the current version of CN in 2005by Statistics Sweden. There was also a demand for a distinction for round wood and pulp wood in the key, which was not provided by the CN classification. The product group was split into fractions by identifying differences in the SEK per tonne prices in the Foreign Trade Statistics.

Available data

[27] For a more complete description of the different commodity classifications referred to here, see final section.

A number of data sets that could be used for creating new keys have been identified, and are briefly described here.

Production of commodities and industrial services (Industrins varuproduktion, IVP)

The Production of commodities and industrial services (IVP) is a yearly survey that has been conducted since 1996. The IVP reports production of commodities using CN on an 8-digit level registered in values and quantities. The population of entities that are being surveyed are defined by the following criteria's

- Workplaces with a minimum of 10-20 employees with their main branch in the industry sector.
- Workplaces with their main branch in the industry sector with less than 10-20 employees and a net turnover that is 50 million SEK or more per year.
- Workplaces with their secondary branch in the industry sector if the secondary branch have a minimum of 10-20 employees.

In 2011 the population consisted of approximately 4000 workplaces and the survey had a weighted response frequency of 99 percentages. The production of commodities in those workplaces that do not fulfil the criteria's are model estimated, since they are not a part of the survey population. For each workplace the main and secondary industry branch can be identified on an SNI 5-digit level. Given that the IVP reports production of commodities on a CN 8-digit level for each workplace a key can be generated, that links economic activity on a SNI 5-digt level with production of commodities and industrial services on a CN 8-digit level.

Foreign trade statistics

For export flows, the foreign trade statistics could be used the same way as has been done for the existing key (except that only the export part of the statistics should be used). The composition of commodities in the production for export is assumed to differ from the production for domestic trade. Therefore, a separate key for export will be derived from the foreign trade statistics. The data is reported using both industry codes (SNI) and commodity codes (CN) and consequently provides a key between these classifications.

Eurostat correspondence table

The Eurostat Metadata Server RAMON provides correspondence tables for several statistical classifications. The NST/R to NST 2007 conversion table could be used here. NST 2007 is the new version of NST/R, but since these two versions are derived from different origins[28], the conversion has to be made via CN and an additional commodity classification, CPA. The NST/R-NST 2007 conversion table thus includes a correspondence table between CN and NST/R, and between CN and NST 2007 (via CPA)[29]. However, the correspondence between NST/R and NST 2007 are not 1:1, i.e. one NST/R code corresponds to several NST 2007 codes, and vice-versa. The advantage of using the detailed information of CN 8 is that each commodity group only correspond to one group in NST/R and NST 2007, hence the correspondence goes from many to one.

Suggestion for new key(s)

As mentioned above, the existing Samgods commodities are based on the NST/R classification. However, this classification has been replaced by NST 2007, meaning that NST/R is being abandoned for transport statistics. We therefore suggest that the new key will use NST 2007 (at some aggregation level that also may be slightly adjusted to fit the purposes of the Samgods model) as commodity classification.

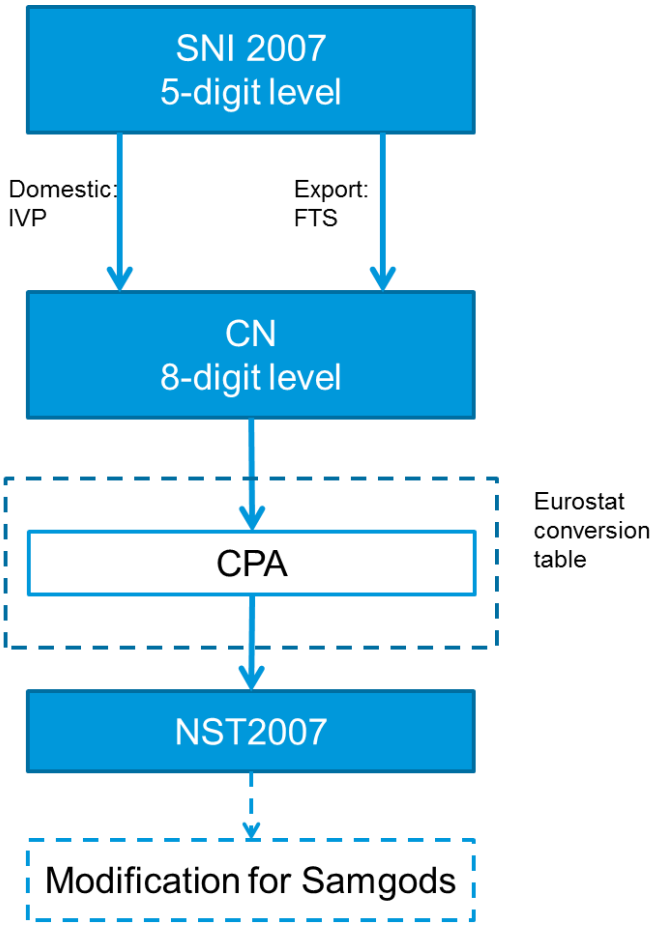
[28] For more information, see final section

[29] "Maintenance of the NST/R – NST 2007 conversion table", Artemis Information Management, September 2008

There is no key available for converting NST 2007 data to the Samgods commodities and vice-versa. The conversion between NST/R and NST 2007 is not straight-forward, see the final chapter. This means that if NST 2007 is chosen for the new key, the option to adjust the Samgods commodity classification should be considered. The drawbacks would be that some of the tables and parameter values in the Samgods model have to be re-calculated, and it will be more complicated to compare new model results to previous results. In return, model results may be validated using new transport statistics directly without the need for transformation between product classifications, but also, extra uncertainties in the calculation of the PWC matrices due to additional conversion between product classifications are avoided.

The calculations of the new key could then be done according to the following schematic picture:

- SNI 2007 values on the 5-digit level are converted to CN on the 8-digit level. The 5-digit level for SNI 2007 is detailed enough to take most regional differences within industries into account. This step is done separately for the domestic part of the key, using the statistics for Production of commodities and industrial services (IVP), and for the part concerning exports, using the foreign trade statistics (FTS).
- CN is converted to NST 2007 using the Eurostat conversion table as a basis.
- Depending on the demands for the Samgods model, the NST 2007 classification can be adjusted e.g. by aggregating some codes and leaving other codes disaggregated.



More technically, the keys could be described the following way.

SNI industries, 5-digit level:	821 industries, index *h*
CN commodities, 8-digit level:	9600 commodities, index *u*
CPA commodities, 6-digit level:	3142 commodities, index *w*
NST 2007 commodities, 3-digit level:	81 commodities, index *z*

Sub-key SNI – CN, domestic: *SNICN*^{D} 821 × 9600

Element (*h*, *u*) gives the fraction of the production value (for domestic trade) of industry *h* that consists of CN commodity *u*, so that $\sum_{u=1:9600} SNICN^D(h,u) = 1$ for all *h* = 1:821

Sub-key SNI – CN, exports: *SNICN*^{X} 821 × 9600

Element (*h*, *u*) gives the fraction of the production value (for exports) of industry *h* that consists of CN commodity *u*, so that $\sum_{u=1:9600} SNICN^X(h,u) = 1$ for all *h* = 1:821

Sub-key CN – CPA: *CNCPA*
9600 × 3142

Element (*u*, *w*) gives the fraction of CN commodity *u* that is included in CPA commodity *w* (in terms of economic value), so that $\sum_{w=1:3142} CNCPA(u,w) = 1$ for all *u* = 1:9600

Sub-key CPA – NST 2007: *CPANST* 3142 × 81

Element (*w*, *z*) gives the fraction of CPA commodity *w* that is included in NST 2007 commodity *z* (in terms of economic value), so that $\sum_{z=1:81} CPANST(w,z) = 1$ for all *w* = 1:3142

"Total key" SNI – NST 2007, domestic: *SNINST*^{D} 821× 81

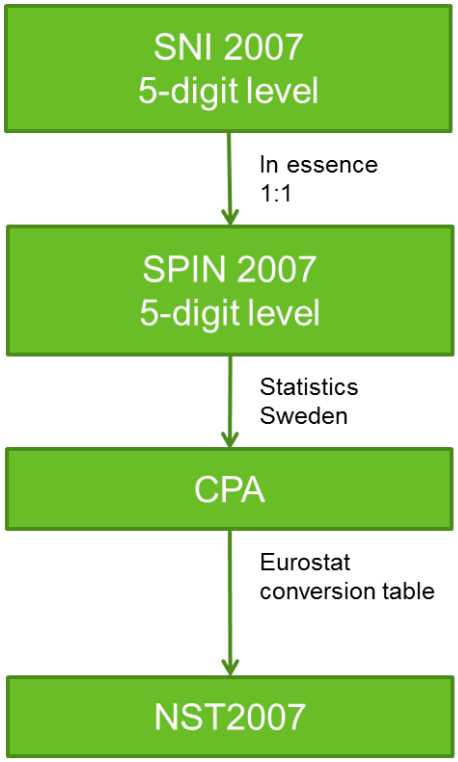
Element (*h*, *z*) gives the fraction of the production value (for domestic trade) of industry *h* that consists of NST 2007 commodity z, so that $\sum_{z=1:81} SNINST^D(h,z) = 1$ for all *h* = 1:821
SNINST^D = *SNICN*^{D} *CNCPA CPANST*

"Total key" SNI – NST 2007, exports: *SNINST*^{X} 821× 81

Element (*h*, *z*) gives the fraction of the production value (for exports) of industry *h* that consists of NST 2007 commodity z, so that $\sum_{z=1:81} SNINST^X(h,z) = 1$ for all *h* = 1:821
SNINST^X = *SNICN*^{X} *CNCPA CPANST*

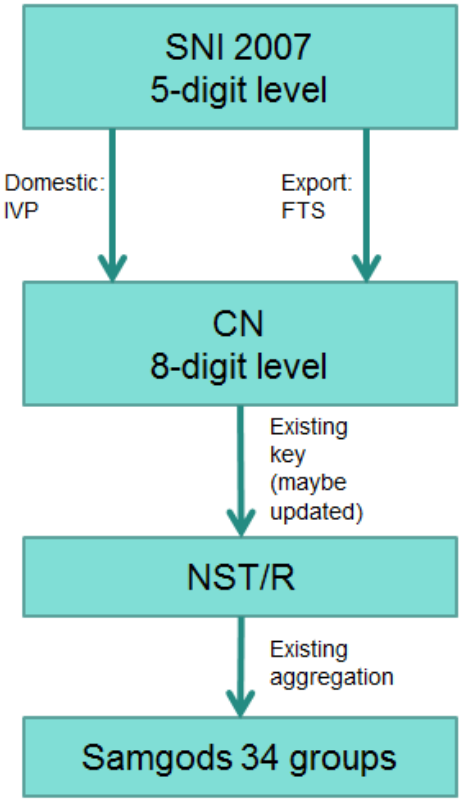
Alternative calculation methods

If, for some reason, one or more steps in the calculation described above are not possible to make, an option could be to skip the use of the CN classification. Instead, the connection between the SPIN and CPA classifications could be utilised. SPIN is a commodity classification that is tightly connected to the industry classification SNI[30] and is the Swedish version of the CPA classification. The correspondence between SPIN and CPA is not 1:1, but differences could maybe be solved if necessary. A description of the correspondence can be found at Statistics Sweden. From CPA to NST 2007, the same method as described above can be used. The drawback here is that the detailed information of IVP and FTS is not utilised.



A third option could be to use the same method that has been used for the existing key, but with updated values and completed with the IVP statistics. The first step of the calculation is thus the same for this option as for the first option described above. Keeping the existing key for the following steps would imply that the NST/R connection to Samgods is preserved, meaning that no update of the Samgods commodities is needed. The drawback is that any conversion to NST 2007 (e.g. for validating model results to current transport statistics) would be more complicated.

[30] See final section



Classifications of industries and commodities

SNI – classification of industries

SNI (Standard för Svensk Näringsgrensindelning) is the label for the Swedish Standard Industrial Classification. According to the webpage of Statistics Sweden [31], SNI is based on the EU standard NACE Rev.2. Production units are classified according to the type of activity pursued. Those units can be companies or other units registered as workplaces. A unit can have several SNI codes, if more than one activity is pursued there. 2008 a new set of SNI codes were introduced; SNI2007. Earlier versions are SNI2002 and SNI92. There are relatively big differences between SNI2007 and the earlier versions.

The SNI2007 classification divides industries into 21 sections, denoted by letters. Those are divided into 88 main groups (2-digit codes), 272 groups (3-digit codes), 615 subgroups (4-digit codes) and 821 detail groups (5-digit codes)[32].

CN – Combined Nomenclature

The Combined Nomenclature is used by all EU countries in their foreign trade statistics and common custom tariff, in Sweden it is also used in the production of commodities and industrial services, IVP. CN 8 is the most detailed level of commodity classification consisting of 8 digits. In 2009 the CN 8 consisted of 9 600 commodity groups. In Sweden a large share of all production and foreign trade is connected to relatively few CN 8 commodity groups, while the rest of the commodity groups have very small values or are equal to zero. The CN 8 is summed hierarchically to the more aggregated commodity groups CN 6, CN 4 and CN 2 which consist of 6 respectively 4 and 2 digits.

[31] http://www.scb.se/Pages/List___257409.aspx, 2013-05-02
[32] http://www.scb.se/Pages/List___257220.aspx, 2013-05-02

Every year small changes are made in the description and classification of the CN 8 commodity groups. Approximately every fifth year the CN 6, CN 4 and CN 2 are revised which implies larger changes to the CN 8 commodity groups. The changes are made to harmonise the commodity classification with technical development and changes in trade patterns.

NST/R

NST/R is the version of the standard goods classification for transport statistics, which was in use from 1967 to 2007, by e.g. member states of the EU. According to the documentation[33], *"the NST/R takes the form of a list with 176 heading for goods which are classified as far as possible on the basis of their nature, processing stage, methods of transport and total tonnage transported; (...) The analytical structure of the NST/R divides the 176 headings of the classification into 10 chapters and 52 main groups, according to a system which consists of:*

- *one digit for the chapters,*
- *two digits for the groups,*
- *three digits for the headings."*[34]

The commodity type classifications used for the Samgods model is based on NST/R, with some modifications, see the Samgods/STAN chapter below.

NST 2007 – classification of commodities

According to the webpage of Eurostat[35]:

"The Standard goods classification for transport statistics abbreviated as NST (2007), is a statistical nomenclature for the goods transported by four modes of transport: road, rail, inland waterways and sea (maritime).

As NST 2007 considers the economic activity from which the goods originate, each of its items is strongly connected to an item of the European Union product and activity classifications Classification of products by activity (CPA) and Statistical classification of economic activities (NACE), which themselves are consistent with their counterparts at UN level, CPC and ISIC."

NST 2007 divides commodities into 20 main groups and 81 subgroups[36] and has been in use since 2008. As described above, NST 2007 is based on the production process where the goods are coming from, while NST/R is based on the physical characteristics of the goods[37]. Therefor the conversion between the two versions is not straight-forward, but has to be done via CN and CPA[38].

---

[33] "Standard goods classification for transport statistics – NST/R", can be found in "Introduction to the classification" at http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NSTR_1967&StrLanguageCode=EN&StrLayoutCode=HIERARCHIC

[34] The classification can be found at http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NSTR_1967&StrLanguageCode=EN&IntPcKey=&StrLayoutCode=HIERARCHIC

[35] http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:NST, 2013-04-15

[36] The classification can be found at http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NST_2007&StrLanguageCode=EN&IntPcKey=&StrLayoutCode=HIERARCHIC&IntCurrentPage=1

[37] http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/en/road_go_esms.htm, 2013-04-15

[38] For more information, see "Maintenance of the NST/R – NST 2007 conversion table", Artemis Information Management, September 2008. The conversion table can be found at http://ec.europa.eu/eurostat/ramon/relations/index.cfm?TargetUrl=LST_REL&StrLanguageCode=EN&IntCurrentPage=8 (NST/R 1967 – NST 2007)

Samgods34 and STAN12 – classification of commodities

The Samgods model operates with 34 commodity groups separately. Those 34 groups can be directly aggregated to the 12 commodity types of the old STAN model. As mentioned above, the Samgods34 commodities are based on the NST/R classification. More specifically, the Samgods34 classification is an aggregation of NST/R on the most detailed level – the 176 headings have been aggregated into 33 groups (the last group of the 34 is for air freight). The Samgods34 classification can be found in the Samgods documentation[39].

SPIN – classification of commodities

SPIN is a product classification for goods, services and other products such as electricity. It is based on the production activity that the products come from and is closely connected to the industry classification SNI. With the exception for the education sector, the five first digits of the SPIN2007 codes are identical with the 5-digit codes if the SNI activity the product group is derived from. The SPIN classification is also connected to its European counterpart, CPA (see the next subsection). The 2007 version of SPIN contains quite big differences compared to the 2002 version.[40]

CPA – classification of commodities

CPA – Classification of Products by Activity – is EU's product classification based on the production process that results in the products. The products can be goods or services. CPA is linked to the EU industry classification the same way as SPIN is linked to SNI. The current version is CPA2008. SPIN is adjusted to CPA so that the first four digits in the respective codes are equal.

CN on the 8-digit level are linked to the 6-digit CPA2008 codes which are linked to the classification of commodities NST 2007on a 3-digit level. The linkage is done through existing Eurostat correspondence tables[41].

[39] See e.g. "Representation of the Swedish transport and logistics system", VTI notat 17A-2009, page 14-15 (please note that the NSTR codes given in the table on page 15 do not equal the official NST/R codes given by Eurostat – the groups are an aggregate of NST/R but the notation used for the codes is another). The document can be found at e.g. http://www.trafikverket.se/PageFiles/64819/representation-av-det-svenska-godstransport--och-logistiksystemet-logistikmodell-version-200.pdf

[40] http://www.scb.se/Pages/List____303898.aspx, 20130515

[41] More information and an index of existing and downloadable correspondence tables can be found at http://ec.europa.eu/eurostat/ramon/relations/index.cfm?TargetUrl=LST_REL

APPENDIX 2

NST 2007 classification

Code	Description
01	Products of agriculture, hunting, and forestry; fish and other fishing products
02	Coal and lignite; crude petroleum and natural gas
03	Metal ores and other mining and quarrying products; peat; uranium and thorium
04	Food products, beverages and tobacco
05	Textiles and textile products; leather and leather products
06	Wood and products of wood and cork (except furniture); articles of straw and plaiting materials; pulp, paper and paper products; printed matter and recorded media
07	Coke and refined petroleum products
08	Chemicals, chemical products, and man-made fibers; rubber and plastic products; nuclear fuel
09	Other non-metallic mineral products
10	Basic metals; fabricated metal products, except machinery and equipment
11	Machinery and equipment n.e.c.; office machinery and computers; electrical machinery and apparatus n.e.c.; radio, television and communication equipment and apparatus; medical, precision and optical instruments; watches and clocks
12	Transport equipment
13	Furniture; other manufactured goods n.e.c.
14	Secondary raw materials; municipal wastes and other wastes
15	Mail, parcels
16	Equipment and material utilized in the transport of goods
17	Goods moved in the course of household and office removals; baggage and articles accompanying travellers; motor vehicles being moved for repair; other non market goods n.e.c.
18	Grouped goods: a mixture of types of goods which are transported together
19	Unidentifiable goods: goods which for any reason cannot be identified and therefore cannot be assigned to groups 01-16.
20	Other goods n.e.c.

NST/R classification

Code	Description
0	Agricultural products and live animals
1	Foodstuffs and animal fodder
2	Solid mineral fuels
3	Petroleum products
4	Ores and metal waste
5	Metal products
6	Crude and manufactured minerals, building materials
7	Fertilizers
8	Chemicals
9	Machinery, transport equipment, manufactured articles and miscellaneous articles

APPENDIX 3

NST2007 code	Description CFS 2009	CFS 2009 code
01	Spannmål	10
01	Potatis	11
01	Rundvirke	12
01	Obearbetad mjölk	13
01	Levande djur	14
01	Färsk fisk, färska fiskeriprodukter som tex. delar av fisk	15
01	Andra skogsråvaror än rundvirke, tex grenar, toppar, stubbar, skogsplantor	16
01	Andra råvaror av vegetabiliskt eller animaliskt ursprung tex färsk frukt/grönsaker, kryddor, levande växter och frön/blommor, oljeväxter, sockerbetor	17
02	Stenkol och brunkol	20
02	Råolja	21
02	Naturgas	22
03	Järnmalm	30
03	Annan malm än järnmalm	31
03	Kemiska och mineraliska (naturliga) gödningsmedel och salt	32
03	Jord, sten, grus och sand	80
03	Torv	83
04	Kött, köttprodukter, oberedda hudar och skinn	40
04	Beredd hållbarhetsbehandlad fisk och beredda fiskeriprodukter	41
04	Beredda och hållbarhetsbehandlade frukter, bär och köksväxter	42
04	Animaliska och vegetabiliska oljor och fetter	43
04	Mejerivaror, drycker och glass	44
04	Kvarnprodukter, beredda djurfoder, stärkelse och produkter därav	45
04	Övriga livsmedel och tobaksvaror	46
05	Textilvaror, kläder, pälsvaror, läder och lädervaror	100
06	Varor/produkter av trä och kork dock (ej möbler)	50
06	Papper, papp och varor därav	51
06	Pappersmassa	52
06	Tryckt och inspelad media	101
07	Stenkolsprodukter, t.ex. koks, koksbriketter	60
07	Flytande raffinerade petroleumprodukter, t.ex. bensin eldningsolja, och spillolja	
07	Gasformiga (som vätska eller komprimerade) petroleumprodukter	62
07	Fasta raffinerade petroleumprodukter	63
07	Etanol	64
07	FAME	65
08	Kemikalier, kemiska produkter, konstfibrer, gummi- och plastvaror, t.ex. plast i obearbetad form, kväveföreningar, handelsgödsel (ej naturgödsel) och syntetgummi	70
08	Läkemedel och fädigvaror från kemisk industri	102
08	Bearbetade gummi- och plastvaror	103
09	Cement, kalk och gips	81
09	Annat byggnadsmaterial (ej metall och trä) tex isolering och byggelement av betong	82

09	Glas och glasvaror, porslins- och keramiska produkter	104
10	Järn och stål t.ex. tackjärn, råstål, bandstål, valsat stål och järnlegeringar	90
10	Andra metaller än järn samt produkter därav	91
10	Byggnadsmetallvaror och rör, rörledningar, ihåliga profiler och tillbehör	92
10	Pannor, järnvaror, vapen och andra metallvaror (ej maskiner och apparater)	105
11	Jord- och skogsbruksmaskiner	107
11	Övriga maskiner samt apparater; kontorsmaskiner, maskinverktyg, maskindelar, datorer, radio, TV, elektriska maskiner, samt medicinska och optiska instrument	108
12	Transportmedel (-utrustning) t.ex. bilindustriprodukter, bilar, motorfordon och andra transportmedel	106
13	Möbler och övriga tillverkade varor	109
14	Hushållsavfall, annat avfall och returavfall t.ex. returpapper, skrot och rivningsmaterial	110
20	Varor ej tidigare uppräknat	111