

Report

VALUE OF FREIGHT TIME VARIABILITY REDUCTIONS

Results from a pilot study for the Swedish Transport Administration

Niklas Krüger and Inge Vierth (VTI)

Gerard de Jong (Significance, NL)

Askill Halse and Marit Killi (TØI, NO)

Table of Contents

1 Introduction

- 1.1 Short summary
- 1.2 Background and motivation
- 1.3 The scope of the project
- 1.4 The parts contained in this report
- 1.5 Comments on policy and future research

References

Part One: Evidence and implications for Swedish VTTV from recent SP in Netherlands and Norway

Part Two: New methods for estimating the value of freight time variability

1 Introduction

1.1 Summary

This project studies the calculation of the 'Value of Transport Time Variability' (VTTV) with regards to freight transports. Different approaches to calculate the cargo component of the VTTV (not the components related to staff and vehicle costs) are presented and the possibilities to use them in the Swedish context are discussed.

Part One presents the SP-studies to calculate VTTS ('Value of Transport Time Savings') and VTTV that have been carried out recently in The Netherlands and Norway. Both the Dutch VOTVOR-study (Significance et al, 2012) and the Norwegian GUNVOR-study (Halse et.al, 2010) comprise door to door transports for all modes and use the standard deviation of the transport time to express the variability in transport time. The PUSAM-study (Halse et al, 2012) is a follow up study of the GUNVOR-study with the intention to obtain better values for rail freight transports. The PUSAM study uses the expected delay between railways stations to express the reliability. VTTV-estimates could be derived although the response rates were low: in GUNVOR ca.6 percent and in PUSAM ca. 14 percent; the majority of the big forwarders were however represented.

Part One also discusses to what extent the findings in these SP-studies can be transferred to Sweden. The easiest way would be a *direct value transfer*. The freight reliability ratios (RR) express the importance of the variability of the transport time relative to the transport time. RR for road freight of 0.9 (in the Dutch VOTVOR-study) and 1.3 (in the Norwegian GUNVOR-study) could be transferred to Sweden. However, the mix of goods transported by road in Sweden needs to be taken into account; the value density is probably somewhat lower in Sweden than in Norway and The Netherlands. Also transport distances and the level of congestion differ and an empirical study in Sweden would be needed to get VTTV for Swedish road transports. An SP study would be an obvious candidate. Concerning rail, the choice between the two reliability measures (standard deviation or expected delays) should to a large extent depend on which measure can be implemented with the least effort in the Swedish transport model SAMGODS. The values for the RR are 0.8 in the Dutch VOTVOR-study and 1.8 in the Norwegian GUNVOR-study, whereas expected delay has a value of 72 NOK per tonne-hour in the Norwegian PUSAM-study. Also the VTTV for rail transports would need to be adjusted to the mix of goods transported and are probably lower for Sweden (and The Netherlands) with a high share of bulk transports with a

relatively low value than for Norway. As for road an empirical study for Sweden would be needed.

Another purpose of Part One is to *make use of the experience in The Netherlands and Norway*. Should Sweden decide to carry out a SP study, design and many other things can be learnt from the Norwegian and Dutch freight SP studies: features that they have in common can be used for Sweden as well and where the studies differ, one can choose the feature most appropriate for Sweden.

In Part Two three alternative approaches to calculate VTTV for Swedish rail freight transports are demonstrated, that to the best of our knowledge have not been implemented before.

1) The *precautionary cost approach* uses the hypothesis that a company reacts to a stochastic delivery time by taking precautionary measures which are a function of the standard deviation in transport time: $s=f(\sigma)$ ceteris paribus. Different types of precautionary measures can be applied; i.e. holding of safety stocks, using more expensive modes that are more reliable and localisation close to suppliers or customers. The cost for holding a safety stock and hence the cost of variability in transport time is the cost of physical storage of the goods and the capital costs of the goods stored. Hence we can compute the societal cost of variability under certain simplifying conditions. We show i) that the marginal precautionary costs measure marginal VTTV and that ii) a precautionary stock approach can in principle be made operational by aggregating all companies in Sweden with freight transport exposure, computing a virtual safety stock and using key aggregate figures about transport time variance, inventory costs, freight flows and required service level. Required service levels should be obtainable from companies since they are key figures used in practice. Further on other precautionary costs than inventory costs (i.e. for perishable fresh fruit or goods that cannot be stored as demand is not known in advance) need to be included in the approach. More research should be done on how to incorporate the extremeness of empirical delays that tentatively increase the firms' precautionary costs and VTTV.

2) The *case study approach* identifies the amount and type of the additional costs that (Swedish) shippers face due to the variability of the rail transport time with help of company cases. Within this pilot project the shuttle train run by the grocery company COOP is studied. We measure the degree of variability that the company faces during a 16-month period, and identify and estimate the precautionary costs COOP were willing to accept to manage (the costs of) transport time

variability and the additional operative costs that the company pays in case of major train delays or cancellations. We show i) that by doing a case study it is possible to get an estimate of VTTV valid for a specific company and that ii) in conclusion, given the high degree of market concentration with regard to shippers and forwarders, just a few case studies for key companies in the market might be sufficient to get a representative VTTV-measure in Sweden.

3) The *market based approach* is built on the hypothesis that publicly traded company stocks accurately reflect the steady stream of information and hence that delays for freight trains should have an effect on stock prices. We show i) that stock prices for companies that use rail transports react on train delays, ii) that changes in company market value per hour delay can be used as a VTTV-estimate given figures about rail freight flows (volume and variability of transport time) for a certain company and that iii) the method can be used to discern between costs for relatively small delays and the very large delays. In other words, the method has the potential to identify costs of transport system vulnerability not covered by VTTV. The approach presented in this pilot study needs to be developed by using information about quantities transported for the examined companies.

Table 1 summarizes the cargo-related VTTV for rail transports calculated with help of the different approaches. The values derived in the Dutch and Norwegian SP-studies (se Part One) are not adjusted to the Swedish commodity mix. In Part Two new approaches are tested, that make use of existing data. The computed VTTV have to be seen as preliminary values; the values are probably lower bounds as not all precautionary costs are included and as the market value approach does not take into account the train delays for specific companies. The importance of how to measure reliability is illustrated in the COOP-case where the VTTV based on the delays over one hour is higher than the VTTV based on the standard deviation. The VTTV that are recommended in the Swedish CBA-guidelines (Trafikverket, 2012) are included as comparison.

Table 1: VTTV for the cargo component in rail transports (SEK per tonne-hour)

	COOP:s goods	Consoli- dated goods	Other goods	Food and feed	High value goods	All goods
Part One						
Dutch VOTVOR-study, (standard deviation for door-to-door transports)						2
Norwegian GUNVOR study, (standard deviation for door-to-door transports)						50
Norwegian PUSAM study, (expected delay between railway stations)		317	40			82
Part Two						
Precautionary cost approach (standard deviation, inventory costs)						2 (prel.)
Case study approach (COOP, standard deviation for goods transported)	18 (prel.)					
Case study (COOP, precautionary costs per delay tonne hour)	34 (prel.)					
Market value approach (standard deviation)						2 (prel.)
Swedish ASEK 5-values (excl. VAT), based on commodity value				4	14	3

We show that the VTTV calculated in the Dutch and Norwegian SP-studies in principle can be transferred to Sweden. However, empirical studies that are adjust with respect to the commodity mix, the transport distances, level of congestion etc. in Sweden are needed. When it comes to rail the question how to handle early arrivals (that are included in the standard deviation but not in the expected delay) needs to be addressed. The question of excessive slack is also important from a policy point of view. Another question is how the variation in transport time for empty trains should be taken into account. If Sweden decides to carry out a SP-study to calculate VTTV (and VTT'S) it is possible to make use of the Dutch and Norwegian experience.

We also demonstrate the use of three approaches using crude figures as an input. We advocate further research on getting more realistic inputs. Moreover, the theoretical and empirical methods should be developed. Also, more research should be done on how to incorporate the extremeness of empirical delays in models and definitions of VTTV and how different policy measures can reduce the transport time variability. Last but not least, new methods, for example based on micro-level data on company inventories, the standard SP-method and the three new

approaches suggested here, should be used in combination with in order to validate VTTV-estimates.

1.2 Background and motivation

There is to date limited knowledge about the impact of different policy measures and the monetary valuation of improved reliability for freight transports. This means that the benefits of reduced variability in transport time are not taken into account properly in cost-benefit analysis (CBA), for example related to infrastructure measures. The ‘Values of Transport Time Variability’ (VTTV) are in comparison to the ‘Values of Transport Time Savings’ (VTTS) hardly addressed in cost benefit analysis despite the intentions in several countries (see i.e. OECD/ITF, 2009).

The Swedish CBA-guidelines recommend provisional commodity specific VTTV for the cargo component. These values are assumed to be two times the VTTS and expressed in SEK per tonne-hour (Trafikverket, 2012). The benefits for the vehicle and staff component are assumed to be part of the transport costs and quantified elsewhere in the CBA. Definitions of the VTTS and VTTV and the practise in relation to CBA differ between countries; for an overview see i.e. Vierth, (2010). The Dutch values refer to the cargo as well as the staff and vehicle costs providing transport services since this is how CBA in The Netherlands is used. The Norwegian and Swedish VTTS and VTTV relate only to the cargo related costs. In this study we focus on the VTTV for the cargo component.

Unreliable rail transports are the main contributor to the transport time variability. In 2009 the Swedish Transport Administration registered around 80 000 delay hours for freight trains, see (Krüger, Vierth & Fakhraei Roudsari, 2013). This corresponds roughly to 40 million tonne-delay-hours, given that 500 tonnes per train is assumed to be the average load. This figure is 20 times higher than the corresponding figure for road transports. The Swedish Transport Administration registered circa 1,300,000 road vehicle-delay-hours due to unplanned stops over five minutes (and at least for all lanes in one direction) in 2010 (Trafikverket, 2013). The number of truck-delay-hours is ca. 200 000 hours (assuming that 15 per cent of the vehicles are trucks) and the number of the “tonne-delay-hours” is ca. two million (assuming a load of ten tonnes per truck). The delays for the sea and air transports are probably negligible.

The information about the reliability of the rail transports and how different policy measures influence i.e. the number and length of delays and or the risks for delays has been limited. The former has been improved in recent years but the latter is still a problem. The lack of underlying

data is one reason why the recommended VTTV have only been applied in a few rail infrastructure projects (Vierth & Nyström, 2013). Nevertheless, the industry complains about the extensive problems caused by train delays or cancellations and that the reduction of these problems is not taken into account in a proper way in the CBA. One example for such a problem is the derailment in Grötingen in January 2011 that led to a two day stop of the SSAB steel plant in Borlänge, additional transport costs of SEK 1.5 million and lost/delayed sales of SEK 60 million. Another example is the derailment and destruction of a seven km single track in Jutland/Denmark in November 2012. The track was blocked and circa 60 freight trains to/from Sweden had to find alternative routes and modes during a period of more than two weeks. ScandFibre Logistics claimed to have additional transport costs of circa SEK 20 million to carry paper (worth ca. one billion SEK) to mainland Europe. Another question that is debated is how much more high value products would be transported by rail if the rail transports would be more reliable. According to the latest Swedish Commodity Flow Survey 2009 (Trafikanalys, 2010), the value of the outgoing goods transported by road (SEK 9.5 per kg) is about three times the value of the outgoing goods transported by rail or rail in combination with other modes (SEK 3.7 per kg).

The Transport Administration and the work group for the CBA-guidelines in transport (ASEK) are aware of the shortcomings in the existing guidelines and routines and have among others asked VTI to suggest research (Vierth, 2010; Vierth, 2012). In the beginning of 2013 the Transport Administration has commissioned WSP and VTI to perform pilot-studies addressing the valuation of freight time variability reductions. These pilots shall serve as basis for further decisions.

1.3 The scope of the pilot project

Starting point for this pilot project are the research and development activities that VTI has recommended in the pre-studies concerning the calculation of VTTV in 2010 and 2012. The pilot study is used to bring the proposed activities a step further.

The CBA guidelines need to cover VTTV for all transports in, to/from and through Sweden and all modes. In this pilot project we particularly focus on transport chains that include rail as the rail mode causes most of the reliability problems for the Swedish freight transports.

1.4 The parts contained in this report

The project contains two main parts:

- Part One: SP studies from Norway and The Netherlands and their implications for Swedish freight VTTV
- Part Two: Three other approaches to calculate the Swedish freight VTTV are demonstrated: a precautionary cost approach, a case study approach and a market based approach.

1.5 Comments on policy and future research

The starting point for our comments on policy and future research are the recommendations in VTI's pre-studies from 2010 and 2012¹ See Table 2 below. Lastly we would like to stress that use of the VTTV in cost benefit analysis requires information about the causes for variation in transport time², how delays, cancellations and early arrivals are distributed over the network, the dispersion of delays and how different policy measures influence the transport time and the variation in transport time. Ideally the variability of the transport time should be included in the Samgods model (as the transport time). Given that also information about the stock out costs is available the buffer stock approach could be integrated in the Samgods model.

Within this project we will based on the final seminar:

- finalise Introduction and Part One and publish it as a 'VTI-Notat'
- finalise Part Two and publish it as a CTS-working paper on Swopec.

When it comes to further developments, we need to be sure what to value and how to use the VTTV in CBA. According to the fact that most of the delays for the Swedish freight transports are related to rail transports, there is a need to improve a) the knowledge about transport time variability in transport chains that include rail and b) the firms' valuation of (the lack of) reliability. The following questions need to be covered in research and development projects:

- How should transport time variability be expressed and measured? - i.e. inclusion of early arrivals and slack or not, handling of large versus small delays (the standard deviation is

¹ The remaining recommendations are not that directly related to the estimations of the VTTV: improvement of the knowledge of the less well studied opportunity costs, a clarification of the whole CBA structure for freight value of time and reliability as well as an application of the logistics model within the Samgods model with its existing mechanisms for the development of transport demand and costs for transport resources

² In the COOP case ca. 50 per cent of the delays are related to the infrastructure

not that relevant in the COOP case), risks for very large delays versus vulnerability of the transport system

- How does the variation in transport time influence companies? - thresholds for delays and early arrivals could be derived with SP-study, companies' costs related to average costs in case of (major) delays are higher than the costs for average delays
- How do different policy measures influence transport time variability? - i.e. analysis of causes for delays, evaluation of impacts of measures that have been applied in recent years, identification of "weak links" in the rail network
- What are the possibilities to transfer the VITV derived in the Dutch and Norwegian SP-studies to Sweden - empirical studies of Swedish freight transports and their conditions compared to Norwegian and Dutch transports
- How do share of rail transports, service levels, type of precautionary costs, backups etc. differ between companies from different sectors? - Collection of information from large shippers and forwarders in order to extend the approaches described in Part Two. One aspect is the separation of variation of transport time and variation of demand. Another aspect is to what extent could less variability in transport time contribute to a higher share of rail transports.

Table 2: Activities to estimate VTTV in earlier pre-studies, this pilot study and recommended as next steps

	Recommendations in VTI:s pre-studies	Covered in this pilot study	Recommended future research
Part One	to conduct thorough assessments of the Norwegian SP-study and the Dutch SP-study	The Dutch VOTVOR-study, the Norwegian studies GUNVOR and PUSAM have been analysed. Results: VTTV are mainly relevant for shippers and forwarders and vary over commodities. Value transfer is in principle possible. Sweden can learn from the experience in The Netherlands and Norway	For value transfers adjustments and further empirical studies are needed For rail there is also a need to agree on how to express variability
Part two (a)	to develop mechanisms and data that are missing in the logistics model outside the model	Precautionary approach for all Swedish transports (that are included in the Samgods model) For simplicity reasons standard deviation and normality are assumed and only buffer stock costs included as precautionary costs.	The implication of the heavy tails in the distribution of the delays and early arrivals needs to be studied More information about the required service levels is needed. A limited SP-study could be carried out Different types of precautionary costs need to be included
Part Two (b)	to conduct RP-studies	COOP case study	More case studies should be performed – preferably for large shippers and forwarders (to cover a great share of the market)
Part Two (c)	To study further alternative approaches (in addition to SP studies).	Market based approach	Approach needs to be developed taking into account specific companies´ rail transports

References

- Halse, A. H., Samstad, H., Killi, M., Flügel, S., & Ramjerdi, F. (2010). *Verdsetting av framføringstid og pålitelighet i godstransport*. Oslo: Transportøkonomisk institutt.
- Halse, A. H.; Killi, M. (2012). *Verdsetting av tid og pålitelighet for godstransport på jernbane*. Oslo: TØI (TØI rapport 1189/2012).
- Krüger, N.A.; Vierth, I; Fakhraci, F.R. (2013). Spatial, Temporal and Size Distribution of Freight Train Delays: Evidence from Sweden. *Forthcoming CTS working paper*.
- OECD/ITF. (2009). *Improving Reliability on Surface Transports*. Paris: OECD/ITF.
- Significance, VU University, John Bates Services, TNO, NEA, TNS NIPO and PanelClix . (2012). *Values of time and reliability in passenger and freight transport in The Netherlands, Report for the Ministry of Infrastructure and the Environment*. The Hague: Significance.
- Trafikanalys. (2010). *Varuflödesundersökningen 2009*. Östersund: Trafikanalys (Statistik 2010:16).
- Trafikverket. (2012). *Samhällesekonomiska principer och kalkylvärden för transportsektorn: ASEK 5, Kapitel 8, Tid och kvalitet i godstrafik*. Borlänge: Trafikverket (Version 2012-05-16).
- Trafikverket. (2013, 6 19). Statistik over totalstopp. Trafikverket (Johansson, Arne).
- Vierth, I. (2010). *Värdering av minskad transporttid och minskad variation i transporttid för godstransporter - Förstudie*. Stockholm: VTI (VTI-rapport 683/2010).
- Vierth, I. (2012). *Värdering av tidsvinster och högre tillförlitlighet för godstransporter*. VTI (VTI Notat 24-2012).
- Vierth, I., & Nyström, J. (2013). *Godstransporter och samhällesekonomiska kalkyler*. Stockholm: VTI (VTI Notat 2013-13).

Part One

Evidence and Implications for Swedish VTTV from recent SP in Netherlands and Norway

SP studies from Norway and The Netherlands and their implications for Swedish freight VTTV

1 Objectives of the studies

In both Norway and The Netherlands Stated Preference (SP) studies in freight transport have been carried out recently (2009-2013) that provide – among other things- monetary values for reliability in freight transport for use in project appraisal.

In Norway this actually concerns two studies (besides the separate Norwegian passenger value of time and reliability study) :

- The multimodal GUNVOR project that was carried out to gain more insight into the valuation of freight reliability and develop methods to assess this value using SP studies as well as to obtain unit values for transport time and reliability in freight for application in cost-benefit analysis (CBA).
- The rail-based PUSAM project that aims at improving rail transport reliability by developing decision support tools, including an SP study on the value of time and reliability in rail freight transport.

The Dutch project, VOTVOR, was carried out to establish values of time and reliability for all modes in freight transport for use in CBA. The same VOTVOR study also treated passenger values of time and reliability.

In Norway, these were the first freight SP studies of this kind; in The Netherlands the study replaces older SP-based values of freight transport time savings and adds values of reliability to this.

In Annex 1 and 2 the two Norwegian SP freight studies and the Dutch one are described in more detail. These annexes also include references to the original material.

2 Definition of reliability

2.1 Reliability in the model

In the models estimated on the SP data and in the values recommended for use in CBA that were derived from those models, reliability is defined as the standard deviation of the transport time distribution (though scheduling terms were sometimes tried in the modelling as well). The same definition was used for the recommended values in passenger transport. The main reason for choosing this definition was that transport models are needed to supply quantity changes in reliability, and that the standard deviation is relatively easy to integrate in these models.

An exception to this rule is rail freight transport in Norway, where the recommended values for reliability are in terms of expected delays.

2.2 Reliability as presented to the respondents in the survey

Since many respondents would not understand the concept of standard deviations, the presentation of reliability to the respondents in the SP experiments is different. GUNVOR,

PUSAM and VOTVOR include at least one choice experiment where reliability is presented within a single choice alternative as a series of five transport times that are all equally likely to happen. These are presented verbally not graphically, which worked best in extensive pilot surveys in The Netherlands.

GUNVOR also used a presentation format with a certain delivery in one of the two alternatives on a screen and a length of the delay with some probability in the other alternative. This format was also used in PUSAM.

The Dutch study provides values of time and reliability that refer to the transport personnel and vehicle costs of providing transport services as well as the cargo-related values (e.g. interest costs on the goods in transit, disruption of the production process due to missing inputs), since this is how the CBA for transport projects in The Netherlands works. The recommend values from the Norwegian study refer to the cargo component only. For CBA in Sweden the recommended values should also only relate to the cargo component, not the transport services. The Dutch results can however be split into both components (and in fact, for the value of reliability, the transport service component was not significant and only the cargo component matters). GUNVOR and VOTVOR are about door-to-door transports and consequently about transport time and its variation at the receiver of the goods. PUSAM on the other hand deals with transport time between railway terminals only.

3 The SP survey

3.1 Why use SP data here?

In the survey design stage in both Norway and the Netherlands, other approaches than SP were considered. Revealed Preference (RP) surveys were not chosen as the main data base, because it is difficult to get RP data where time, costs and reliability are not heavily correlated and where there is still sufficient variation in these variables (that should relate to chosen and non-chosen alternatives) for estimation.

The buffer stock approach (or: 'logistics costs function' approach) was mentioned in both countries, but not selected because no examples were known of empirical applications for estimating the value of reliability using this approach.

3.2 Design of the SP survey

In GUNVOR, PUSAM and VOTVOR the questionnaires first asked questions about the firm and then about a specific transport/shipment that was carried out in practice by or for the firm. The attribute values are based on the reported attribute values for this reference transport/shipment, which can be considered good practice in SP survey design.

Then follows a series of SP choice experiments. All of these use binary choices between choice alternatives that both refer to the same mode. GUNVOR, PUSAM and VOTVOR first have an experiment with only two attributes: transport time and transport cost. This is only relevant for the VTTS.

After that all three surveys continue with an experiment with transport costs and with reliability in the form of five equi-probable transport times. These experiments are similar but not quite the same:

- The Dutch experiment also contains as a separate attribute presented in the SP the usual transport time; the Norwegian GUNVOR study left it out (showing that it is not needed) and inferred the mean transport time from the series of five transport times.
- The Dutch study also presents departure time and five arrival times (corresponding with the five transport times and the departure time), so that the respondents can easily see the

scheduling consequences of delays. In one experiment the most likely arrival time is fixed, in another it varies.

- The statistical design used is different.

Whereas the third and final Dutch SP experiment is a variant of the second (initially the most likely arrival time varies, later it is fixed), the two Norwegian studies use a different format with risk of a specific delay instead of the five transport times.

The Dutch study has 19 SP choice situations in total (6+6+7), the two Norwegian ones have 20 (8+6+6). The Norwegian studies always present the reference cost and time in one of the alternatives; in VOTVOR this is not necessary. VOTVOR and GUNVOR present both early and late arrivals, but in PUSAM only late arrival is considered.

In the Dutch survey, there were specific experiments for carriers in sea and inland waterway transport that did not use the context of a door-to-door transport, but the context of waiting for a lock/bridge or of waiting to be loaded/unloaded at a quay in the port.

3.3. Who are the respondents?

GUNVOR and VOTVOR were targetted at:

- Shippers that contract transport out
- Shippers with own account transport
- Companies providing transport services, such as carriers.

PUSAM only looked at customers (shippers and transport companies that act on behalf of the shippers, but do not operate the trains) of the rail operator CargoNet.

Shippers are in the best position to provide the components of the value of time and reliability that are related to the goods themselves, whereas carriers have the best knowledge to supply the transport services components. In the Dutch questionnaire, the shippers that contract out were specifically asked to only consider the aspects related to the goods. Similarly the carriers were specifically asked to only think about the transport services, not about the goods. This set-up helped to obtain values with a clearer interpretation than previous surveys (including older Dutch SP surveys) that were ambiguous on this.

3.4 Recruitment and interview method

The shippers and carriers in GUNVOR and VOTVOR were recruited from various national company registers. The PUSAM respondents were taken from the customer data base of CargoNet. The Norwegian surveys approached the firms by email; VOTVOR used approach by phone.

The VOTVOR interviews were carried out at the offices of the firm by professional interviewers as computer-assisted personal interviews (CAPI). This interview method has good possibilities for explaining the questionnaire and motivating the respondent, but is expensive. GUNVOR and PUSAM were carried out online, which is considerably less expensive.

3.5 Sample size obtained

GUNVOR obtained responses from 117 transport firms and 640 shippers (including those that contract out and those that do own account transport), whereas VOTVOR had 315 carriers and 497 shippers as respondents.

These sample sizes are, in as far as we are aware, the largest ever achieved in SP research in freight transport. Nevertheless, compared to the SP sample sizes that are common in passenger transport, these are rather small samples and many of the more sophisticated models that are used in the analysis of passenger transport SP data cannot be supported by these freight SP data sets.

GUNVOR has a majority of observations from road transport (almost 80%), whereas VOTVOR has just over 50% for road transport, but also many respondents for inland waterways and sea transport (not so many for air and 50 respondents for rail, of which 35 shippers). The GUNVOR data contains 42 respondents in the shipper segment which have used rail transport for their shipment, making it possible to study this mode separately.

PUSAM by definition only has respondents (34 in total) that use rail transport. Most are forwarders or consolidators, some are shippers.

4 Analysis of the SP survey

4.1 Data checks

All three data sets were checked for outliers and missing values, and these respondents were removed before estimation. In GUNVOR also respondents who had answered the questionnaire in less than 10 minutes were discarded (such situations did not occur in VOTVOR or PUSAM). Respondents that do not involve in trade-offs between the attributes were kept in the estimation sample in all three surveys. But in the Norwegian studies attributes that are ignored by the respondent are eliminated from the model. These surveys asked the respondents at the end which attributes they had considered, and this information is used to identify attribute non-attendance. Such questions also appeared at the end of the VOTVOR questionnaire, but were not used in estimation.

4.2 Model specifications

A typical feature of freight transport is the large degree of heterogeneity, for instance in the time and costs attributes. This needs to be taken into account in the analysis.

The recommended values from GUNVOR, PUSAM and VOTVOR all come from multinomial logit models (MNL). More sophisticated models (e.g. mixed logit, latent class), as used in the Norwegian and Dutch passenger SP surveys, were not successful and stable when estimated on the freight SP data.

The number of interaction variables to explain heterogeneity in the coefficients for time and reliability in these models remained very limited.

The chosen MNL models in GUNVOR use a multiplicative error specification, as do the models in VOTVOR for carriers in road transport (the latter is a log-willingness-to-pay space model). All Norwegian freight models use preference (utility) space. The chosen MNL models in PUSAM and those for all other carriers and for shippers in VOTVOR use an additive error term (and are formulated in preference space). These VOTVOR models all use a relative specification (all attributes measured relative to its reference value), which is one way of dealing with heterogeneity. This specification can only provide values of time and reliability when combined with information on the transport costs per hour.

In order to correct for repeated measurements (multiple SP choice observations on the same respondent), the Norwegian studies specify the user ID as a panel variable, whereas VOTVOR used the Jackknife method to correct for the possible bias (especially in the t-ratios).

In VOTVOR, joint models (with scaling factors) were estimated on all three SP experiments together. The Norwegian studies have different models for each SP experiment. The chosen specification in GUNVOR and VOTVOR includes reliability in the form of the standard error (besides time and cost). The Norwegian studies have an alternative specification, used for the data from the third experiment that includes the expected delay as reliability variable, which is the preferred definition for rail transport (also from PUSAM) for the Norwegian studies.

4.3 Results

Below we focus on results for the VTTV (the various studies also give outcomes for the VTTS) for shippers (cargo component only), which is the appropriate VTTV for the Swedish CBA context (though this does not guarantee transferability). The reliability ratio in the table below is the ratio of the VTTV (using the standard deviation) to the VTTS.

Table 1 Comparison of outcomes for the VTTV (as the reliability ratio) from GUNVOR and VOTVOR

	Reliability ratio road transport; Cargo component (shipper) only	Reliability ratio rail transport; Cargo component (shipper) only
GUNVOR (from 2 nd experiment)	1.3	1.8
VOTVOR (from all 3 experiments; weighted average)	0.9	0.8

We see that the reliability ratios (RRs) for the cargo component only (as given by the shippers) is reasonably similar, with somewhat lower values for The Netherlands. For the carriers VTTV, both GUNVOR and VOTVOR found estimated coefficients for the standard deviation that were not significantly different from zero. (However, the valuation of delays in the third experiment in GUNVOR was somewhat higher than the VTTS.) If we would add the carriers VTTV of 0 to the shippers VTTV and divide this sum by the summed VTTS of the shipper and the carrier (where the carrier component is substantially larger than the shipper component), we get an overall RR that is much smaller than 1 (0.1 to 0.4).

The Norwegian team has so far recommended to use the PUSAM results for rail. The key result for the VTTV here is the value of expected delay, which for rail (weighted average) is 72 NOK per tonne-hour.

5 Current practice

5.1 What is used for VTTV (P-side of reliability)?

The current practice of CBA of transport projects in Norway is reported in the handbooks by the Norwegian Public Roads Administration and the Norwegian National Rail Administration. The latter has recently been revised, while the former is due for revision.

In road transport, only known transport time savings are valued in the current practice and there is no cargo component in the VTTS. In rail transport, delays are given a higher weight than changes in known transport time both for passengers and freight. The current values of transport time savings and delays for freight are taken from the Norwegian freight model. These are considerably lower than those in the results reported in this document, and the National Rail Administration is considering replacing them with the new values.

We expect that the new Dutch VTTs and VTTVs will become the official transport CBA values in August 2013. At the moment, reliability benefits in The Netherlands are usually calculated as simply 25% of the travel time (passengers) or transport time (freight) benefits. More

differentiated, though still preliminary, guidelines for the VTTV have been available since 2004/2005.

5.2 Variability forecasts (Q-side of reliability)

Concerning the prediction of variability, this has not been given very much attention in Norway. In the case of road transport, the current practice is as mentioned not to value changes in variability. In rail transport, the CBA tool of the National Rail Administration contains a formula which estimates the amount of delays in hours based on the percentage of trains which are late. However, since the percentage late trains is not something which is estimated in the transport models normally used, this figure has to be based on some analysis which is specific to the project. Furthermore, for passenger trains, delays are assumed to be the same for passengers getting off at all stations on the line.

The more detailed preliminary VTTVs in The Netherlands referred to in section 2.5.1 have been used in very few studies, because of the difficulty to predict changes in reliability and the impact of a project on reliability (only some prototype forecasting models are available to do this). Most project appraisals have used the 25% surcharge on the time benefits for the reliability gains. The new VTTVs can also only be used in conjunction with forecasts of how infrastructure projects influence variability (in The Netherlands this is called the 'Q-side of reliability'). The development of such models for passenger and freight transport is planned, but still needs to be carried out. Some models explaining variability of transport time in road transport (passengers and freight) have already been estimated in The Netherlands on speed data from induction loop measurements on motorways.

5.3 Swedish freight SP experience

Before moving to the lessons that Sweden might learn from the Norwegian and Dutch experience from freight SP surveys that include reliability, it is good to look at freight SP surveys carried out in Sweden in the past.

The SP surveys carried out by Staffan Widlert (who was in close contact at the time with Hague Consulting Group that was doing a similar study in the Netherlands) in the early 90ties (Transek (1990) for road and rail freight transport and Transek (1992) for road freight transport) focused on the VTTTS. However an attribute referring to reliability was also included in the SP: the frequency of shipments arriving late (either on the same day or the next day). The outcomes for this definition of reliability are hard to translate to a value of the standard deviation of transport time or a reliability ratio. Both studies interviewed shippers (so in our interpretation the results can best be interpreted as for the cargo component only) and used face-to-face interviews. The models estimated are MNL model using a relative specification (attributes levels relative to those of the reference shipment).

The data from the 1992 Transek study have later been analysed by Erik Bergkvist, using different models (absolute levels for the attributes instead of relative) and estimation methods (weighted exogenous maximum likelihood). See for instance Bergkvist and Westin (2000).

The SP study by Inregia that also involved Mogens Fosgerau from Denmark (Inregia, 2001) also focused on the VTTTS (for road, sea, rail and air transport), but included a measure of reliability in the SP. This measure was presented as the fraction of shipments (how many in a 1000 shipments) that is delayed. The respondents were shippers, so again the results should probably be interpreted as referring to the cargo component only. The interviews were done by phone. An MNL model (amongst other models) was estimated and it gives a value for the risk of delay expressed as a change of 1 per 1000 shipments. It is difficult to translate the outcomes for this

definition of reliability into a value of the standard deviation of transport time or a reliability ratio.

Xing Liu estimates VTTS for four commodity groups based on data from the Swedish Commodity Flow Survey (CFS 2001) and costs information from the Swedish national freight transport model Samgods as part of her PhD at Örebro University (which is planned to be finished in 2013).

5.4 Implications for the Swedish VTTV

The recent Norwegian and Dutch freight SP studies have shown that it is possible to derive plausible monetary values for the cargo-related component of reliability measured as the standard deviation of transport time by means of SP interviews with shippers and models estimated on these data. The finding in these studies that the transport services component of reliability should be equal to zero is somewhat unexpected (if reliability increases this would increase the predictability of the deployment of transport vehicles and staff, which should have some positive value for the carrier). On the other hand, we expect that this value will be rather low and that the shipper component in reliability would dominate the picture (the reverse picture as for the full VTTS).

The easiest way for Sweden to use these findings would be a direct value transfer from these two countries to Sweden. If this would not be considered sufficiently reliable, the SP studies carried out so far in freight transport and especially recently in Norway and The Netherlands provide guidance on how a freight SP study can be carried out in Sweden. Below we discuss both options, one by one.

Direct value transfer

Road

In line with the practical recommendations from the Norwegian and Dutch SP studies and the German feasibility study (Significance et al. 2012b), reliability in road transport can best be expressed in the form of the standard deviation. What is required then for valuation of reliability is a direct money value for the standard deviation or a reliability ratio (in the latter case the monetary value can be derived using the VTTS).

Current overview studies on the VTTS in freight transport (Zamparini and Reggiani, 2007; de Jong, 2008; Feo-Valero et al., 2011) contain many results for the VTTS, but not for the VTTV. The European Project HEATCO recommended using a reliability ratio for freight transport of 1.2 (note that this refers to the sum of the cargo and transport services component). This ratio was not directly based on empirical research but came from an international expert workshop convened at Schiphol airport and reported in Hamer et al. (2005). Most of the freight SP research before 2007, that included some measure of reliability, used the probability of late delivery for this. Translating such results into a reliability ratio is very hard and requires many assumptions (de Jong et., 2009).

Outcomes for the cargo component for road transport in terms of a reliability ratio have been obtained in the Norwegian GUNVOR study and the Dutch VOTVOR study. Fowkes (2006) also obtained reliability ratios for the UK. These refer to the sum of the cargo and the transport services component and at that level his RRs are broadly comparable to the Norwegian and Dutch results. So the road freight RR (cargo component only) that might be transferred to Sweden should be based on the RR of 1.3 for Norway and 0.9 for The Netherlands. The mix of types of goods transported by road in Sweden has a more bulky character (and lower value densities) than in Norway and The Netherlands. This makes direct value transfer a risky prospect.

All one can say is that a value at the lower end or slightly below these values (say 0.9, 0.8) would be rather plausible. To get a value for the Swedish road transport, one needs to do an empirical study in Sweden, and an SP study would be an obvious candidate.

Rail

For rail freight the Norwegian study (PUSAM) has provided a recommended value of expected delay to be used within the current CBA framework of the National Rail Administration. It is however possible to derive an RR by combining the results from both studies. In the Netherlands the recommendation is to use the standard deviation also for rail freight. The choice between the two measures for Sweden should to a large extent depend on the question which measure can with the least effort be implemented in the Swedish transport forecasting models.

Again, as far as we are aware, the only potentially transferable values for the RR (using the standard deviation) for the cargo component in the rail VITTV come from the Norwegian GUNVOR study (1.8 and the Dutch VOTVOR study (0.8).

Alternatively for the expected delay, the preferred study in Norway for rail is PUSAM, that obtained a value of 72 NOK per tonne-hour. Rail freight in Norway contains a large share of general cargo (in containers). In Sweden there is more focus in rail transport on bulk products. The same is true for The Netherlands, but these are often different bulk products (e.g. oil products, waste) than in Sweden. An RR for Sweden close to the Dutch value of 0.8 would therefore be plausible, but empirical work in Sweden itself would be required to obtain a value one could have more confidence in. The value of 72 NOK might be transferred as well, but a somewhat lower value for Sweden is not unlikely.

For all value transfers a good idea is to do sensitivity analyses: carry out the CBA for a range of VITTVs around the most likely value.

Implications for the design of an SP freight survey in Sweden

For obtaining values for the transport service component of the VTTS, SP studies are not strictly needed: these values can also be derived from the transport costs calculations (assuming that in the long run all staff and vehicle costs are time-varying). The cargo component of the VTTS could be calculated on the basis of interest cost calculations, but it is likely that the cargo-related VTTS will contain more components than just capital costs (such as deterioration of the goods, disruptions of the production process or being unable to serve demand due to lack of stock). An SP survey among shippers then is a feasible way to find this component (more extended cost functions, preferably estimated on RP data, could be another).

With regards to the VITTV, a considerable simplification can be achieved by assuming that the transport service component of the VITTV equals zero and that there only is a cargo component that can be identified from the behaviour of the shippers. In the Norwegian and Dutch SP studies, the transport service component was **not found to be significant** and it is not likely to be a large component.

For identifying the VITTV, the standard cost functions that are used in transport models are not helpful as they do not vary with reliability. One can also try extended RP-based logistics costs functions (including buffer stocks, see chapter 3) for this, or launch an SP survey among shippers.

So the purpose of a possible SP freight survey for Sweden would be twofold: to obtain the cargo component in both the VTTS and the VITTV.

Should Sweden decide to carry out a freight SP study, many things can be learnt from the Norwegian and Dutch freight SP studies: aspects that they have in common can be used again and where the studies differ one can try to choose the feature most appropriate for Sweden and so have the best of both worlds.

Recommended features of such an SP survey would be:

- Base all SP experiments on the attribute values of a transport actually carried out for/by the shipper.
- Do binary choice within-mode experiments.
- Start with an experiment with time and costs only, then do an experiment in which reliability is added, presented in the form of a series of equally likely transport times (also present departure and possible arrival time). Mean or most common transport time does not have to be presented separately in this experiment. A third experiment could use the length of delay with some probability versus a certain delivery time.
- Explain to the shippers that they should only take into consideration the implications for the cargo itself.
- Include questions about attribute attendance at the end of the questionnaire, so that these can be used in the modelling.
- Collect data for at least a few hundred shippers.
- Sample firms from company registers, recruit by phone, confirm by email and interview on-line (this is much less expensive than CAPI, and has proven to produce credible results).
- Test different MNL model specifications in the estimation phase (absolute versus relative models, additive versus multiplicative error term models, utility space versus willingness-to-pay space).
- Combine data from several experiments in the same model, by a logit scaling approach
- Correct for repeated measurements by using a panel specification in the model.

References

- Bergkvist, E. and L. Westin (2000) Regional valuation of infrastructure and transport attributes in Swedish road freight, Umeå Economic Studies No. 546, Umeå.
- Feo-Valero, M., L. Garcia-Menendez and R. Garrido-Hidalgo (2011) Valuing freight transport time using transport demand modelling: a bibliographical review, *Transport Reviews*, 201, 1-27.
- Fowkes, A.S. (2006) The design and interpretation of freight stated preference experiments seeking to elicit behavioural valuations of journey attributes, ITS, University of Leeds.
- Hamer, R., G.C. De Jong, and E.P Kroes. (2005) The value of reliability in Transport – Provisional values for the Netherlands based on expert opinion, RAND Technical Report Series, TR-240-AVV, Netherlands.
- HEATCO (2006) Developing Harmonised European Approaches for Transport Costing and Project Assessment, Deliverable 5, Proposal for harmonized guidelines. IER, University of Stuttgart.
- Inregia (2001) Inregia (2001) Tidsvärden och transportkvalitet, Inregia's studie av tidsvärden och transportkvalitet för godstransporter 1999. Background report of SAMPLAN 2001:1, Stockholm.
- Jong, G.C. de (2008) Value of freight travel-time savings, revised and extended chapter for Handbooks in Transport, Volume 1: *Handbook of Transport Modelling* (Eds: D.A. Hensher and K.J. Button), Elsevier.
- Significance, Goudappel Coffeng and NEA (2012b) Erfassung des Indikators Zuverlässigkeit des Verkehrsablaufs im Bewertungsverfahren der Bundesverkehrswegeplanung: Schlussbericht, Report for BMVBS, Significance, The Hague (see: <http://www.bmvbs.de/SharedDocs/DE/Artikel/UI/bundesverkehrswegeplan-2015-methodische-weiterentwicklung-und-forschungsvorhaben.html>).
- Transek (1990) Godskunders värderingar, Banverket Rapport 9 1990:2.
- Transek (1992) Godskunders transportmedelsval, VV 1992:5.
- Zamparini, L. and A. Reggiani (2007) Freight transport and the value of travel time savings: a meta-analysis of empirical studies, *Transport Reviews*, 27-5, 621-636.

Annexes: SP-studies in Norway and The Netherlands

A) Norwegian studies on the value of freight time variability

B) VTTV in the recent national stated preference study on values of time and reliability in freight transport in The Netherlands

Annex A) Norwegian studies on the value of freight time variability

This working paper is written as a part of a joint project between VTI, TØI and Significance. In this pilot project we develop methods to value reductions in freight time variability for rail freight in Sweden. One of the inputs consists of the results from the stated preference (SP) studies in the Netherlands and Norway, which can be compared with each other, with the values derived using the buffer stock approach and with the experiences from the case studies in this project.

TØI has conducted two recent SP studies on the valuation of transport time and variability in freight – one including all transport modes (GUNVOR) and one targeted directly at railway freight. In this working paper we cover both studies. The latter study was part of a project (PUSAM) which also consists of other parts, specifically the development of a web-based decision support tool visualizing rich data on railway reliability. In the following we refer to the two SP studies as the GUNVOR study and the PUSAM study.

1 Objectives of the study

The first of the two recent SP studies on freight in Norway was conducted as a part of the research project GUNVOR¹, which was granted to TØI in 2008 by the Research Council of Norway with co-funding by the Norwegian Public Roads Administration. The stated objectives of the study were (1) to gain more insight into the valuation of reliability in freight and develop methods to assess the value using SP studies (2) to obtain actual unit values representing the values of transport time savings and reliability which could later be applied in cost-benefit analysis. The SP survey was conducted in 2010 and the results reported later the same year. All modes of transport were considered.

The second SP study represented one of the work packages in the project PUSAM², which was launched in 2010 and finished in 2013. This project was also funded by the Research Council of Norway, but as an innovation project fostering cooperation between research institutes, public agencies and businesses. The project partners were the Norwegian National Rail Administration, the rail operators CargoNet, NSB and Flytoget, the research institutes SINTEF and TØI and the Norwegian University of Science and Technology (NTNU).

The aim of PUSAM was as to improve railway reliability through developing decision-support tools based on socioeconomic calculations. The tools were meant to support decisions on all levels of railway operation and management, not just infrastructure investment decisions. The two main contributions of the project are (A) a web-based software which visualizes statistics on the reliability level in the railway network and (B) an SP study on the values of transport time savings and reliability in railway freight, followed by supplementary analysis and recommendations how to apply the values. We consider both how to apply the values in the decision-support software developed in PUSAM and in traditional cost-benefit analysis.

¹ Godstransport og Usikkerhet, Norsk Value Of Reliability

² PUnktlighetsforbedring for godstrafikk på bane gjennom beslutningsstøttesystem basert på SAMfunnsøkonomiske kostnader

Our decision to conduct a new SP study in PUSAM was based both by the fact that we had gained more experience in designing such a study, that the survey could be tailored better when limiting the scope to railway freight, and that we were unable to obtain any meaningful and robust results for railway freight based on the data from the GUNVOR study. As later discovered, we were able to obtain more reliable results also using this data and an improved model specification (see section 4.2). This gives us the opportunity to compare the results of the two studies, which we do in this document.

2 Definition of reliability

2.1 Definition in the model

Our work is strongly inspired by recent developments within the methods to assess the value of reliability in personal travel. The measures of reliability considered for freight were therefore the same as those applied for people. As of now, the measures of reliability on which there is most consensus are the (1) standard deviation of travel time and (2) scheduling costs (Significance et al 2012).

In both our studies on freight, we therefore report values of reduced variability, measured as unit changes in the standard deviation of transport time. This is also what we have recommended to use in the case of road freight in Norway (Halse et al 2010). This corresponds with the recommendations for personal travel in the latest Norwegian valuation study (Samstad et al 2010). In the case of rail, the current practice in Norway (see chapter 5) is however to use the amount of delays as the measure of reliability. It is therefore relevant to include this measure as well,

Given this and since the SP experiments involving this measure seemed to work out well, we have provided recommended unit values to be used within the existing framework. Here the value of reliability in railway freight is measured as the value of the expected delay. By ‘expected’ we mean average, not that the occurrence or the length of the delay is in any sense known on beforehand. Further recommendations could however be made once we have more knowledge about what the standard deviation represents in the case of rail.

Originally, the idea behind the expected delay approach in the SP studies was to also obtain a value of expected early arrival and relate the results to the scheduling model described by Small (1982) and Fosgerau and Karlström (2010). We soon however learned that most freight customers do not consider early arrival as costly. The value of expected delay should hence not be interpreted as an underlying preference parameter, one should rather view expected delay as an alternative measure of uncertainty. In the case of passenger transport, Börjesson and Eliasson (2009) discuss the use of expected delay as a measure of reliability for rail.

Concerning the possible use of the standard deviation as the measure of reliability, an issue in the case of railway freight is that many of the freight trains arrive early and that this imposes no cost to the shipper/receivers or the consolidator for container transport. It only implies that the goods are kept at the terminal longer before they are transported to the final destination. (This uncertainty about arrival time could however be costly for the train and terminal operators.) How this affects valuation depends on how transport times are presented and interpreted in the SP experiment. Note also that if we measure reliability at the final destination, early arrivals will be part of the uncertainty considered by the shipper and receiver.

In addition to the choice of measure, an important question is at which level of the network the measure is to be applied. Most rail transports involve road transport between the rail terminals

and the shipper and receiver. In the GUNVOR study, we considered reliability measured at delivery to the receiver of the goods, while in PUSAM the object of study was the railway part of the transport chain. The results of the latter study should hence be the most suitable for valuing railway reliability measured at the terminal.³

2.2 Presentation of reliability

The way reliability was presented to the respondents is essentially the same in the two SP studies. In the first choice experiment involving reliability, the variability of transport time was presented in terms of five different transport times with the same probability (i.e. 20 percent). Mean transport time was not presented explicitly and there is also no information about departure or arrival times.⁴ The other attribute was the cost of the transport.

The screenshot shows a choice experiment interface from the Transportøkonomisk institutt (TOI) and Stiftelsen Norsk senter for samferdselsforskning. The question is: "Gitt at alt annet er likt, hvilken jernbanetransport vil du velge?" (Given that everything else is equal, which railway transport would you choose?).

Two options are presented:

- Transport A (Yellow background):**
 - Kostnad: 2000 kr
 - Framføringstid: Anta at disse framføringstidene har lik sjansse for å forekomme.
 - 5 timer og 20 minutter
 - 5 timer og 20 minutter
 - 5 timer og 40 minutter
 - 5 timer og 50 minutter
 - 6 timer og 10 minutter
- Transport B (Orange background):**
 - Kostnad: 1371 kr
 - Framføringstid: Anta at disse framføringstidene har lik sjansse for å forekomme.
 - 5 timer og 10 minutter
 - 5 timer og 20 minutter
 - 5 timer og 40 minutter
 - 6 timer og 10 minutter
 - 7 timer og 0 minutter

At the bottom, there are "Forrige" and "Neste" buttons, and a progress bar showing 0% to 100%.

Figure 2.1. Presentation of reliability in the choice experiment involving variability (CE2)

In the second choice experiment, reliability was presented by a certain probability of delay in one of the alternatives, and the length of the delay should it occur. The other alternative was always without risk of delay. Here, transport time was not presented at all, but respondents were instructed to take their actual transport as the point of departure. As before, transport cost was the other attribute.

In the GUNVOR study, the respondents also faced one or three (out of six) choice situations where there was a risk of early arrival instead of delay. As most firms apparently did not consider early arrival as costly, these results are left out in the following. In the PUSAM study, only late arrival was considered.

³ Another important subject is how reliability at arrival is related to measures of reliability at the railway link level. We do not discuss this in this working paper.

⁴ In the second study, where transport time is the the transport time of the railway part of the transport, it is hence a bit unclear whether the respondents would picture early train arrivals as being part of the distribution.



Figure 2.2. Presentation of reliability in the experiment involving risk of delay (CE3)

3 The SP survey

3.1 Why use SP data?

Following Bruzelius (2001), we can distinguish between three ways of obtaining values of goods for which there is no market for the good itself:

1. Deriving values from market prices
2. Estimating values based on actual behaviour (RP data)
3. Estimating values based on hypothetical choices (SP data)

Concerning the value of transport time savings in freight, an example of (1.) the capital-value approach, where the values are calculated based on the value of the goods. The idea is that as long as the goods are in transport, they are not available for consumption or as input to production. This approach typically gives very moderate values, and furthermore there is no obvious way to calculate values of reliability using this approach.

Another market price-based approach is to use costs of holding buffer stocks. This has been suggested several times but so far not implemented.

Using actual behaviour – revealed preference (RP) is attractive, but the availability of data is scarce. To estimate values of reliability, we would need quite detailed data on the choices made by shippers of goods, not just aggregate data on freight flows. Concerning the level of reliability, there is data on railway reliability for as good as all of the Norwegian railway network, although it has not been utilized for socioeconomic calculations to a large extent. In the case of road transport, there is less systematic data collection.

Although RP studies would be most welcome, it is difficult to picture how such studies could be made representative for the freight market as a whole. Typically the sample would depend on what data is available and different explanatory variables would need to be included for different contexts. In an SP study, on the other hand, one can recruit all relevant firms as respondents and treat other factors affected the choices as fixed. Weighting the results based on observable

characteristic, one can in principle obtain values which are quite representative for the market in question.

3.2 Design of the SP survey

3.2.1 Questionnaire and choice experiment design

In both studies, the main purpose of the survey questionnaire was to prepare the respondents for the choice experiments (CEs). The choice experiments take as the point of departure an actual transport or shipment reported by the respondent, and the attribute values in the choice experiments are based on the actual costs and transport time of the shipment or transport ('pivoted design').

In addition, the questionnaires contained other questions, particularly about experiences with unreliability in transport. The structure of the questionnaire can be summarized as follows:

- A. Introduction and questions about the firm
- B. Questions about a specific shipment/transport and its characteristics
- C. Choice experiment with deterministic transport time (CE1) and follow-up questions about choice behavior
- D. One or two contingent valuation questions about transport time savings
- E. Questions about experienced unreliability and its consequences
- F. Choice experiment with variable transport time (CE2) and follow-up questions
- G. Choice experiment with risk of delay (CE3) and follow-up questions
- H. Additional questions about the firm and the possibility to comment on the questionnaire

The choice experiment design was to a large extent based on that which was developed in the Norwegian value of time study for personal travel (Ramjerdi et al 2010) and which is similar to the design in the Danish, Swedish and Dutch value of time studies. The following features were common in the studies in GUNVOR and PUSAM:

- There were eight choices in CE1 and six in each of CE2 and CE3
- The reference cost occurred in one of the alternatives
- The reference transport time occurred in one of the alternatives in CE1 (and as basis for the distribution of transport times in one of the alternatives in CE2)
- CE1 involved two willingness-to-pay (WTP) choices, two willingness-to-accept (WTA) choices, two equivalent gain (EG) choices and two equivalent cost (EL) questions. (See explanation later in the text.)
- In CE2, two of the three attributes cost, mean transport time and variability co-varied in each choice, based on a fixed pattern
- The order of the choices was randomized
- Which alternative was on which side (left or right) was randomized

The attribute values were set as follows: First a percentage deviation in one of the attributes (cost or time/delay) was set. This deviation was drawn randomly from different intervals, once from each interval. Then, a price parameter – a 'prior' value of transport time savings or delay – was set, also based on a random draw from intervals. The value of the other attribute (time/delay or cost) was then determined based on these two other values, but with restrictions on how much it could deviate from the reference value.

In the case where a percentage deviation in cost is drawn first and the other attribute is transport time, the value of the time attribute is calculated as follows:

$\Delta \text{time} = \Delta \text{cost} / \text{price parameter} = \% \text{ dev. cost} \cdot \text{reference cost} / \text{price parameter}$

In CE2, the value of the ‘time attribute’ in the design is an underlying transport time which is the basis for the distribution of transport times displayed. It is not exactly equal to the expected value of this distribution of transport times. More on these distributions follows.

Also note that in CE2, both the ‘time attribute’ and cost will be higher in one of the alternatives in two out of six choices. (But variability could be higher in the other alternative.) In this case, the price parameter cannot be interpreted as the price of transport time saving, because this ‘price’ would then be negative.

The following features differed between the two studies.

- In the GUNVOR study, the cost attribute was set first in all CEs, with the time/delay attribute depending on this and the price parameter. In the PUSAM study, the time attribute was set first, ensuring that all respondents faced some options with very similar and some with very different transport times
- In the GUNVOR study, the price parameter in CE1 was drawn from intervals with slightly higher values if the respondent reported that the shipment or transport was ‘time sensitive’
- In the GUNVOR study, CE3 involved one or three choices where there was a risk of early arrival in one of the alternatives. In PUSAM, the choice was always between risk of delay and no risk of delay.
- In the GUNVOR study, the length of the possible delay in CE3 depended (negatively) on the probability of delay, while in PUSAM there was no dependence
- In the PUSAM study, there was a programming error in CE2 such that the percentage deviations of the time attribute and the degree of variability in the five-point distributions were not distributed quite as intended.

For some respondents, the first of these changes between the two studies would make a significance impact on the presented choices. In the case where cost is drawn first, some respondents might only face relatively low differences in transport time between the alternatives and some only relatively high differences. In the latter group, the maximum restriction on the difference will often be binding, implying that the actual price of transport time savings offered is actually higher than what is intended in the design.⁵ This could lead to more lexicographic choice behavior, but we have no evidence on this. In any case, it seems more important to us to have control over the time attribute than the cost attribute. Hence, this was changed in the PUSAM study.

Presenting all the attribute levels and corresponding values would take up too much space here. Instead, we present a table which briefly summarizes the setup and the range of values.

⁵ This might not be that unfortunate, given that those who are likely to experience this are those firms which pay relatively high transport fees, which presumably are those who care more about transport time and less about costs.

Table 3.1. Attribute values in the choice experiments

Experiment	Attribute	Values in GUNVOR	Values in PUSAM
CE1	Cost	Drawn from 8 intervals, range 5-60 % for decreases, 5-300 % for increases	Dependent. Minimum -70 %
	Time	Dependent. Minimum -50 %, maximum +200 %	Drawn from 8 intervals, range 5-50 %
CE2	Cost	Drawn from 6 intervals, range 5-35 %	Dependent. Minimum -70 %
	Time	Dependent. Minimum -50 %, maximum +100 %	Drawn from 6 intervals, range 1-30 %
	Distribution	Five different degrees of variability	Five different degrees of variability
CE3	Cost	Drawn from 6 intervals, range 3-50 %	Dependent. Minimum -70 %
	Probability of delay	0, 5, 10, 20, 25, 30 or 40 %	0, 5, 10, 20, 25, 30 or 40 %
	Delay length (or early arrival)	Dependent. Minimum 3 % of reference transport time, maximum 100 %*	Drawn from 6 intervals, range 3-60 % of reference transport time

All percentages represent differences from the reference value (of time or cost) unless otherwise specified

* Restrictions on the length of delays vary between long and short transports

We now explain how the attribute levels are combined. In the transport time savings experiment CE1, there are

- Two choices between the reference transport and a transport which is faster and more expensive (WTP)
- Two choices between the reference transport and a transport which is slower and cheaper (WTA)
- Two choices between a transport which has the reference transport time and is cheaper and a transport which is faster and has the reference cost (EG)
- Two choices between a transport which has the reference transport time and is more expensive and a transport which is slower and has the reference cost (EL)

In the variability experiment CE2, there are effectively three attributes: Cost, the underlying transport time ('time attribute') and the applied distribution of transport times. These are combined using four fixed patterns ('design blocks') randomly assigned to the respondents. Each pattern implies that

- In two choices, the alternative which is more expensive than the other involves a lower value of the time attribute and a distribution with a lower degree of spread
- In two choices, the alternative which is more expensive involves a lower value of the time attribute but a distribution with a higher degree of spread
- In two choices, the alternative which is more expensive also has a higher value of the time attribute, but a distribution with lower degree of spread.

The distribution of transport time is generated by multiplying the time attribute with a set of five factors slightly below or above 1. One of these sets represents the 'base' distribution, two represent a decrease in variability and two represent an increase in variability. The sets of factors are shown below. Note that if the difference in the time attribute is large, the alternative which has lower variability in terms of the factors applied might have higher absolute variability in terms of the actual distribution of transport times presented (measured by e.g. the standard deviation).

Note also that the distributions do not have the exact same shape. If they were standardized to have zero expectation and variance equal to one, they would still look different. As shown by Fosgerau and Karlström (2010), the shape of the distribution could matter for valuation.

Table 3.2. Factors used to generate distribution of transport times in CE2 in GUNVOR, road transports taking more than one hour

Level	Level -2	Level -1	Base level	Level 1	Level 2
Transport time 1	0.97	0.97	0.9	0.8	0.7
Transport time 2	1	1	1	1	0.8
Transport time 3	1	1	1	1	1
Transport time 4	1	1.05	1.15	1.2	1.3
Transport time	1.03	1.07	1.2	1.6	2

Table 3.3. Factors used to generate distribution of transport times in CE2 in GUNVOR, other modes than road

Level	Level -2	Level -1	Base level	Level 1	Level 2
Transport time 1	0.99	0.99	0.98	0.95	0.9
Transport time 2	1	1	1	0.98	0.95
Transport time 3	1	1	1	1	1
Transport time 4	1	1.05	1.15	1.2	1.25
Transport time 5	1.03	1.07	1.2	1.5	1.7

Table 3.4. Factors used to generate distribution of transport times in CE2 in PUSAM

Level	Level -2	Level -1	Base level	Level 1	Level 2
Transport time 1	0.98	0.97	0.95	0.92	0.90
Transport time 2	0.99	0.98	0.97	0.95	0.92
Transport time 3	1.00	1.00	1.00	1.00	1.00
Transport time 4	1.00	1.03	1.05	1.10	1.20
Transport time	1.03	1.05	1.10	1.25	1.70

In the delay experiment CE3, attribute levels are also combined according to four fixed patterns. As already explained, in all choices one alternative has a risk of delay (or early arrival) while one has not. In three of six choices, the reliable alternative has the reference cost while the other is cheaper. In the other three, the unreliable alternative has the reference cost while the other is more expensive.

As mentioned earlier, the CE3 experiment in the GUNVOR study was originally designed to involve three choices between a riskless alternative and an alternative with a risk of early arrival. As we learned in the pilot study that most respondents did not consider early arrival a problem, we manipulated the design such that those who answered earlier that early arrival would not be costly faced only one such choice situation. In the two other choices, the early arrival was changed into a delay. In the PUSAM study, all six choices involved a risk of delay.

3.2.2 Targeted sample

The two SP studies are quite different with respect to the targeted sample and how respondents were recruited. In the GUNVOR study, the main objective was to gain more insight in the area. Hence, we targeted the study at all firms which could potentially be relevant and designed the survey to be as generic as possible.

When doing the PUSAM study, we were only interested in rail transport. Also, we knew somewhat more about how the survey should be designed to fit this group. We therefore only recruited firms which were already rail customers, constituting a more homogenous sample than that of the GUNVOR study.

The different samples are:

- a) Shippers which buy transport services (GUNVOR)
- b) Shippers using own-account freight (GUNVOR)
- c) Transport companies, i.e. trucking firms, forwarders, shipping firms, rail operators (GUNVOR)
- d) Rail customers having direct contracts with the rail operator CargoNet (PUSAM)

In this document, we focus the most on the results from the studies targeted at (a) and (d), but we review some experiences also from the two other parts of the GUNVOR study.

Shippers which buy transport services (a) were targeted because they presumably are the best to represent the interest of fast and reliable transport related to the cargo itself. Alternatively the receivers of the goods could have been targeted, but these are likely to have less information about how long the transport takes and what it costs. A possible exception is when the receivers pay for the transport, something which is not uncommon (Hovi and Hansen 2010).

Transport companies (c) were targeted because we expected that an SP study targeted at this group would inform us about the benefits of transport time savings and improved reliability related to the operating costs of the transport. As we will see later, the way the questions were framed makes it difficult to say whether this is actually what we measure. Shippers using own-account freight (b) were included in the study for comparison with the two other samples.

Rail customers (d) include both shippers and transport companies (consolidators) which organize transports on the behalf of shippers (but which do not run their own trains), among which the latter constitutes the largest share. These were targeted because they could presumably better inform us about the tradeoffs between cost, time and reliability related to the rail transport in itself, while the shippers in (a) only experience changes in these factors for the transport chain as a whole (including terminals and connecting road transport). The transport companies could be seen as a surrogate for the shippers because they only care about the departure and arrival of the goods, not the train operation. A critical assumption is however that they take the interests of their customers into account when valuing changes in reliability

In the GUNVOR study, we recruited respondents in all groups (a)-(c) from two sources. The first was a commercial company database called Kompass Norge (<http://no.kompass.com>). The other was a database of firm data collected by Bodø Graduate School of Business at the University of Nordland. The latter database only contained firms from the western, middle and northern parts of Norway.

From these databases, we selected 9826 firms which were considered as potentially relevant respondents. Among these were 883 firms within the transportation industry and 8943 firms within manufacturing, wholesale, retailing, construction or other sectors dealing with physical goods. Not all firms were actually involved in the shipping of goods, but we could not identify which firms were and were not. We did not have any contact with the respondents prior to the survey.

In the PUSAM study, respondents were recruited from the customer database given to us by the rail operator CargoNet. The databases contained 340 contacts. Most of these were within forwarders which manage transports on the behalf of the owners of the goods, but there were also a few shippers which have direct contracts with the rail operator. In the case of the big forwarding companies, there were several contacts within the same company, often representing different subsidiaries.

Before launching the survey, we sent an e-mail to the potential respondents explaining the purpose of the study and asking whether they wanted to participate. After removing those who answered that they did not want to participate and the invalid e-mail addresses we were not able to correct, we were left with 227 contacts. Among these, 42 had confirmed that they would like to participate in the survey while the rest had not answered anything. Most of those who answered that they did not want to participate informed us that they were no longer using rail transport services.

3.2.3 Data collection and response rate

Both surveys were conducted using online questionnaires, with no assistance given other than to those who contacted us with questions about the questionnaire. The GUNVOR survey was launched in March and April 2010. Before that, we had conducted a small pilot study for all segments.

The firms were invited to participate in the survey via an e-mail containing a link to the questionnaire. In the e-mail we explained that the purpose of the study was to learn about the demand of businesses for fast and reliable transport and that the research would constitute a basis for public policy. The purpose was to motivate firms to respond to the survey. Of course, it might also have driven them to attempt to respond strategically (although it is somewhat difficult for the respondent to perceive how his or her responses affect the results).

Concerning the response rate, we only know how many e-mails reached a valid e-mail account and not how many of these corresponded to firms which were actually shipping goods (or doing business at all). There was a question in the introduction asking whether the firm had shipped goods or conducted transports during the last two weeks (one week for transport companies), but the answers here are not very informative as many of those who were not shipping goods would probably not log in at all after reading the e-mail. In total, 12.5 percent of the invited shippers and 18.8 of the transport companies logged in. Respectively 7.2 percent and 13.4 percent participated in the first choice experiment, and 5.6 percent and 5.9 percent answered the parts of the questionnaire down until and including the third choice experiments.

Table 3.5 shows the how different sectors are represented among the shippers who have responded to the survey up until and including the first choice experiment. Note that the figures also include those who transport their own goods themselves. Among the manufacturing firms, the industries constituting the largest shares are ‘metallic goods excluding machines and equipment’ (13.6 %), ‘food and beverages’ (12.3 %) and ‘machines, equipment and electrical devices’ (11.9 %). Among the wholesalers, the largest shares are made up by ‘machines, production equipment, boats and airplanes’ (15.0 %), ‘furniture, tools , paint, wallpaper and other household goods’ (11.0 %) and ‘electrical devices’ (11.0 %). A broad variety of other industries and branches were however also represented.

Table 3.5. Sectors represented among the shippers in the GUNVOR survey

Sector	Number	Share
Manufacturing	290	45.3 %
Wholesale	127	19.8 %
Construction	89	13.9 %
Other	134	20.9 %
Total	640	100.0 %

The PUSAM survey was tested twice in a computer-assisted personal interview (CAPI) with a respondent representing one of the major forwarding agents in Norway. As no substantial changes were made in the questionnaire after the second testing, we decided to include the responses of this respondent in the sample used for analyses. This respondent also answered the survey online; hence the data contains three responses which are from the same respondent but representing different shipments.

The other respondents were recruited via e-mail. Out of the 227 e-mails sent, 32 respondents answered the whole questionnaire. This corresponds to a response rate of 14.1 percent. Including the two CAPI responses, this gives 34 responses in the data used for analysis. Only one respondent who started on the first choice experiment did not complete the questionnaire. This respondent was left out of the analysis. Table 3.6 shows the respondents by different segments.

Table 3.6. Respondent segments in the PUSAM survey on rail transport

	Invited contacts		Completed questionnaires		Response rate
	Number	Percent	Number	Percent	
Forwarding agent/consolidator	149	65.6	26*	76.5	16.1 %
Shipper buying transport services	19	8.4	4	11.8	21.1 %
Shipping firm	51	22.5	4	11.8	7.8 %
Other	8	3.5	0	0	0 %
Total	227	100	34	100	14.1 %

*Including two CAPIs. These are not included when calculating the response rate.

15 out of the 26 responses from forwarding agents or consolidators come from respondents representing the four companies Bring, DB Schenker, Tollpost and DHL. This is a reasonable feature, as these companies dominate the market for consolidated rail transport.

We have not undertaken any measures in the analyses in any of the two studies to deal with representativity with respect to firm type or industry. In the PUSAM study, we have however taken into account the distribution between consolidated and non-consolidated goods, which we will show proved important. Also, the choice of model specification, which is particularly important with the GUNVOR data, can have an impact on to which extent the results reflect the preferences of different groups of firms.

4 The analysis of the SP survey

4.1 Data checks and processing

A central characteristic of the data from the GUNVOR study is the vast heterogeneity in the quantitative shipment characteristics – transport cost and transport time – which are used to generate the attribute values in the choice experiments. This is an issue which has to be handled in the model specification. However, we excluded three respondents with a reference cost of 7 NOK or less, because the choice experiment design would work very poorly with such low cost values. Table 4.1 shows the range of reference values of all three groups of respondents.⁶ (The shipments also varied a lot with respect to other quantitative features like shipment weight and value and shipping distance.)

Table 4.1. Summary statistics for reference attribute values in the GUNVOR study

Type of firm	Mode	Variable	No.	Min	Max	Mean	Std. D
Shippers, hired transport	All	Ref. cost (NOK)	504	25	100000	3957	9331
		Ref. time (min.)	504	20	18720	2364	2228
	Road only	Ref. cost (NOK)	395	25	100000	3759	8553
		Ref. time (min.)	395	20	10142	2068	1902
Shippers, own-account freight	All	Ref. cost (NOK)	114	50	30000	2410	3945
		Ref. time (min.)	114	5	5265	245	582
	Road only	Ref. cost (NOK)	112	50	30000	2442	3972
		Ref. time (min.)	112	5	1800	203	340
Carriers	All	Ref. cost (NOK)	119	29	140000	10075	20402
		Ref. time (min.)	119	15	8640	1187	1756
	Road only	Ref. cost (NOK)	107	29	45000	6899	7862
		Ref. time (min.)	107	15	8640	1037	1643

Furthermore, we excluded 15 respondents who had answered the whole questionnaire in less than ten minutes, which is very little time given the length of the questionnaire. In the analyses reported in this document, those respondents who did not report the shipment weight are also excluded, as well as those who did not answer the questions about choice behavior following the choice experiments

As mentioned earlier, a considerable share of those who completed the first choice experiment did not answer the whole rest of the questionnaire. We have not excluded these respondents,

⁶ The data shown in this table includes some observations which were later left out of the analyses due to the reasons mentioned in the text.

such that those who completed CE1 (and the follow-up questions) are included in the analysis of the value of travel time savings and likewise in the case of CE2. In the data from the PUSAM survey we have not excluded any respondent except one who just started answering CE1 and did not continue. The respondents used for analysis have answered the whole questionnaire.

An important and potentially problematic characteristic of the data in both studies is that a considerable share appears not to have taken all choice attributes into account in their choices. We refer to this type of choice behavior as ‘lexicographic’ or ‘non-trading’. The presence of such behavior can be seen both by looking at the actual responses and the answers to the follow-up questions about choice behavior. In Table 4.2, Table 4.3 and Table 4.4 we compare the self-reported and observed choice behavior of the respondents in the GUNVOR survey. Note that we here also include the choice situations involving risk of early arrival in CE3. To avoid having too many tables and because the sample size is smaller in the PUSAM survey, we only show the self-reported choice behavior from this study, in Table 4.5.

The tables show that on overall, what respondents report about their choice behavior is in line with the choices they actually make. However, for some the answers to the follow-up questions are clearly not consistent with their choices in the CE. The most obvious inconsistency in the responses is that some respondent who stated that they had only considered one attribute (e.g. cost) have chosen in a way that implies that they care about both attributes (e.g. by sometimes choosing the faster and more expensive alternative in CE1). In Table 3.1 Table 4.2, we see that this is the case for 57 (38.3 %) of those 149 who stated that they only cared about cost.

We define the observed choice behavior as ‘lexicographic’ if a respondent always chose the alternative which was better with respect to a certain attribute, for instance always the alternative with the lower cost. We can however not know for certain that such a respondent only cared about cost and not about e.g. transport time, because he or she might have chosen differently had the attribute values been different. One should therefore be careful drawing conclusions about other forms of inconsistency, especially in the two choice experiments where there are only six choices.

The share who does seem to take both/all attributes into account is about 65-75 percent in both studies, both when we consider self-reported and actual choice behavior. The most common type of lexicographic behavior is to only consider costs, but this seems to change somewhat when the other attribute is risk of delay. We return to the question about how to treat this type of responses in the analysis in the next section.

Table 4.2. Comparison of self-reported and observed choice behavior in the first choice experiment of the GUNVOR study (shippers with hired transport).

Self-reported behavior, CE1 (8 choices)	Number	Share	Actual (seemingly) lexicographic choices with respect to an attribute?		
			No	Yes, cost	Yes, time
"Considered both attributes"	315	64.7 %	263	52	0
"Considered only cost"	149	30.6 %	57	92	0
"Considered only transport time"	23	4.7 %	22	0	1
Total	487	100.0 %	342	144	1

Table 4.3. Comparison of self-reported and observed choice behavior in the second choice experiment of the GUNVOR study (shippers with hired transport).

Self-reported behavior, CE2 (6 choices)	Number	Share	Actual (seemingly) lexicographic choices with respect to an attribute?			
			No	Yes, cost	Yes, mean time*	Yes, variability
"Considered both attributes"	304	71.0 %	223	56	18	11
"Considered only cost"	101	23.6 %	31	70	0	0
"Considered only the transport time distribution"	23	5.4 %	15	1	6	3
Total	428	100.0 %	269	127	24	14

*This column and the one to the left includes six respondents who six times chose the alternative which had *both* lower mean time and variability, due to weaknesses in the experiment design. (See section 3.2.1.)

Table 4.4. Comparison of self-reported and observed choice behavior in the third choice experiment of the GUNVOR study (shippers with hired transport).

Self-reported behavior, CE3 (6 choices)	Number	Share	Actual (seemingly) lexicographic choices with respect to an attribute?		
			No	Yes, cost	Yes, reliability
"Considered both attributes"	293	74.9 %	254	23	16
"Considered only cost"	67	17.1 %	28	39	0
"Considered only late/early arrival"	31	7.9 %	20	0	11
Total	391	100.0 %	302	62	27

Table 4.5. Self-reported choice behaviour in the PUSAM study.

Reported choice behavior	CE1 (8 choices)	CE2 (6 choices)	CE3 (6 choices)
"Considered both attributes"	67.6 %	73.5 %	73.5 %
"Considered only cost"	20.1 %	20.6 %	14.7 %
"Considered only transport time/reliability"	11.8 %	5.9 %	11.8 %
Total	100 %	100 %	100 %

4.2 Model specifications

The results reported in this document are all based on estimation using multinomial logit (MNL) models. Furthermore, with one exception⁷ the models do not contain any explanatory variables for the value of transport time savings or reliability. This contrasts with the methodology used to estimate the value of travel time savings in the Norwegian value of time study for personal travel

⁷ See the analysis of CE3 from the PUSAM study, where there are two time coefficients which are interacted with dummies representing the type of shipment.

(Ramjerdi et al. 2010). In that study, the model is specified as a mixed multinomial logit (MMNL) model in the willingness-to-pay space as advocated by Fosgerau (2007), using covariates which help explain whether the respondent accepts the ‘price’ of transport time savings implicit in the more expensive alternative. Some attempts were made with such a model using the GUNVOR data but it was not very clear from the results which covariates should be included and not.⁸ Therefore, the models are specified in the conventional ‘preference space’ manner where the utility of alternatives A and B depend on the values of the choice attributes of each alternative. Ignoring the structure of the error term for now, the utility function of each alternative in CE1 is:

$$(1) \quad U_i = \alpha_i L_i + \beta_{COST} C_i + \beta_{TIME} T_i$$

where C_i is transport costs, T_i is transport time and L_i is equal to one if the alternative is shown on the left side and zero otherwise. Furthermore, we discarded using a mixed multinomial logit model with β_{TIME} as a random parameter in the analyses used in our recommendations. Halse and Ramjerdi (2012) shows that model fit is clearly higher in the MMNL model with a log-normally distributed β_{TIME} , but the values of transport time savings are unreasonably high unless we censor the distribution at some cut-off point. It is not clear where this cut-off point should be, especially considering the large heterogeneity in the freight market. If we assume that the scale of the error terms is the same regardless of the attribute values, the model is then

$$(2) \quad U_i = \alpha_i L_i + \beta_{COST} C_i + \beta_{TIME} T_i + \varepsilon_i$$

where ε_i is assumed to be independently, identically Gumbel distributed. However, the large heterogeneity in the reference values of transport time and costs in the GUNVOR data implies that the model is very sensitive to respondents with large C_i or T_i . If such a respondent makes a choice which is relatively inconsistent with the estimated utility parameters, the error term will be estimated to be very large. We therefore instead assume the multiplicative model⁹

$$(3) \quad U_i = e^{\alpha_i L_i} + (C_i + \beta_{VTTS} T_i)^\mu e^{\varepsilon_i}$$

which on logarithmic form becomes

$$(4) \quad \ln U_i = \alpha_i L_i + \mu \ln(C_i + \beta_{VTTS} T_i) + \varepsilon_i$$

In this model β_{VTTS} directly gives the value of transport time savings. This model proved to yield substantially higher model fit than the additive error specification (2) when estimating on the GUNVOR data. In the case of the PUSAM study, there was not much of a difference, probably reflecting the fact that the attribute heterogeneity is considerably smaller in this data. The results reported from the PUSAM study are based on additive specifications like (2).

Finally, the models are estimated taken into account the panel structure of the choice data, namely that the eight or six choices made by the same respondent are likely to be correlated. In practice this is done by specifying the user ID as the panel variable in the software BIOGEME, as recommended by Daly and Hess (2010)

We now continue to the specification of the utility functions. We have already shown this in the case of the value of transport time savings in CE1: Utility is assumed to depend linearly (and

⁸ The findings are reviewed by Halse and Killi (2012). WTP for lower transport time increases in shipping costs and decreases in shipping distance. It is also significantly higher for textile goods than for chemical goods and timber and higher for air transport freight than for road freight. However, on overall the covariates add very little explanatory power to the model.

⁹ This specification was proposed by Mogens Fosgerau in a lecture given at the Kuhmo Nectar summer school in transport economics in 2011 and is also used in parts of the Dutch value of time study on freight.

negatively) on cost and transport time. In the case of CE2, we assume the following model of the utility of costs, transport time and variability measured by the standard deviation of transport time. In the additive error term specification, the model is

$$(5) \quad U_i = \alpha_i L_i + \beta_{COST} C_i + \beta_{MEAN} E(T_i) + \beta_{STD} Std(T_i) + \varepsilon_i$$

Using the multiplicative formulation, the model becomes

$$(6) \quad \ln U_i = \alpha_i L_i + \mu \ln [C_i + \beta_{VETTS} E(T_i) + \beta_{VTTV} Std(T_i)] + \varepsilon_i$$

This specification was chosen because the standard deviation is currently the most widely applied measure of travel time variability for personal travel (see e.g. Significance et al, 2012), and we wanted to test whether this model would work well in the case of freight. Halse and Ramjerdi (2012) show the results when using a rank-dependent utility model to analyze the same data. In the case of the third choice experiment (CE3), we first tried specifications including in the spirit of the scheduling model by Small (1982), where expected utility depends negatively on both early and late arrival (delivery). However, we soon realized that we would not obtain any meaningful results with respect to scheduling costs of early arrival, as most respondent did not consider this as costly. We therefore omitted the choice situations involving the risk of early delivery and analyzed the model

$$(7) \quad \ln U_i = \alpha_i L_i + \beta_{COST} C_i + \beta_{DELAY} E(D_i) + \varepsilon_i$$

where $E(D_i)$ denotes expected delay, in this case delay multiplied by probability. Using the multiplicative error specification, the model is

$$(8) \quad \ln U_i = \alpha_i L_i + \mu \ln [C_i + \beta_{DELAY} E(D_i)] + \varepsilon_i$$

We also tested specifications including a constant penalty of delay as suggested by Small, but with the GUNVOR data this estimate turned out as negative¹⁰, which is counter-intuitive. This is probably related to the choice experiment design, where the alternative involving delay is always the cheapest and cost is highly correlated with the delay probability.

Using the PUSAM data, we obtained some reasonable results with alternative specifications. These point in the direction that for the aggregate sample, the cost of delay is decreasing in the delay length. At the same time, the WTP to avoid delay is lower when delay is below some threshold. We think this is very reasonable, thinking of the situation of an individual firm: Very short delays presumably have few consequences. And once delays exceed a certain length, both the forwarding agent and its customer will have to do some rescheduling, implying that the benefits of reducing the delay by a marginal amount might not be very big. In the survey we asked the respondents about their critical threshold. As we see on Figure 4.1, these are clustered around delays of a few hours.¹¹

¹⁰ This analysis was performed using an additive error specification. It would not work with the specification (8), because the expression inside the logarithm would be negative for some parameter estimates and attribute values.

¹¹ We also asked whether some or all of the goods in the shipment were to be delivered at a specific time the day the train arrives. 35.3 percent answered yes. 44 percent answered that they only needed to be delivered at some time during this day. 21.1 percent answered no.

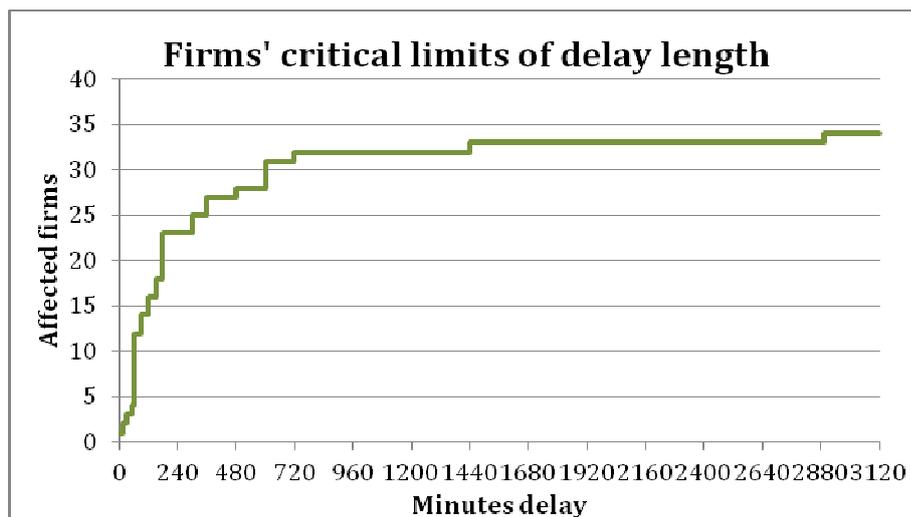


Figure 4.1. The critical limits of delay length reported by the firms in our sample

Taking these results into account, assuming a linear disutility of expected delay which is the same for all firms is clearly a simplification of reality. It is however very convenient for aggregation, and we would also not feel confident in choosing one of the other specifications as the true one, given the small data sample.

In the analysis of the PUSAM data, we omit the side-preference parameter a_i because we did not want arbitrary correlation to affect the results, given that we had a rather small sample size.¹² We tested including it in the analysis of CE1, and it had little impact. We also do not observe any clear evidence of respondents choosing alternatives displayed on one side rather than the other. As noted earlier, the degree of involvement by the respondents seems to have been higher in this survey than in the GUNVOR survey.

As pointed out in section 4.1, an important issue is how to handle the observations of respondents which do not take both/all attributes into account. Several authors have suggested ways of taken this into account, for instance Hensher and Rose (2009) and Hess et al (2012). We choose a simple approach and eliminate an attribute from the utility function if this attribute was ignored. The remaining question is then how to identify attribute non-attendance.

We showed in the previous section that the statements of respondents about their own behavior is not always consistent with observed behavior. In the case of CE1, we have only two attributes and eight consecutive choices, so inferring non-trading choice behavior based on actual choices is clearly an option. However, in the two other choice experiments we have only six observations of each respondent and in CE2 it is not clear which characteristic(s) of the presented transport time distributions should be used to define non-attendance. We therefore chose to use stated behavior (from the answers from the follow-up questions) to define whether an attribute was ignored or not.¹³ In the case of CE2, the 'time attribute' is then the whole transport time distribution. (See Table 4.3.)

We would strongly advice against excluding respondents which choose lexicographically, as this makes the sample less representative. It is also likely to bias the results because some attributes

¹² Strictly speaking, this also applies to the rail segment in the GUNVOR study.

¹³ In the results from the GUNVOR study previously reported by Halse et al (2010), observed (inferred) behavior is used instead.

are ignored more often than others. In the appendix in the report by Halse et al (2010), it is shown excluding non-traders result in a higher estimated VTTS than the other approaches tested. A more practical issue is whether attributes and derived values of transport time savings and reliability should be expressed in per shipment or per ton units. In the PUSAM study, this is not too important since a ‘shipment’ here represents the whole loading unit, so it would be relatively easy to compare the results with those of other results no matter which unit is chosen. In the GUNVOR study however, the shipment is the cargo unit sent by the shipper, which could be anything from a small package to several truck or wagon loads.

In order to make the results comparable and comparison more convenient, we estimated VTTS and values of reliability per ton for both data sets. This introduces some measurement error in the GUNVOR case because here we only have weight reported by intervals (0-35 kilos, 36-99 kilos etc.). The model specifications in the analysis of CE1 in the two studies are, respectively

$$(9) \quad U_i = e^{\alpha_i L_i} + (C_i / weight_i + \beta_{VTTS} T_i)^{\mu} e^{\varepsilon_i}$$

$$(10) \quad U_i = \alpha_i L_i + \beta_{COST} C_i + \beta_{TIME} T_i \cdot weight_i + \varepsilon_i$$

where C_i is measured in kroner (2011-NOK), T_i in minutes and $weight$ in tons. The specifications are in principle equivalent (except for the way the error term enters), but for some reasons a model where we divided cost by weight performed slightly worse and did not give the exact same results when used on the PUSAM data. In the following tables we also show the results from the PUSAM study when attributes are measured per shipment.

Halse et al (2010) find that the relationship between weight and shipment VTTS is less than one-to-one in the GUNVOR data, but we interpret this as an effect of other characteristics of small and large shipments rather than an effect of shipment size in itself. In the PUSAM study and especially concerning delays, it could however be that forwarding agents have some fixed costs of delays which are independent of shipment size, causing the relationship between weight and values of reliability to be truly non-linear. Halse and Killi (2011) find some evidence of this, but the results must be viewed with caution given the small sample size.

4.3 Results

Tables Table 4.6-Table 4.8 show the results of the estimations. We divide the GUNVOR data into road transport and rail transport, leaving out those shipments which were carried by other modes (ship etc.). All coefficient estimates representing valuation have the expected sign and are statistically significant, except for the coefficient representing the utility of the standard deviation in the analysis of CE2 from the PUSAM study. This estimate is positive and not statistically significant; hence it cannot be used for valuation. It could be explained by poor experiment design, a small sample size or inconsistent choice behavior. However, the value of reducing expected transport time ($\beta_{MEAN} / \beta_{COST}$) from these results is in a reasonable order of magnitude.

In the PUSAM study, we segment between general cargo (consolidated shipments) and pallet cargo in the analysis of CE1 and CE3. (Doing this for CE2 did not help in retrieving meaningful results). That is, we estimate different time or delay coefficients for the two groups of shipments. Model fit (adjusted rho-squared) is satisfactory in all estimations, except for the analysis of CE3 in the road transport sample of the GUNVOR data. In this model this statistic is very low, but the coefficient estimates of interest are nevertheless statistically significant and of reasonable magnitude.

Table 4.6. Multinomial logit results, road transport in GUNVOR study (shippers with hired transport).

	CE1 (multiplicative error)		CE2 (multiplicative error)		CE3 (multiplicative error)	
	Estimate	Rob st. error	Estimate	Rob st. error	Estimate	Rob st. error
a_i	0.00144	0.0471	-0.0763	0.0582	-0.0951*	0.0534
μ	-4.24***	0.281	-6.44***	0.435	-1.73***	0.366
β_{VTTs}	0.198***	0.0343				
β_{VETTs}			0.235***	0.0579		
β_{VTTV}			0.30***	0.0757		
β_{VED}					0.829**	0.403
Obs.	3072		2010		1525	
Respondents	384		335		319	
Null Log L (LL ₀)	-2129		-1393		-1057	
Final Log L (LL _F)	-1483		-1054		-1021	
Adj. rho-square	0.302		0.24		0.032	

Note: All models were estimated using BIOGEME (Bierlaire 2003). Robust t-tests were computed taking into account the repeated observations nature of the data.

* p<0.1, ** p<0.05, *** p<0.01

Table 4.7. Multinomial logit results, rail transport in GUNVOR study (shippers with hired transport)

	CE1 (multiplicative error)		CE2 (multiplicative error)		CE3 (multiplicative error)	
	Estimate	Rob st. error	Estimate	Rob st. error	Estimate	Rob st. error
a_i	0.004	0.142	-0.351	0.212	-0.142	0.202
μ	-4.50***	0.876	-9.99***	1.67	-5.00***	1.3
β_{VTTs}	0.444**	0.212				
β_{VETTs}			0.399***	0.0952		
β_{VTTV}			0.727***	0.256		
β_{VED}					1.49**	0.562
Obs.	336		228		167	
Respondents	42		38		35	
Null Log L (LL ₀)	-232.9		-158.0		-115.8	
Final Log L (LL _F)	-161.2		-97.9		-94.1	
Adj. rho-square	0.295		0.355		0.161	

Note: All models were estimated using BIOGEME (Bierlaire 2003). Robust t-tests were computed taking into account the repeated observations nature of the data.

* p<0.1, ** p<0.05, *** p<0.01

Table 4.8. Multinomial logit results, rail transport in PUSAM study

	CE1 (additive error)		CE2 (additive error)		CE3 (additive error)	
	Estimate	Rob st. error	Estimate	Rob st. error	Estimate	Rob st. error
β_C	-0.00161***	-0.000443	-0.00176***	-0.000418	-0.00166***	0.000636
β_t^G	-0.00127***	-0.000332				
β_t^P						
	-0.000198**	-0.000102				
β_{MT}			-0.000573***	-0.000189		
β_{ST}			-0.000094	0.000258		
β_d^G					-0.00770***	0.00242
β_d^P						
					-0.00097***	0.000398
Obs.	272		204		204	
Null Log L (LL ₀)	-188.5		-141.4		-141.4	
Final Log L (LL _F)	-140.3		-116.5		-100.1	
Adj. rho-square	0.24		0.155		0.271	

Note: All models were estimated using BIOGEME (Bierlaire 2003). Robust t-tests were computed taking into account the repeated observations nature of the data.

* p<0.1, ** p<0.05, *** p<0.01

Based on these results, values of transport time savings (VTTS) and reliability can easily be obtained. In Table 4.9, we show both the value of a standard deviation measured in NOK (VTTV), the ratio between the utility coefficients of the standard deviation and mean of transport time ('reliability ratio' – RR) and the value of an expected delay (VED). The values in the last row are weighted averages of the values of general and pallet cargo, based on market shares reported by the largest rail operator in Norway (CargoNet).

Table 4.9. Values of transport time savings and reliability (NOK/ton-hour)

Sample	VTTS (CE1)	VTTV (CE2)	RR (CE2)	VED (CE3)
Road, GUNVOR	12	18	1.3	50
Rail, GUNVOR	27	44	1.8	89
Rail general, PUSAM	47	--	--	278
Rail pallet, PUSAM	7	--	--	35
Rail all (weighted), PUSAM	13	--	--	72

We see that the unit values are quite comparable in magnitude. As expected, the value of avoiding an hour delay which is uncertain is higher than the value of an hour of known transport time savings. Whether the results from CE2 and CE2 are consistent depends on what assumptions one makes about the shape of the transport time distribution and the share of shipments arriving late, but the results seem reasonable.

Concerning rail transport in particular, we see that time and reliability is much more important for general cargo. This is as expected, because this is a concept which in Norway has attracted shippers of many categories of goods, also rather time-critical goods. Pallet cargo, on the other hand, will presumably not be sent by rail unless transport time is of relatively low importance.

Given that the shipment already amounts to a truckload, sending it by rail will result in longer transport time due to terminal handling. (On-board transport time is similar for road and rail.)

We would primarily recommend using the results from the PUSAM study for socioeconomic calculations involving rail transport, not the GUNVOR results. We believe this sample is more representative, and it also allows us to segment between the two important segments general cargo and pallet cargo. Furthermore, the valuation can be related directly to the rail part of the transport chain, while respondents in the GUNVOR study considered the door-to-door transport time when they answered the CE questions.¹⁴ It is however reassuring that the results are not too different. It also allows us to derive an approximate value of the standard deviation for railway freight based on the results from the PUSAM study, although some assumptions would have to be made.

Regarding the values of reducing delays, there is some uncertainty about whether these can be applied to an improvement in reliability which only implies avoiding delays which would only have been of a very short length, say, 5 minutes. As seen on Table 3.1, the shortest delays presented in the CE are 3 percent of transport time, and most are much longer. Furthermore, on Figure 4.1 we see the vast majority of firms report that a delay only has consequences for delivery when it exceeds an hour or two. This does not imply that delays of a few minutes are not costly at all, but we are not confident that their costs are well reflected in our estimated values of delay.

We do not report the full results of the analyses of the two other groups in the GUNVOR study – shippers with own account freight and transport companies. These results have not been used in any recommendations for valuation. In Table 4.10 we present the main results for comparison, including all transport modes. The value of changes in the standard deviation had the unexpected sign and/or was not statistically significant for both groups. The ratio of the estimated value of expected delay and the VTTS is lower for transport companies here than in the results shown earlier for shippers and rail customers, but still higher than unity.

Originally, the idea was that the SP values of transport companies would represent the economic benefits related to time- and reliability-dependent transport costs. We are however not sure if this is actually the case in our study. First, the respondents were not instructed not to consider the consequences of changes in transport time and reliability for their customers, so they might have done this to some extent. Secondly, transport cost was an attribute in the choice experiments, so if the respondents read this literally they should not consider the benefits of saved time- or reliability dependent transport costs. It could however be that they interpreted the cost attribute more as a direct payment (e.g. road toll). In fact, the estimated VTTS is close to the one currently used in CBA for road freight, where the VTTS is based on factor costs (mainly labor costs).

¹⁴ Whether this is a drawback or a benefit depends on the object of the study.

Table 4.10. Further results from the GUNVOR study, shippers with own account freight and transport companies (all modes of transport)

5	6	7	Shippers, own account freight (N=114)	Transport companies (N=118)
CE1	VTTS (NOK/hour)	331		449
CE2	8 Value of expected transport time (NOK/hour)	1444		305
	9 Value of standard deviation (NOK/hour)	Not significant		Not significant
CE3	10 Value of expected delay (NOK/hour)	1361		872
Approximate average weight		3,9 tons		20,6 tons

5 Current practice in Norway

The current practice of CBA of transport projects in Norway is reported in the handbooks by the Norwegian Public Roads Administration (Statens vegvesen 2006) and the Norwegian National Rail Administration (Jernbaneverket 2011). The latter has recently been revised, while the former is due for revision.

In road transport, only known transport time savings are valued in the current practice. On behalf of the Public Roads Administration, TØI has estimated values of reliability for personal travel (Ramjerdi et al 2010), but these have so far not been used in official calculations.

In rail transport, delays are given a higher weight than changes in known transport time both for passengers and freight. The values of transport time savings and delays for passenger travel are based on those estimated by Ramjerdi et al, while the values for freight are taken from the Norwegian freight model (Madslie et al 2012). The model uses a cost function based on expert advice. The values are considerably lower than those in the results reported in this document, and the National Rail Administration is considering replacing them with the new values. As the freight model does not include uncertainty, the values of delay are estimated by multiplying the VTTS by a factor based on expert judgement.

Concerning predicting variability, this has not been given very much attention. In the case of road transport, the current practice is as mentioned not to value changes in variability. In rail transport, the CBA tool of the National Rail Administration contains a formula which estimates the amount of delays in hours based on the percentage of trains which are late. However, since the percentage late trains is not something which is estimated in the transport models normally used, this figure has to be based on some analysis which is specific to the project. This could be either simulations or judgement-based predictions. Furthermore, for passenger trains, delays are assumed to be the same for passengers getting off at all stations on the line.

References

- Börjesson M. og J. Eliasson 2009: On the use of "average delay" as a measure of train reliability, *Transportation Research Part A* 45 (2011) 171–184
- Bruzelius N 2001: *The Valuation of Logistics Improvements in CBA of Transport Investments; A Survey*. Underlagsrapport till SAMPLAN 2001:1
- Daly, A.J. & Hess, S. (2010): *Simple approaches for random utility modelling with panel data*. Working paper
- Fosgerau, M. (2007) Using nonparametrics to specify a model to measure the value of travel time. *Transportation Research Part A: Policy and Practice*. Volume 41, Issue 9, November 2007, Pages 842–856
- Fosgerau M og A Karlström (2010): The value of reliability. *Transportation Research Part B: Methodological*, Volume 44, Issue 1, s. 38-49
- Halse, A. H., Samstad, H., Killi, M., Flügel, S., & Ramjerdi, F. (2010). *Verdsetting av framføringstid og pålitelighet i godstransport*. Oslo: Transportøkonomisk institutt.
- Halse, A. and Ramjerdi, F. (2012). Value of reduced transport time and variability in freight transport. Paper presented at the INSTR2012 conference in Hong Kong.
- Hensher, D.A. og J. Rose 2009: Simplifying choice through attribute preservation or non-attendance: Implications for willingness to pay. *Transportation Research Part E*, Volume 45, Issue 4, s. 583-590
- Hess, S., Stathopoulos, A. & Daly, A.J. (2012), Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies, *Transportation* 39(3), pp. 565-591
- Hovi, Inger Beate og Wiljar Hansen (2010): *Logistikkostnader i norske vareleverende bedrifter. Nøkkeltall og internasjonale sammenlikninger*. TØI-rapport 1052/2010. Oslo, Transportøkonomisk institutt.
- Jernbaneverket 2011: *Metodehåndbok JD 205. Samfunnsøkonomiske analyser for jernbanen*. Versjon 3.0, juli 2011. Oslo, Jernbaneverket Utredningsavdelingen.
- Madslie, A. C. Steinsland and S. E. Grønland (2012). Nasjonal godstransportmodell. En innføring i bruk av modellen. TØI-rapport 1247/2012
- Ramjerdi, F, S Flügel, H Samstad og M Killi (2010): *Verdien av tid, sikkerhet og miljø i transportsektoren. Dokumentasjonsrapport B: Den norske tidsverdistudien*. TØI-rapport 1053B/2010. Oslo, Transportøkonomisk institutt
- Significance, Goudapel Coffeng, NEA 2012. Erfassung des Indikators Zuverlässigkeit des Verkehrsablaufs im Bewertungsverfahren der Landesverkehrswegeplanung: Schlussbericht. Bundesministerium für Verkehr, Bau und Stadtentwicklung

Samstad, H., F. Ramjerdi, K. Veisten, S. Navrud, K. Magnussen, S. Flügel, M. Killi, A. H. Halse, R. Elvik, O. San Martin 2010: *Verdien av tid, sikkerhet og miljø i transportsektoren – Sammendragsrapport*. TØI-rapport 1053/2010.

Small K 1982: The scheduling of consumer activities: work trips. *American Economic Review*, Volume 72, No. 3, s. 467-479.

Statens vegvesen (2006): Håndbok 140 Konsekvensanalyser. Statens Vegvesen Vegdirektoratet

Annex B) VTTV in the recent national stated preference study on values of time and reliability in freight transport in The Netherlands

1. Objectives of the study

The freight values of transport time (VTTs) currently used in cost-benefit analyses (CBA) of transport projects and policies in The Netherlands are based on stated preference (SP) research carried out in 2003/2004 (RAND Europe et al., 2004). For the VTTV, reliability ratios (RRs; expressing the value of reliability relative to the value of time) have been derived from the same survey, but this required many additional assumptions (de Jong et al., 2009).

Therefore, a new study (Significance et al., 2012a) was carried out for the Dutch Ministry of Infrastructure and the Environment to obtain up-to-date, evidence-based monetary VTTs and VTTVs in freight transport for use in CBA. This project also dealt with passenger transport, but this is not treated in this chapter.

The modes for freight transport covered in this project were:

- road;
- rail;
- air;
- inland waterways;
- sea.

In August 2013, the VTTs from this study will replace the existing values; the VTTVs will be the first of their kind for the Netherlands.

2. Definition of reliability/variability

a. Reliability in the model

In earlier projects (RAND Europe, 2004; Hamer et al., 2005; HEATCO, 2006), it was decided that the variability of transport time in this project should be measured by the standard deviation of the travel time distribution. The main reason behind this choice was the assessment that including travel time variability in transport forecasting models would be quite difficult, and that using the standard deviation would be the easiest option. Any formulation that would go beyond the standard deviation of transport time (or its variance) would be asking too much from the national and regional freight transport models, also as they might look like in a few years from now, that are regularly used in CBA in The Netherlands.¹⁵ An extensive review on the use of the

¹⁵Nevertheless, other specifications than the one where unreliability is measured as the standard deviation of transport time, such as the scheduling model have been tried as well, to see which specification performs best on the data obtained. If a scheduling model did a better job in explaining the data, it would still be possible, under certain conditions, to calculate a standard deviation of transport time from the estimated scheduling coefficients (Fosgerau and Karlström, 2010).

standard deviation and other measures of reliability/variability can be found in Significance et al. (2012b).

b. Reliability as presented to respondents in the SP

Many respondents in passenger and freight transport cannot be expected to understand standard deviations, so reliability was presented as a series of five equi-probable transport times, described only verbally, not graphically (see Tseng et al. (2009) for a justification).

Figure 1 contains a screenshot of the original interview in Dutch, showing a choice situation in experiment 2a. ‘Vertrektijd’ is departure time. Then, we say that the respondent has an equal change on any of five transport times with corresponding arrival times. The bottom two attributes are usual transport time and transport cost.

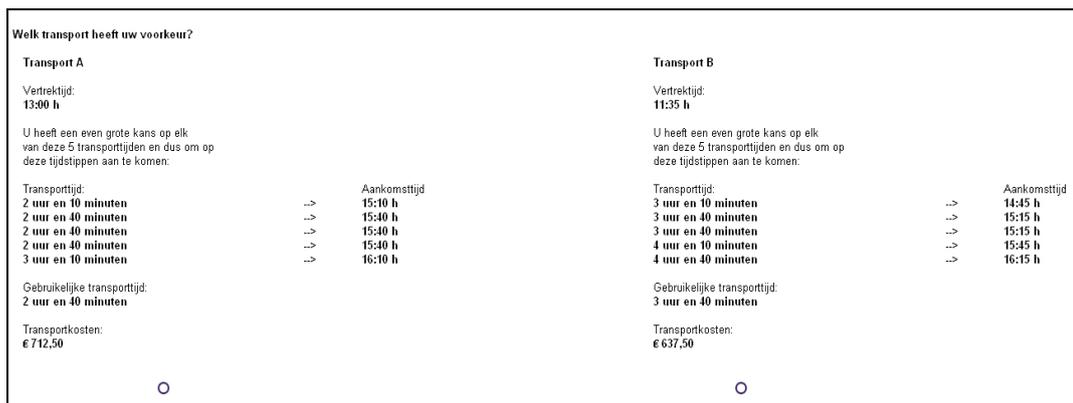


Figure 1: Screenshot of an SP question of experiment 2a for shippers and carriers (excluding carriers using sea and inland waterways)

3. The SP survey

a. Why use SP data here?

In the survey design report (Significance et al., 2007) the following motivation for using SP in this context was given:

‘...For the valuation of reliability in passenger transport and in freight transport, the use of stated preference (SP) data is a logical choice. The estimation of a model that includes a reliability variable on RP data is only possible in exceptional circumstances. In RP data (e.g. for different time-of-day periods or days), reliability, travel time and costs will often be heavily correlated, which hampers the estimation of significant separate parameter values. Furthermore, for these variables, values for the non-chosen alternatives are required, which need to come, for instance, from assignments. Here too it will be difficult to obtain sufficient variation in the variables of interest. In SP surveys, the researcher can exert control over these attributes, for all choice alternatives, and over their correlation...’.

The possibility of using logistics cost functions instead of stated preference for freight transport was discussed in the survey design stage. In principle it would be possible to determine the cost of unreliability by first calculating the costs of transport, distribution and production and then making ad hoc assumptions on the correspondence between cost items in this logistics costs function and the degree of unreliability. As key problem of this approach, researchers and the client in the survey design stage mentioned that the empirical basis is lacking for making plausible assumptions on which cost items (and to which degree) are caused by unreliability. Because of this the SP approach was regarded as looking ‘more fruitful’.

b. Design of the SP survey

The questionnaire consisted of the following parts:

1. Questions regarding the firm.
2. Selection of a typical transport and questions on the attributes of this transport, such as transport time and costs. These values are used as base levels for the attribute levels presented in the SP experiments.
3. Questions on the availability of other modes for this transport and what the attribute levels would be for that mode (as a possible basis for estimating an RP model; however no successful RP models could be estimated on these data). For the carriers this referred to a different route rather than a mode.
4. SP experiment 1 (transport time versus transport cost).
5. Introduction of variable transport times and SP experiment 2a (see below).
6. SP experiment 2b (same as 2a but without the variation in most likely arrival time).
7. Questions in which the shippers or carriers were asked to evaluate the choices they made in the experiments.

Carriers in road, rail and air transport and all shippers took part in three SP experiments (called: 1, 2a and 2b):

- Within-mode experiment 1 with six choice situations of two alternatives and two attributes: transport time and transport cost (choice alternatives with only two attributes were previously used in the SP experiments on the national passenger VTT studies of 1988/1990 and 1997/1998).
- Within-mode experiment 2a and 2b with two alternatives per choice situation, where each alternative is being described by four attributes: time, cost, reliability and arrival time (departure time is also presented, but is not independent of transport time and arrival time). In the first six choice situations, the expected arrival time is varied (see Figure 1), in the following seven choice pairs (of which one had a dominant alternative), expected arrival time is kept constant .

Experiment 1 can only give a VTT. For this two-attribute experiment, a so-called 'Bradley-design' was used. This type of design does not have any dominant questions by default. The third (out of five) levels of the both the cost and the time attribute are regarded as the base level.

Experiment 2a and 2b can give both VTT and VTTV, also distinguishing between model specifications with and without explicit scheduling terms. Experiment 2a used five attribute levels for each of the four attributes (see Significance et al., 2007; Significance et al., 2012a) and was set up as an orthogonal design (but minimising the number of possibly dominant alternatives). For the essentially three-attribute experiment 2b, a design similar to the Bradley-design of experiment 1 was used.

For transporters in sea and inland waterways transport, preliminary discussions with professionals from the sector led us to choose a different setting. The main uncertainty in transport times for these modes does not occur on a river/canal or sea link, but at locks, bridges and ports. Therefore we used an innovative setting for the SP experiments, where a ship is waiting for a lock, bridge or to be loaded/unloaded at a quay in the port. An example of an SP choice situation

is presented in Figure 2. The attributes here are five equi-probable waiting times, average waiting time (in this example waiting for a lock) and total transport cost. Note that no departure and arrival times were presented in these choice alternatives.



Figure 2: Screenshot of SP question of experiment 2b for carriers using sea and inland waterways

c. Respondents and sample size targets per segment

Table 1 summarizes the assumptions (a priori hypotheses) we make on the extent to which particular actors take into account different components of the freight value of time and variability – and should do so, when responding to our stated choice questions.

Table 11: Hypotheses on the aspects that freight respondents include in their VTT and VTTV

	VTT and VTTV related to cargo	VTT and VTTV related to vehicles and staff
Carrier	Not included	Included
Own account shippers	Included	Included
Shipper that contracts out	Included	Not included

Carriers are in the best position to give the VTT or VTTV that is related to the costs of providing transport services. If the transport time would decrease, vehicles and staff would be released for other transports, so there would be vehicle and labour cost savings.

Shippers that contract out are most interested in other aspects, as expressed by the VTT or VTTV that is related to the goods themselves. This includes the interest costs on the capital invested in the goods during the time that the transport takes (only important for high-value goods, but we did not impose a definition of high value on the respondents), the reduction in the value of perishable goods during transit, but also the possibility that the production process is disrupted by missing inputs or that customers cannot be supplied due to lack of stock. The latter two arguments are possibly more important for the VTTV.

Shippers with own account transport can give information on both the VTT and VTTV that is related to the costs of providing transport services and the VTT and VTTV that is related to the goods themselves. If both components are properly distinguished, the carrier VTT and shipper (contract out) VTT can be added to obtain the overall VTT for use in societal cost-benefit analysis (and likewise for the VTTV).

In this study VTTs and VTTVs were sought that include both components (not just the goods-related but also the services-related component), since in CBAs for transport projects in The Netherlands the user benefits of savings in vehicle and staff cost are included in the time savings of the project. Previous studies have not tried to disentangle the two components, but in the current study we obtained estimates for both components separately.

Of course there may be exceptions to the general pattern depicted in Table 11, but in the freight questionnaires we steered the shippers that contract out only to answer on the components they generally know most about (bottom-left), and likewise for carriers (top-right). We did this by giving very explicit instructions and explanations to get clearly defined component values from each type of agent. In other words, we:

1. Explained to all respondents that the changes in time, costs and reliability are generic: these apply to all carriers using the same infrastructure, and are not competitive advantages for their specific firm.
2. Explained to carriers (and logistics service providers) that a shorter transport time might be used for other transports: the staff and vehicles/vessels can be released for other productive activities. A higher reliability means that the carrier can be more certain about such re-planning/re-scheduling. Also explain that they do not have to take into account what would happen (deterioration, disruption of production process, running out of stock, etc.) to the goods if they were late.
3. Explained to the shippers that contract out that they only have to take into account what would happen (deterioration, disruption of production process, running out of stock, etc.) to the goods if the transport time or its reliability would change (whether these things would occur and how important they are was left to the respondent (shipper).
4. Explained to shippers with own account transport that they have to take all of this (=cargo and vehicle) into account.

As sample stratification variables for freight transport we used mode, (non)-container and shipper versus carrier (the latter also including logistics service providers). The target total sample size was 520 interviews.

Table 2: Target sample sizes for freight survey

		Road	Inland waterways	Sea	Rail	Air
Container	Carrier/own account shipper	50	40	20	20	
	Contract out shipper	50	20	30		
Non-container	Carrier/own account shipper	50	40	30	30	30
	Contract out shipper	50	20	20		
Total		200	120	100	50	50

d. *Methods used for recruitment of respondents; methods of doing the interviews*

Shipper and carrier firms were recruited from existing registers of firms (e.g. from Chamber of Commerce) and approached (mostly by phone) to seek firms that were prepared to participate in

the interviews. Within the firm, we searched for the director or head of logistics or operations (at carrier firms) or head of distribution (shippers).

The subsequent interviews were carried out as face-to-face interviews where a professional interviewer visited the firm and the questions were shown on a laptop computer.

e. Sample size obtained per segment

Table 3 shows the actual number of respondents for each of the questionnaire types (by means of different colours – see below) and for each mode. With 812 successfully completed interviews, this survey must be, together with the recent Norwegian VoT survey (Halse et al., 2010; see Annex A), one of the largest SP surveys ever carried out in freight transport.

Table 3: Number of freight respondents by (sub)segment

		Road	Rail	Air	Inland waterways	Sea	Total
Container	Carrier	35	10	0	16	18	79
	Own account shipper	10	2	0	0	0	12
	Contract out shipper	41	14	0	18	80	153
Non-container	Carrier	131	5	19	69	12	236
	Own account shipper	36	0	0	0	0	36
	Contract out shipper	162	19	44	22	49	296
Total		415	50	63	125	159	812

Note: the questionnaire types are indicated by a shading colour:

- Questionnaire type A – carrier (road, rail, air).
- Questionnaire type B – shipper that contracts out (all modes).
- Questionnaire type C – own account shipper (road, rail, air).
- Questionnaire type D – inland waterways and sea transport carriers.

4. The analysis of the SP survey

a. Data checks and processing

After having checked the data for outliers and implausible combinations of attribute values, discrete choice models were estimated on the SP data. In the model estimation, data from 724 interviews were used. Since the sample size for the smaller segments is often already quite small, we decided to keep the respondents that choose the dominated alternative (experiment 2b) in all models for freight transport. Overall, only 3.5% of the respondents are non-traders. These non-traders have been kept in the analysis.

b. Model specifications tested

The specification that worked best for carriers in road transport was the multinomial logit model in log-willingness-to-pay (logWTP) space. WTP-space means that the VTI and VTTV are estimated directly, as a single coefficient, instead of inferred from the ratio of the estimated time and cost coefficients. The models in the recent Danish, Norwegian and Swedish VTT projects in passenger transport were estimated in logWTP space, i.e. they used logarithmic utility functions

in their estimation processes (e.g. Fosgerau, 2006; Börjesson et al., 2011; Börjesson and Eliasson, 2011). In a mean-standard deviation model this gives for utility U:

$$U = \lambda \cdot \log(C + VTT \cdot T + VTTV \cdot \sigma) \quad [1]$$

where:

- λ = scale parameter
- C = transport cost
- T = transport time
- σ = standard deviation of the transport time distribution

All three SP experiments were used in a simultaneous model. To correct for repeated measurements (we have up to 19 choice situations per respondent), we used the Jackknife method.

Estimation results for one of the segments for carriers in road transport are given as an example in Table 4 (the full set of estimation results can be found in Significance et al. 2012b). We report on a mean-standard deviation model here, as was used for our final recommendations. Scheduling terms were also tried, but were usually not significant.

For the non-road models and all models for shippers we used a relative model specification, in which the attributes are measured relative to the observed levels, which differ over respondents. So, the utility of a fractional change of each attribute is estimated. However, this cannot be done for the scheduling terms (early and late), since it is not sensible to define a fraction of an arrival time. The relative mean-standard deviation model (using MNL) is:

$$U = \beta_C^{rel} \cdot \frac{C}{C_0} + \beta_T^{rel} \cdot \frac{T}{T_0} + \beta_R^{rel} \cdot \frac{\sigma}{\sigma_0} \quad [2]$$

where:

- C_0 = Base value of the travel or transport cost (BaseCost)
- T_0 = Base value of the travel or transport time (BaseTime)
- σ_0 = Base value of the standard deviation of the travel or transport time distribution.

Relative models were also used (for all the modes) in the Dutch freight VOT studies of 1992 (Hague Consulting Group et al., 1992) and 2003/2004 (RAND Europe et al., 2004) to cope with the heterogeneity in the typical transports in the SP data. When estimating models on the 2010 SP data we again found that these models performed best for almost all segments. An example of the estimation results for one of the relative models is given in Table 5 (see Significance et al. (2012a) for the full set of estimation results).

Table 4: Estimated coefficients and t-ratios (in brackets) for MNL logWTP model for carriers and own account shippers in road transport, non-container, cargo weight between 2 and 15 tonnes

Segment	Road non-container 2 - 15 tonnes		Road non-container 2 - 15 tonnes Jack-knife	
Experiments	1, 2a and 2b		1, 2a and 2b	
Observations	1170		1170	
Respondents	65		65	
Final log (L)	-683.6		-683.6	
D.O.F.	3		3	
Rho ² (0)	0.156		0.156	
	Value	(t-ratio)	Value	(t-ratio)
Lambda (Cost)	-8.938	(-10.8)	-8.747	(-6.2)
VTT	19.14	(3.7)	18.49	(2.6)
VTTVR	30.66	(4.6)	29.62	(2.6)
	Derived value		Derived value	
Reliability Ratio	1.60	(2.8)	1.60	(1.8)

Note:

- Lambda is the scale parameter.
- VTT is the monetary value of a change of one hour in transport time, in Euro per movement.
- VTTVR is the monetary value of a change of an hour in the standard deviation of transport time, in Euro per movement.
- The Reliability Ratio (RR) is the value of reliability (measured as the standard deviation) divided by the value of transport time.

Different characteristics of the shipment were tried as interaction variables (e.g. commodity type, value density), both for the relative models and for the LogWTP models for carriers in road transport, but these did not provide a clear pattern, presumably due to the still limited number of observations. Models distinguishing between modes, container/non-container, shipment weight and shipper/carrier performed best.

Table 5: Estimated coefficients and t-ratios (in brackets) for relative MNL model for shippers that contract out (non-container)

Segment	Shippers non-container		Shippers non-container Jack-knife	
Experiments	1, 2a and 2b		1, 2a and 2b	
Observations	4482		4482	
Respondents	249		249	
Final log (L)	-2623.7		-2623.7	
D.O.F.	5		5	
Rho ² (0)	0.155		0.155	
	Value	(t-ratio)	Value	(t-ratio)
BetaCost Relative	-7.026	(-15.3)	-6.992	(-13.1)
BetaTime Relative	-0.709	(-3.2)	-0.706	(-2.7)
BetaReliability Relative	-0.639	(-8.1)	-0.634	(-5.7)
Scale experiment 1	0.558	(8.5)	0.556	(8.5)
Scale experiment 2b	1.293	(11.2)	1.293	(9.2)
	Derived value		Derived value	
Trade-off ratio time vs cost	0.101	(3.3)	0.101	(2.8)
Trade-off ratio reliability vs cost	0.091	(9.6)	0.091	(6.6)
Trade-off ratio reliability vs time	0.901	(2.8)	0.898	(2.2)

Note:

- Relative Cost: impact of a change in cost (relative to base cost) on utility.
- Relative Time: impact of a change in time (relative to base time) on utility.
- Relative Reliability: impact of a change in reliability (relative to base reliability) on utility.
- t-ratio scale coefficient is with respect to zero.

The ratio of the estimated reliability coefficient to the estimated cost coefficient in a relative model can be treated as a ‘**trade-off ratio**’ (TR) that indicates how relative changes in reliability are traded off against relative changes in costs.

$$TR = \frac{\beta_R^{rel}}{\beta_C^{rel}} \quad [3]$$

By multiplying this ratio by the transport cost per hour for a mode (or vehicle type within a mode), the so-called ‘factor costs’, we obtain the VTTV (and similarly the VTI):

$$VTTV = TR \cdot FactorCost \quad [4]$$

These factor costs were made available by the Ministry (NEA, 2011) and used in our project in combination with the new SP estimates.

More sophisticated models than the above, such as models that include prospect theory effects (Kahneman and Tversky, 1992), mixed logit and latent class (the latter worked very well in this project for the passenger data, see Significance et al., 2012a), did not lead to stable results for freight transport. Probably, even though we have a large database compared to most other SP surveys in freight, our sample is still too small to move beyond variants of MNL.

c. Results obtained for VTTV

On the basis of the estimation results of the 2010 SP models (adding carrier and shipper components) and the external data on the factor cost, new VITs and VITVs per vehicle and vessel were calculated. Results for the trade-off ratio, distinguishing between shippers and carriers, are in Table 6, and the final results for the VTTV in Euro per vehicle or vessel per hour are in Table 7.

Table 6: Partial value of reliability (in 2010 euro/shipment/hour) and trade-off ratios (TR) for reliability versus cost

		Road	Rail	Air	Inland waterways	Sea
Container	Carrier / own account shipper	[2-40t truck]: 0	[full train]: TR=0	Not applicable	[ship waiting for a quay]: TR=0 [ship waiting for a lock/ bridge]: TR=0	[ship waiting for a quay]: TR=0
	Shipper*	[All]: TR=0.06	[All]: TR=0.06	Not applicable	[All]: TR=0.06	[All]: TR=0.06
	Total	[2-40t truck]: TR=0.06	[full train]: TR=0.06	Not applicable	[ship waiting for a quay]: TR=0.06 [ship waiting for a lock/ bridge]: TR=0.06	[ship waiting for a quay]: TR=0.06
Non-container	Carrier / own account shipper	[2-15t truck]: 29.62 [15-40t truck]: 0	[full train]: TR=0	[full freighter aircraft]: TR=0	[ship waiting for a quay]: TR=0 [ship waiting for a lock/ bridge]: TR=0	[ship waiting for a quay]: TR=0
	Shipper*	[All]: TR=0.09	[All]: TR=0.09	[All]: TR=0.09	[All]: TR=0.09	[All]: TR=0.09
	Total	[2-15t truck]: 29.62 + 0.09* factor cost [15-40t truck]: 0.09*factor cost	[full train]: TR=0.09	[full freighter aircraft]: TR=0.09	[ship waiting for a quay]: TR=0.09 [ship waiting for a lock/ bridge]: TR=0.09	[ship waiting for a quay]: TR=0.09
All		[2-15t truck]: 27.84 + 0.09* factor cost [15-40t truck]: 0.09*factor cost	[full train]: TR=0.08	[full freighter aircraft]: TR=0.09	[ship waiting for a quay]: TR=0.09 [ship waiting for a lock/ bridge]: TR=0.09	[ship waiting for a quay]: TR=0.09

Note:

- * Shipper = shipper that contracts out.
- The values for shipper (both for container and non-container) do not differ between modes since the final models do not have terms depending on mode.

Table 7: VTTV for freight transport (Euro/hour per vehicle or vessel, price level 2010)

	Road	Rail	Air	Inland waterways	Sea
Container	[2-40t truck]: 4	[full train]: 101	Not applicable	[ship waiting for a quay]: 18 [ship waiting for a lock/bridge]: 27	[ship waiting for a quay]: 45
Non-container	[2-15t truck]: 34 [15-40t truck]: 6 [all non-container]: 15	[bulk]: 260 [wagonload train]: 240 [all non-container]: 250	[full freighter aircraft]: 1600	[ship waiting for a quay]: 25 [ship waiting for a lock/bridge]: 25	[ship waiting for a quay]: 110
All	[2-40t truck]: 14	[full train]: 220	[full freighter aircraft]: 1600	[ship waiting for a quay]: 25 [ship waiting for a lock/bridge]: 25	[ship waiting for a quay]: 60

Notes:

- All these values are combined values from shippers and carriers and were obtained after rounding off.
- The values for rail are for a train (not a wagon).
- The values for inland waterways and sea refer to a ship.

The main outcomes for the VTTV in freight transport are:

- The VTTV is mainly due to shippers (cargo-related); most carriers have no significant VTTV.
- The Reliability Ratio, RR (value of reliability relative to value of time), is between 0.1 and 0.4, depending on the mode (the higher end is for road transport).

d. Discussion of the VTTV results

For shippers using road transport, Halse et al. (2010) obtain a reliability ratio of 1.3 (where we found a weighted average of 0.9). For the carriers VTTV, Halse et al. (2010) found estimated coefficients for the standard deviation that were not significantly different from zero (however, the valuation of delays in a different SP experiment within the same survey was somewhat higher than the VTTV). If we would add a carrier's VTTV of 0 to the shipper's VTTV and divide this sum by the summed VTTV of the shipper and the carrier (where the carrier component is substantially larger than the shipper component), we would get an overall RR for Norway that is much smaller than 1. In Fowkes (2006) values of 0.19 and 0.34 for the overall RR are reported from SP studies carried out in the UK. We conclude that the overall RRs that we now find are substantially lower than the earlier value of 1.2 (from de Jong et al., 2009), but that this value was based on many assumptions and that the few available empirical values in the literature are also much lower than 1.2. In the current survey, unreliability, its context and its consequences were made much more explicit and the presentation format is much more suitable for measuring unreliability in terms of the standard deviation of transport time (or scheduling terms).

The impact of just-in-time deliveries and perishable commodities on the VTTV should be reflected in the shipper's component of the VTTV. This component is significant in estimation, but usually not very large in money terms. One might have expected higher values for this component to reflect the popularity of just-in-time in modern logistics thinking, but the results that we obtain should also take into account that time-critical segments are still a relatively minor

part of all freight transport (unless we would measure transport in terms of the value of the cargo shipped).¹⁶

The carrier component of the VTTV has to do with the impact of reliability on being able to use vehicles and services for other transports. For this effect we find a coefficient that is not significantly different from zero (except for road non-container, 2-15 ton). This could be due to the small samples that we had to use in estimation and therefore we have to be careful in interpreting and using these results. In principle carriers could take into account that they could lose customers if their transport reliability became worse, but in our freight SP experiments, the changes in reliability are presented explicitly as things that happen to all carriers, so there are no competitive advantages or disadvantages here.

We expect that for carriers the sum of VTT and VTTV will not exceed the total transport costs per hour. The reason for the benefits is that in the presence of time and reliability gains they can use their vehicles and staff for other transports. Their total benefits can never be higher than the transport costs per hour, otherwise they should have been carrying out those other transports instead of the current ones. It may happen that a carrier cannot reap the full benefits from a transport time gain, because there is uncertainty about the transport times.¹⁷ As soon as this uncertainty is reduced (transport times become more certain, as do transport time gains), there could be benefits for the carrier, but according to our estimation results for most segments this is a small and not-significant effect. The VTTV for road-container is lower than for non-container.

This may have to do with the fact that the VTT for road-container is relatively high, leaving less room for a high VOR. The only segment for which we find a significant carrier VOR is road transport 2-15 ton. These vehicles are often used for urban distribution, where the uncertainty of travel times is large due to the heavy congestion in cities. Moreover, the planning patterns for these smaller freight vehicles are often quite complex, unlike those for larger road vehicles and other modes that are not so easily disturbed.

5. Current practice in The Netherlands

We expect that the new VTTs and VTTVs will become the official transport CBA values in August 2013.

At the moment, reliability benefits are usually calculated as simply 25% of the travel time (passengers) or transport time (freight) benefits, following a literature review on this topic and recommendation by CPB (Besseling et al., 2004). More differentiated, though still preliminary, guidelines for the VTTV have been available since 2004/2005. These values have been used in very few studies, because of the difficulty to predict changes in reliability and the impact of a

¹⁶The relatively small monetary values that we find for reliability seem to contradict surveys among shippers that found that reliability is the most important non-cost factor in mode choice (e.g. NERA et al., 1997). These studies however usually compare reliability to scheduled time, not to expected time (as we do), which will be more relevant if this often deviates from scheduled time (and then some of the value of unreliability will transfer to the value of expected transport time). More generally, a ranking study that finds reliability at the top of the list of non-costs attributes provides considerably less information than a stated preference study that gives a value of unreliability in money or transport time equivalents.

¹⁷A way this could be investigated would be to do case studies with qualitative in-depth interviews with firms involved in freight transport: how do they cope with the current uncertainties in transport time, and what might happen if there were shorter transport times, more reliable transport times or both?

project on reliability (only some prototype forecasting models are available to do this). Most project appraisals have used the 25% surcharge on the time benefits for the reliability gains.

The new VITVs can also only be used in conjunction with forecasts of how infrastructure projects influence variability (in The Netherlands this is called the 'Q-side of reliability'). The development of such models for passenger and freight transport is planned, but still needs to be carried out. Some models explaining variability of transport time in road transport (passengers and freight) have already been estimated in The Netherlands on speed data from induction loop measurements on motorways (Kouwenhoven et al., 2005; Tu, 2009; Hellinga, 2011; Peer et al., 2012). These models deal with variability not at the link level, but at the route level (at least the part of the routes that takes place on motorways). An overview of these studies and other studies in Europe and the US on forecasting models for variability can be found in Significance et al. (2012b).

References

- Börjesson, M. and J. Eliasson (2011) Experiences from the Swedish value of time study, CTS working paper, Centre for Transport Studies, Royal Institute of Technology, Stockholm.
- Börjesson, M., J. Eliasson and J.P. Franklin (2011) Valuation of travel time variability in scheduling versus mean-variance models, Centre for Transport Studies, Royal Institute of Technology, Stockholm.
- Fosgerau (2006) Investigating the distribution of the value of travel time savings, *Transportation Research Part B*, 40 (8), 688-707.
- Fosgerau, M. and Karlström, A. (2010) The value of reliability, *Transportation Research B*, 44(1), 38-49.
- Fowkes, A.S. (2006) The design and interpretation of freight stated preference experiments seeking to elicit behavioural valuations of journey attributes, ITS, University of Leeds.
- Hague Consulting Group, Rotterdam Transport Centre and NIPO (1992) De reistijdwaardering in het goederenvervoer, rapport hoofdonderzoek, Rapport 142-1 voor Rijkswaterstaat, Dienst Verkeerskunde, HCG, Den Haag.
- Halse, A., H. Samstad, M. Killi, S. Flügel and F. Ramjerdi (2010) Valuation of freight transport time and reliability (in Norwegian), TØI report 1083/2010, Oslo.
- Hamer, R., G.C. De Jong, and E.P. Kroes. (2005) The value of reliability in Transport – Provisional values for the Netherlands based on expert opinion, RAND Technical Report Series, TR-240-AVV, Netherlands.
- HEATCO (2006) Developing Harmonised European Approaches for Transport Costing and Project Assessment, Deliverable 5, Proposal for harmonized guidelines. IER, University of Stuttgart.
- Hellinga, B. (2011) Defining, Measuring, and Modelling Transportation Network Reliability. Final report, Delft University of Technology, the Netherlands.
- Jong, G.C. de (2008) Value of freight travel-time savings, revised and extended chapter for Handbooks in Transport, Volume 1: *Handbook of Transport Modelling* (Eds: D.A. Hensher and K.J. Button), Elsevier.
- Jong, G.C. de, M. Kouwenhoven, E.P. Kroes, P. Rietveld and P. Warffemius (2009) Preliminary monetary values for the reliability of travel times in freight transport, in *European Journal of Transport and Infrastructure Research*, Issue 9(2), pp. 83-99.
- Kahneman, D. and A. Tversky (1992) Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty*, 5 (4), 297-323.

Kouwenhoven, M.L.A., A. Schoemakers, R. van Grol and E.P. Kroes (2005) Development of a tool to assess the reliability of Dutch road networks. Proceedings of the European Transport Conference – 2005, Strasbourg.

McKinsey (1986) Afrekenen met files, McKinsey & Company, Amsterdam.

NEA (1990) Rekening Rijden en het goederenvervoer over de weg, Rapport 90183/12526, NEA, Rijswijk.

NEA (2011) Kostenbarometer goederenvervoer, rapport voor DVS, NEA, Zoetermeer.

NERA, MVA, STM and ITS Leeds (1997) The potential for rail freight, A report for the Office of the Rail Regulator, ORR, London.

Peer, S., Koopmans, C. and Verhoef, E.T. (2012) Prediction of travel time variability for cost-benefit analysis, Transportation Research Part A pp. 79-90.

RAND Europe (2004) De Waardering van kwaliteit en betrouwbaarheid in personen- en goederen vervoer (The valuation of quality and reliability in passenger and freight transport). AVV/RAND Europe, Rotterdam.

RAND Europe, SEO and Veldkamp/NIPO (2004) Hoofdonderzoek naar de reistijdwaardering in het goederenvervoer, rapport TR-154-AVV voor AVV, RAND Europe, Leiden.

Significance, VU University Amsterdam and John Bates (2007) The value of travel time and travel time reliability, Survey design, Final Report prepared for the Netherlands Ministry of Transport, Public Works and Water Management, Significance. Leiden.

Significance, VU University, John Bates Services, TNO, NEA, TNS NIPO and PanelClix (2012a) Values of time and reliability in passenger and freight transport in The Netherlands, Report for the Ministry of Infrastructure and the Environment, Significance, The Hague.

Significance, Goudappel Coffeng and NEA (2012b) Erfassung des Indikators Zuverlässigkeit des Verkehrsablaufs im Bewertungsverfahren der Bundesverkehrswegeplanung: Schlussbericht, Report for BMVBS, Significance, The Hague (see: <http://www.bmvbs.de/SharedDocs/DE/Artikel/UI/bundesverkehrswegeplan-2015-methodische-weiterentwicklung-und-forschungsvorhaben.html>).

Tseng, Y.Y., E.T. Verhoef, G.C. de Jong, M. Kouwenhoven and A.I.J.M. van der Hoorn (2009) A pilot study into the perception of unreliability of travel times using in-depth interviews, *Journal of Choice Modelling*, 2(1), 8-28.

Tu, H. (2009) Monitoring travel time reliability on freeways, PhD thesis, Delft University of Technology.

Part Two

New Methods for Estimating the Value of Freight Time Variability

NEW METHODS FOR ESTIMATING THE VALUE OF FREIGHT TIME VARIABILITY

Working paper

Preliminary draft: August 1, 2013

Niclas A Krüger
Inge Vierth

Abstract: This paper suggests and demonstrates three different approaches to calculate VITV for Swedish freight transports: A precautionary cost approach, a case study approach and a market based approach.

1 Introduction

There is up to date limited knowledge about the impact of policy measures and the monetary valuation of improved reliability for freight transports. This means that the benefits of reduced variability in travel time are not taken into account properly in cost-benefit analysis (CBA), for example related to infrastructure measures. The Value of Transport Time Variability (VTTV) is in comparison to the Value of Transport Time Savings (VTTTS) hardly addressed in cost benefit analysis despite the intentions in several countries.

Regardless of the organisation of the transport, the benefit of reduced transport time variability is related to the impact on costs of the sender and/or receiver of the goods. There are basically two issues here when it comes to the VTTV for an infrastructure measure: The first issue is to find the change of time variability for a certain link of the transport chain that accrues to a specific project. Little guidance is available for how changes in transport time variability due to various policy measures may be calculated. There are of course various sources for time variations during a transport from a sender to a receiver (in many cases via transshipment points like terminals and ports). The sum of these variations adds up to the total variations during the transport chain.

The second issue is to calculate a VTTV for the average goods transported on a certain link for the sender and/or receiver. The goods flows (in a base scenario and a reference scenario) are typically calculated with help of transport models. Both the Norwegian and the Swedish freight model include an aggregated-disaggregated-aggregated logistics module. This module makes it in principle possible to calculate the benefits due to less deviations in travel time by making use of the trade-off between transport costs and inventory costs. However information on demand, standard deviations for lead/transport times and stock out costs are needed as input.

When it comes to VTTV several definitions are possible to use. Reliability can be calculated as the *standard deviation* of the transport time distribution (comprising too late and too early arrivals). One reason for choosing this definition is that standard deviations are relatively easy to integrate into transport models. Reliability can also be defined as *expected delays* (excluding early arrivals) or as *risk for delays*. One question is if cancelled transports should be considered as very long delays.

Several approaches have been suggested (in Sweden and other countries) to estimate VTTV since variability itself is not directly traded in the marketplace: i) based on calculations of costs for safety stocks required because of stochastic transport times (Bruzelius, 1996; Minken, 1997) ii) based on revealed preference (RP) techniques and iii) based on stated preferences (SP) techniques. The safety stock approach has never been implemented empirically to derive values

for VTTV. RP-techniques have failed due to lack of data. Up to now, only SP-techniques have been employed to derive monetary values.

This paper aims at improving the calculation of the cargo-related benefits related to improved reliability in CBA. Main focus is on door-to-door-transport chains that include rail, as the rail causes most reliability problems in Sweden (in contrast to other countries).

In the paper different methods are applied to calculate the cargo-related VTTV for Swedish rail freight transports. In section 2 we present and use the so-called safety stock approach, which is a type of precautionary costs approach. Assuming a required (average) service level, measured variability for Swedish freight transports on rail and data for volume and value of goods transported in Sweden, we derive an estimate of VTTV. In section 3 we calculate the cargo-related VTTV for the grocery company COOP that has transferred transports from road to rail. We measure the reliability that COOP faces and compute the precautionary costs that the company pays for minimizing the stockout costs and the operative costs that they need to pay in case of major delays. We calculate the VTTV also based on COOP's precautionary costs and the standard deviation for the transport time of their shuttle train. Section 4 studies how the reactions of the stock markets on the train delays can be used to compute a VTTV and shows that it is possible to discern between reliability and vulnerability problems using this approach. Section 5 concludes the paper with discussions and an outline for future research.

2 Precautionary cost approach for estimating VTTV

2.1 The relationship between precautionary, operational and stockout costs and VTTV

We distinguish three different kind of costs associated with the variation in transportation time (in most cases delays): precautionary costs in order to avoid or mitigate the consequences of delays in general, operational costs that are caused by measures that have to be taken in case of delays and stockout costs.

Precautionary costs arise when measures are taken in advance that mitigate the consequences of delays and therefore have insurance like characteristics. The specific precautionary costs we examine in this section is keeping a safety stock, so that if goods arrive too late at the receiver, there are still enough units on stock to satisfy the demand in production or sales, so that these processes do not come to a halt. Another type of stockout costs can occur when a producing firm runs out of inputs and can't continue the production process.

Of course, keeping a safety stock is not costless, since costs for staff, physical storage capacity and capital arise. Hence, companies will try to minimize safety stock while maintaining a sufficient service level. However, inventories are not the only precautionary measure that can be taken. The acceptance of additional transport and or planning costs can be precautionary costs as well (see COOP case study in the next section). Physical agglomeration between buyer and seller of intermediate goods and the choice of more reliable (and more expensive) modes¹ to transport on road and sea instead of rail are other major precautionary costs to prevent adverse effects due to rail delay. In summary, precautionary measures might target both the probability of delays (e.g. maintenance measures), the size of delays (e.g. physical agglomeration) or consequences of the delays (e.g. insurance or safety stocks). Generally two components of uncertainty can be identified for freight transports: variations in the lead time taken between order and delivery (the transport time is part of the lead time) and variations in demand. Both can be dealt with by using precautionary measures. In this paper we assume that demand is deterministic.

Operational costs are other costs associated with delays that can be associated with a specific delay after the delay materializes. These costs can for example be the costs for the booking of other freight carriers or overtime compensation for workers.

Stockout costs are costs that arise when the inventory is exhausted. For retailers this means loss of sales (and hence less revenues) and for suppliers this means loss of revenues or fines. For producers it means that production has to stop because they run out of input, causing higher production costs, lost sales or fines.

Precautionary costs, operational costs and stockout costs are closely related. When determining the level of precautionary measures, the operational and stockout costs due to delays are taken into account. The more precautionary measures are taken, the fewer the operational costs will be later on. However, once precautionary measures are fixed, operational and stockout costs arise for every delay that occurs (for the part of delay cost not covered by the precautionary measures).

Hence, there is a trade-off between precautionary costs and future operational and stockout costs, incurring more precautionary costs now means lower future cost of variability; the company will buy insurance in form of precautionary measures as long as the marginal cost of doing so is lower than the expected benefits of saved variability costs due to the precautionary measures.

If we assume that demand is non-stochastic², that the cost of variability is a function of the standard deviation σ mitigated by the precautionary measure B (denoted as $v(B(\sigma))$, where v consists of both operational and stockout cost), and that the cost for the precautionary measure³ B is denoted by $c(B(\sigma))$,⁴ the company will minimize the expected costs of variability:

$$c(B(\sigma)) + E[v(B(\sigma))] \quad (1)$$

which yields, by rearranging the first order condition, in optimum:⁵

$$c'(B(\sigma)) = E[v'(B(\sigma))] \quad (2)$$

where $c'(B(\sigma)) > 0$ and $v'(B(\sigma)) < 0$. Equation 2 says that the last monetary weighted unit spent on mitigation of the effect of transport time variability, should be equal to the expected reduction of operational and stockout costs due to the chosen level of precautionary measure. Per definition, precautionary costs are always certain, but operational and stockout costs are always uncertain since they are future costs, hence the expectation operator in Equation 2.

Let us assume that there is an exogenous change in σ and that the company optimally adjusts B , which gives:

² Given stochastic demand for a perishable good we get the classical newsboy problem. For a non-perishable good it corresponds to a situation where we substitute the salvage cost per unit unsold with the per unit storage and depreciation costs. We focus here on stochastic transportation times.

³ It can be seen as the price paid for a delay related put-option bought by the company for each transport: the pay-off is zero as long as actual time < planned time, but has a positive pay-off as actual time > planned time. The pay-off of the put option is: $\max [f(t_{\text{actual}} - t_{\text{planned}}), 0]$ Hence, the company pays a price for a put-option now to insure it against future delay related downside risk. A central insight from option theory is that the option price is increasing with respect to variability.

⁴ It would also be possible to model the problem as $c(\sigma(B))$, that is, the company can instead of mitigating the effects of variability try to change the level of variability, for example by increased maintenance.

⁵ The expectation operator E is an integral and the unconditional expectation is an integral over an infinite interval. Hence, a change in σ will not change the integration interval. Thus the Leibniz rule for differentiation under an integral in this case implies that $E'[v(\sigma)] = E[v'(\sigma)]$.

$$c'(B) \cdot B'(\sigma) = E[v'(B) \cdot B'(\sigma)] \quad (3)$$

Since $B'(\sigma)$ is deterministic, we have that:

$$[c(B(\sigma))]' = E[v'(B)] \cdot B'(\sigma) \quad (4)$$

Since σ changes, delay cost in total will necessarily change, but the company can adapt by adjusting B . Note that the right hand side of Equation 4 can be interpreted as $VITV$ for a marginal change in σ , as it is the change of expected delay costs with respect to changes in variability (given that the precautionary measures are adjusted optimally).

This implies that $VITV$ can be estimated using the left hand side, that is, $VITV$ for a small reliability improvement is equal to the marginal costs for precautionary measures in optimum, given the estimated variability σ for either a specific combination of route and mode (conditional $VITV$) or for the transportation network in general (unconditional $VITV$).

Whereas future marginal costs of variability are difficult to estimate, we will show in the next section a feasible way to estimate the marginal cost of precautionary measures for a given variability in the transport system.

2.2 Aggregating a disaggregate precautionary cost model: assumptions and model

In the previous section we derive the optimal precautionary level for a single company. In order to derive a $VITV$ valid for the Swedish transport system, we built a virtual company called Sweden Freight AB (SFAB), consisting of all senders, receivers and carriers of freight in Sweden. The basic idea is that in order to calculate an unconditional $VITV$, averages of key aggregate figures are sufficient as input and that there is no need to study micro level data on inventories, transport times, order size, mode choice etcetera. Since the division of delay costs in the supply chain can freely be determined by contractual arrangements and thus can be considered vertically integrated, SFAB is the horizontal sum of vertically integrated companies. In order to avoid double-counting, we assume that all variability related costs accrue to the receivers of the goods. Naturally, SFAB has to deal with the variability of the Swedish transports as a whole, over all possible modes and routes. In the previous section we saw for a given variability how much companies have to invest in precautionary measures in order to provide a sufficient service level.

Let us assume that holding a (virtual) safety stock is the only measure for SFAB to deal with σ . The safety stocks approach rationale is as follows: A representative company (in this case SFAB) calculates a required service level α defined as the following ratio for a given period (see p. 320-324 in Hansmann, 2006)⁶:

⁶ We assume that on time deliveries include early deliveries, hence that the cost of early delivery are negligible compared to cost of very late delivery.

$$\alpha = \text{number of on time deliveries} / \text{total deliveries.} \quad (5)$$

Based on this defined service level the company reacts to a stochastic delivery time t :

$$t = E[t] + e \quad (6)$$

by holding a safety stock, which is a function of standard deviation (or more general: uncertainty) in transport times *ceteris paribus*:

$$B = f(\sigma) \quad (7)$$

The cost for holding a safety stock and hence the cost of variability in transport times is the cost of physical storage of the goods and the capital costs of the goods stored. Hence we can compute the societal cost of variability (captured by the cost for SFAB) under certain simplifying conditions. Based on (4), $1-\alpha$ is the probability of a stockout. If I denotes number of parts in inventory and q denotes the demand for parts per time unit, we will have a stockout if $t > I/q$. Hence, we can write:

$$1 - \alpha = p(t > I/q). \quad (8)$$

Substituting for t and rearranging, gives:

$$1 - \alpha = p(e > I/q - E[t]). \quad (9)$$

If F denotes the cumulative distribution function, we can write the probability for a stockout as:

$$p(e > I/q - E[t]) = 1 - p(e \leq I/q - E[t]) = 1 - F(I/q - E[t]) \quad (10)$$

Comparing (7) and (6), we can see that:

$$\alpha = F(I/q - E[t]) \quad (11)$$

If we assume that the stochastic transport time is $e \sim N(0, \sigma)$, we have:

$$\alpha = N((I/q - E[t]) / \sigma) \quad (12)$$

The inventory I has to be chosen so that the α -percentile equals $I/q - E[t]$. The α -percentile can be written as $z_\alpha \sigma$ (since the mean is zero), where the multiple z_α is determined by α . Hence we get for I :

$$I = E[t]q + z_\alpha \sigma q \quad (13)$$

The right hand side of Equation (10) consists of two parts: $E[t]q$ is the inventory needed during expected transport time, whereas $z_\alpha \sigma q$ is the inventory needed to cover the stochastic time component so that the service level α is ensured.

Hence, we have shown that if transport time is normally distributed, this means that the safety stock required is as follows (z_α is the value of a given percentile α for the standardized normal distribution, σ is the standard deviation in time units and q is the demand in production per time unit):

$$B = z_\alpha \cdot \sigma \cdot q \quad (14)$$

The cost c for holding a safety stock can therefore be written as:

$$c(B(\sigma))=f(B)=f(z_\alpha \cdot \sigma \cdot q) \quad (15)$$

The marginal cost for a small per unit change of σ is:

$$[c(B(\sigma))]’= c’(B) \cdot B’(\sigma)= z_\alpha \cdot q \cdot f’(B) \quad (16)$$

where $f’(B)$ denotes the marginal storage and capital cost per unit. If $f’(B)$ is a constant k , we have for the change in of precautionary costs:

$$[c(B(\sigma))]’= z_\alpha \cdot q \cdot k \quad (17)$$

By comparing Equation (17) and Equation (4), we can see that $z_\alpha \cdot q \cdot k$ is an estimate of VTTV.

2.3 Combining data on variability, freight flows and inventory cost with precautionary cost model to calculate VTTV

A recent study (Krüger, Vierth & Fakhraei Roudsari, 2013) shows that freight train delays in Sweden are characterized by concentration in size, space and time. The 20 per cent largest delays make up about 74 per cent of total delay time. More than 50 per cent of the total arrival delay per year occurs in just seven per cent of the (final) stations. Considering the temporal scale, there seems to be a regime of extreme days differing from the rest of the days. However, they find only a weak link between capacity utilization and average delays. They estimate based on rail freight transport data for 2009, the standard deviation being 76.5 minutes (1.275 hours). We assume that road and sea have negligible reliability problems, since there is no congestion problem on inter-urban road transport (which is a difference to many other European countries).

In 2009 230 million tonnes goods were sent to destinations inside and outside Sweden and arrived from abroad (Trafikanalys, 2010), of these ca. 13 per cent (25,900 thousand tonnes goods) were wholly or partly transported by rail.⁷ Assuming that production and sales are uniformly distributed across 250 days of the year, it implies that 104 thousand tonnes are transported any day on rail or 8.63 thousand tonne per hour (assuming 12 hours per day since freight on rail is mainly transported night time).

If the production is solely based on just-in-time deliveries in the whole economy, the required service level should be about 99 per cent. Many industries use production processes which have a considerable lower required service level. Hence, we assume an average required service level for SFAB of 90 per cent. Based on Equation (10), the safety stock can thus be computed as:

⁷ The rail transport volume of ca. 25,900 thousand tonnes in the Commodity Flow Survey 2009 (Trafikanalys, 2010(a)) differs from the information in the rail statistics 55,436 thousand tonnes 2009 (Trafikanalys, 2010(b)). The volume in the CFS misses volumes from certain sectors while the volume in the rail statistics can include double counts (E-mail 18 June 2013 from Fredrik Söderbaum, Trafikanalys). This means that the volume of ca. 25,900 thousand tonnes can be seen as an lower bound.

$$B = z_{90} \cdot \sigma \cdot q = 1.2816 \cdot 1.275 \cdot 8630 = 14,107 \quad (18)$$

If the standard deviation is 1.275 hours, in 90 per cent of all cases the deliveries will exceed the expected transportation time with at most 1.63 hours. The safety stock is what is needed in production during the 1.63 hours of delay, roughly 14.1 thousand tonnes, since we need 8.63 thousand tonnes per hour as input in production or retail.

If we assume that the average cost to store one tonne is SEK 5,068,⁸ this yields SEK 14.1 million in annual extra storage costs in the economy. Based on Equation (13), we can calculate the marginal safety stock costs with respect to σ as:

$$[c(B(\sigma))]' = z_{\alpha} \cdot q \cdot k = 1.275 \cdot 8670 \cdot 5068 = 55.8 \text{ MSEK} \quad (19)$$

A reduction of the standard deviation in transport time would thus save 55.8 million SEK. Per transported tonne (assuming 25,900 thousand tonnes 2009), this corresponds to ca. 2 SEK per tonne-hour. This figure has to be seen as first preliminary estimate; we see a need to revisit the questions how many tonnes transported on rail and how big the firms' inventory costs are. The VTTV is calculated to be ca. 4 SEK per tonne-hour if we assume inventory costs of SEK 10 000 per tonne.

2.4 Extensions of the precautionary cost approach

From Equation 17 we can see the key value drivers of transport time variability: the amount transported by rail q , transport time variability σ (or more generally uncertainty), service level α and inventory cost k (which in turn depends on the value of goods and physical storage costs). In this section we examine how extensions related to the key value drivers affect the VTTV derived by the precautionary cost approach.

i) VTTV and appropriate risk measure: The above approach is based for ease of calculation on the assumption of normality and hence standard deviation is the appropriate risk measure. However, the skewness and kurtosis in transportation time data suggest that standard deviation is not an appropriate measure. To get more accurate results, the formula in Equation (17) should be extended to cover better risk measures for fat-tailed, skewed distributions, for example, the empirical percentile in the delay distribution that would be a worst-case stockout scenario for the supply chain.

Krüger, Vierth & Fakhraei Roudsari (2013) demonstrate this by using the following hypothetical example for a manufacturer in the vicinity of Hallsberg with a supplier in the Malmö area: it is assumed that the monthly production is 4400 units and that this implies that the

⁸ The inventory costs of 5,068 SEK are a weighted average of the costs used in the Samgods model, see Vierth, Lord & McDaniel (2009).

monthly need for a certain input is 4400 as well, if the ratio is one input unit per one unit output. If the manufacturer produces on average during 22 working days a month, it implies that there is a daily need for 200 parts. If the production is based on Just-in-time deliveries the required service level would be about 99%, that is, 99% of all deliveries should be on time in order to avoid stockout costs. Recall Equation 17 and note that σ and q have to be expressed in the same time unit; hence if the standard deviation is expressed in hours, we need to compute the input demand per hour as well. For the route Malmö-Hallsberg the authors compute from a database of all train movements during 2008 and 2009, that the average planned transport time for rail freight is roughly 6 hours and that the standard deviation is 1 hour. Assuming that the company faces this average transport time and variability, the buffer stock is computed as follows (we need to order 100 parts at each time if the company produces during 12 hours per day, that is 200/12 parts per hour):

$$B = z_{\alpha} \cdot \sigma \cdot q = 2.32 \cdot 1 \cdot 200/12 = 39 \quad (20)$$

If the standard deviation is 1 hour, in 99 % of all cases the deliveries will be within 8.23 hours, that is, at most 2.32 hours delayed. The buffer stock is thus what is needed in production during the 2.32 hours of maximum delay, roughly 39 parts (since we need 200/12 parts per hour as input in production). Thus, in total the manufacturer has to order 139 parts each time, 100 parts for production during the expected transport time (6 hours) and 39 as a buffer for the stochastic part of transport time (at most 2.32 hours).

However, the authors point out that rail transports are not normally distributed and that this affects the calculation of buffer stocks since the standard deviation is not a good measure for variability related to the tail of the distribution. Hence, we suggest that the approach is modified in future research; it should be based on empirical delay distributions instead of normal or lognormal distributions. To do this might be quite straightforward, since it is possible to express any probability distribution with fat tails by appropriately varying σ (Taleb & Haug, 2011). In our case, σ could be expressed as a function of transportation time and service-level, so that we have $\sigma(t, \alpha)$,⁹ meaning that the cost of safety inventory (Equation 15) would become:

$$B = z_{\alpha} \cdot \sigma(t, \alpha) \cdot q \quad (21)$$

According to the empirical delay distribution used in Krüger, Vierth & Fakhraei Roudsari (2013), the 99th percentile between Malmö and Hallsberg is more than 2.32 times the standard

⁹ In finance, a function $\sigma(t, \alpha)$ is called a volatility surface. Time always increases disorder and hence changes σ , e.g. we could adjust σ dependent on transportation length. And the higher the required service level, the farther we are in the tail, the more necessary it is to make an adjustment of σ . If σ would depend on one parameter only, it would be called a volatility smile (since it is often a convex function in finance).

deviation of 1 hour (that is 2 hours and 20 minutes); it is instead almost 4 times the standard deviation (4 hours). Hence, the necessary buffer stock would be for $\sigma(t,\alpha)=1.72$:¹⁰

$$B = z_\alpha \cdot \sigma(t,\alpha) \cdot q = 2.32 \cdot 1.72 \cdot 200/12 = 67 \quad (22)$$

Therefore the safety stock should consist of 67 parts instead of 39 parts, an increase by 70 percent compared to the calculation based on the assumption of normality. If inventory costs were linear, this would translate into a 70% higher VTTV (for a discussion of functional form, see below).

ii) Modal share and shadow cost of variability: The modal share for rail would be higher if reliability was higher for rail transports. Hence, instead the market share of rail might be twice as much as it is today. Hence, there is – at least for some companies – a shadow cost of transporting not on rail, since rail is cheaper than road transportation due to the exploitation of economies of scale. The above value for VTTV has thus to be adjusted to cover this cost of variability in transport time as well. A prerequisite for this is that research has to be done with regard to the potential modal share of rail (given that infrastructure capacity is available), i.e. the long term elasticity of modal split with regard to changes in reliability (*ceteris paribus*, since rail operators probably increase prices if reliability improves).

Moreover, in a similar line of argument, the average value of goods transported on rail is lower (by a factor of 3) than the average transported on roads. Less variability would probably result in a higher average value for the goods transported by rail and hence a higher VTTV.

iii) Required service-level: The derivation of VTTV is simplified, since it is based on the service-level α , that does not take into account the quantity of goods that is delayed (service-level β). Since service-level α is always smaller or equal to service-level β , a simple remedy would be to use a higher service-level α to capture part of the costs associated with service-level β . We think that it is possible to get information from companies about their required service-levels, since they are used in practice. It can be done either by questionnaires or in case studies (see Section 3) and it is probably easier than getting answers in choice experiments.

Stockout cost can also occur because of stochastic demand fluctuations. Naturally this has implications on safety inventory and other precautionary measures. Further research should therefore unveil the implications for estimating VTTV using more realistic assumptions about the relationship between stochastic transport time and demand.

iv) Precautionary measures other than safety inventory: Most goods can be stored; however, it can in some cases be expensive. Hence we will observe other measures that

¹⁰ $\sigma(t,\alpha)$ is simply the value that fits the normal distribution to the empirical distribution, in this case: $\sigma(t,\alpha)=4/2.32=1.72$ instead of $\sigma=1$.

companies use to handle the costs of transport time variability, as long as the marginal cost for this is lower than the marginal storage cost. In optimum, we would expect the marginal costs to be the same.

However, if some goods cannot be stored, the costs might differ from the marginal storage costs, and are probably bigger than the average value we use. In general, also marginal cost of storage can be higher than the average cost we use here (see below).

Many products are transported direct into the production process,¹¹ i.e. for producing firms that require that their suppliers are located close to them (so that they have no possibility to choose between different suppliers); in a sense, this is also a cost of transportation time variability. There might be also a need to construct a logistic system with redundancies (besides safety stock) which increases costs of transportation.¹²

v) Cost function and variability in transportation time: it is important to keep in mind that we are not interested in delays per se, but in a function of delays measuring exposure in terms of costs arising from variability. If the function is non-linear, the difference can be substantial. The difference $f(x)-x$ grows with the size of x if the cost function is convex, which means in the context of delays, that the difference between delays and costs of delay grows with delay size.

Moreover, if the cost function is convex, Jensen's inequality shows that $E[f(x)] > f(E[x])$. Hence, expected cost of delay cost is bigger than the cost of expected delay. The difference $E[f(x)] - f(E[x])$ grows with the degree of variability. This implies that it is not enough to value the average delay (given that a delay occurs; also called mean lateness).

We model the rational reaction to increased variability as $dC/dB \cdot dB/d\sigma$. In the safety stock approach, the term $dB/d\sigma$ is linear; the term dC/dB might be convex however, meaning that additional insurance (safety stock or other precautionary measure) possibly at some point exhibits increasing marginal costs.

Taken points i)-v) into consideration, we conclude that the value derived in 2.3 constitutes a lower boundary for VTTV. Further research on each point raised in this section is necessary to derive a VTTV which is generally applicable.

¹¹ A SSAB factory was closed for two days in January 2011 due to a rail incident, as inputs were not delivered on time.

¹² See Section 3, i.e. expensive backup solution for COOP using trucks.

3 Case-study approach for estimating VTTV – COOP

3.1 The logistic system of COOP

COOP is transporting about 3.5 million cubic metres¹³ of grocery products, refrigerated and frozen food and non-food products per year in Sweden (Håkansson, 2011) (Håkansson, 2012). The company's ware houses are located close to Stockholm: in Bro for colonial and non-food products, in Västerås for refrigerated food and in Enköping for frozen food. For incoming products from outside Sweden (Denmark, Germany, Benelux, Italy, Far East etc.) different predominantly intermodal transport chains are used.¹⁴ The port of Helsingborg in the South of Sweden is one of the main import ports. A high share of the incoming products is transported to the ware houses. The products are assembled in the ware houses and go from there to COOP.s over 700 shops in Sweden. The distance between Helsingborg and the ware houses close to Stockholm is about 600 kilometres.

Traditionally, COOP has bought all products including transport costs. However, in 2009 the company made an effort to ensure control over their transports. Today COOP pays the transport costs directly to the carriers for over 80 per cent of the incoming products. In that context COOP developed the combined road-rail transport concept in order to shift transports from road to rail.¹⁵ To ensure high reliability transports, COOP has personal that manages the transfer from rail to road when the rail transports are not performed as planned. COOP has also invested in trailers that can easily be transferred from rail to road.

Since September 2009 two shuttle trains carrying 36 trailers go five days per week between Helsingborg and Bro. The southbound train stops has a stop in Alvesta. The ware house in Bro has, via an industrial track, access to the rail network.¹⁶ The southbound train is part of an intermodal transport chain, i.e. the trailers are transferred to trucks and transported to the shops in Småland (from the railway station in Alvesta) and Skåne (from the railway station in Helsingborg).¹⁷

¹³ Ca. 9,4 million tonnes (1 cubic metre is ca. 250 – 285 kg)

¹⁴ Circa 70 per cent of COOP: s transports are long haul transports over 300 kilometres, where road, rail and sea compete.

¹⁵ According to COOP: s rough estimations the rail transport costs are up to around five to ten per cent lower than the road transport costs.

¹⁶ The goods are on average about twelve days in the ware house in Bro.

¹⁷ The shops get deliveries two to five times per week.

Assuming an average of 29 loaded trailers per train¹⁸ and around 10 tonnes per trailer¹⁹ the two shuttle trains have transported around 200 000 tonnes or 74 000 cubic metres during the period 2011-06-01 to 2012-09-30.²⁰ COOP planned to use the shuttle train also to carry very time sensitive fresh fruit and vegetables that have to be in the shops before the shops open. A trial was started but discontinued after two major breakdowns. (Today these transports are performed by truck.)

3.2 Transportation time variability data

COOP has kindly provided us with data that describe the performance of the two shuttle trains between Helsingborg and Bro between 1 June 2011 and 30 September 2012. During that 16-month-period both the northbound and the southbound train were cancelled twice due to the storm Berit in the end of November 2011.

The northbound trains arrived on 53 occasions (15 per cent) more than one hour too late in Bro. The southbound train was on 27 occasions (8 per cent) more than one hour too late in Alvesta and on 48 occasions (14 per cent) more than one hour late in Helsingborg. Delays over one hour cause major problems for the ware house in Bro and especially the shops in the South of Sweden. Major delays in Helsingborg or Bro cause also problems for the rail operator as the shuttle train cannot turn around as planned. (As stated above, we focus on the cargo-related component of the VTIV in this study). Figure 1 shows that a lot of trains arrived before schedule – up to nearly two hours. COOP does not experience problems related to too early trains.

¹⁸ From 1 June 2011 to 31 December 2011 the average load factor was 78 per cent for the northbound train and 80 per cent for the southbound train.

¹⁹ The exact figure needs to be confirmed by COOP.

²⁰ Bread and milk products, that stand for high volumes of the products sold in COOP: s shops are transported in separate systems.

Figure 1: Deviations planned-actual time (minutes) for COOP's shuttle train going north

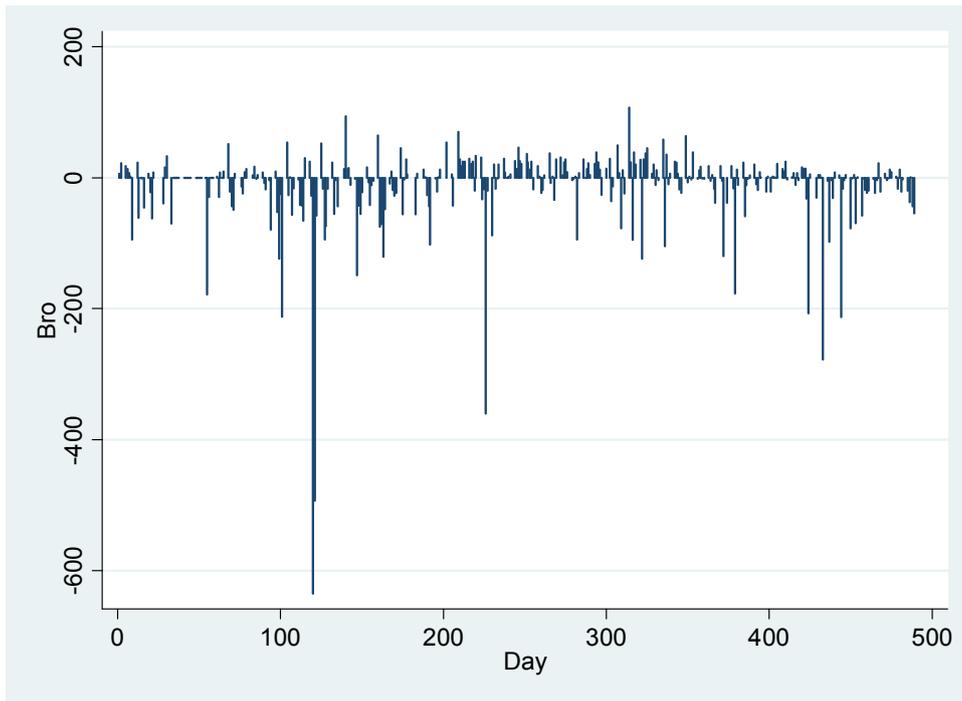


Figure 2: Deviations planned-actual time (minutes) for COOP's shuttle train going north

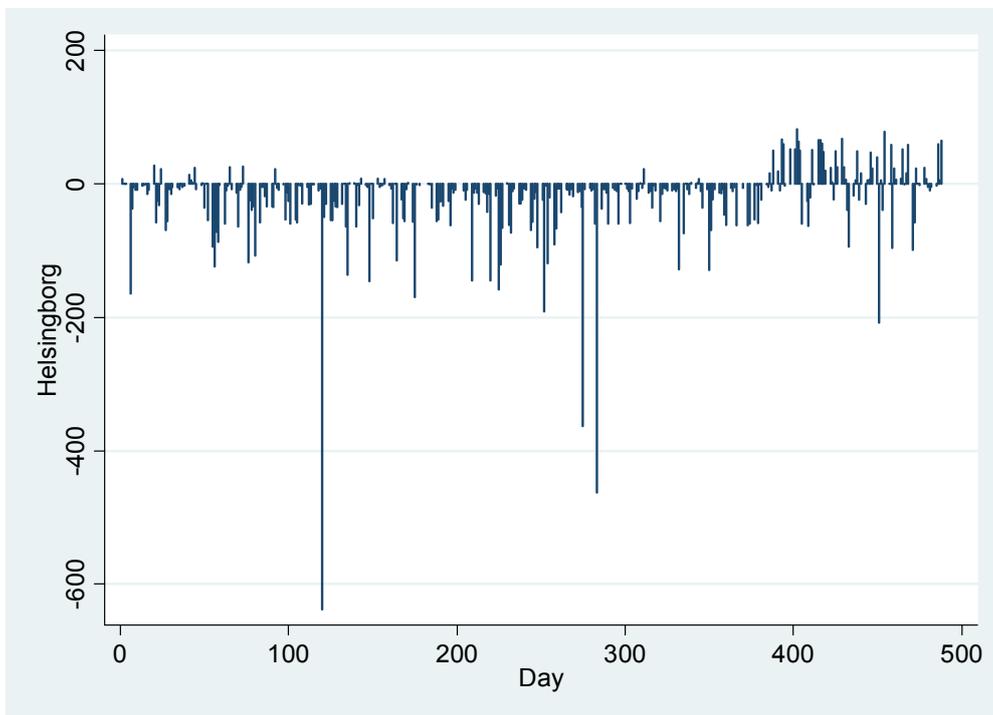


Table 1 shows the mean and the standard deviation for the three stations that the shuttle train serves (only the southbound train stops in Alvesta). The mean exclusive of early arrivals is the conditional expectation (CE) given that we have a delay. The standard deviation exclusive early

arrivals is the so-called lower partial standard deviation (LPSD) or the square root of the semi-variance. It uses only negative deviations from the time-table, squares those deviations to obtain a semi-variance, and then takes the square root to get a standard deviation for the left-tail. The LPSD is therefore the square root of the average squared deviation, conditional on a negative outcome (late arrival). The LPSD captures however not the frequency of delays.

As can be seen, the standard deviation is about one hour for both the northbound and the southbound train.

Table 1: Mean and standard deviation (excl. cancellations) for COOP’s shuttle train

	Alvesta	Helsingborg	Bro
Average delay: mean (minutes incl. early arrivals)	-8.5	-23.5	-13.7
Average delay: mean (minutes excl. early arrivals)	-52.6	-38.3	-48.6
Standard deviation (minutes incl. early arrivals)	61.4	62.2	64.7
Standard deviation (minutes excl. early arrivals)	86.3	63.0	79.9

The figures in Table 1 illustrate that the average delays (including too early arrivals) are relatively small (9 to 24 minutes), the average delays (excluding too early arrivals) are more substantial (38 to 53 minutes). Hence, it is necessary to study the real distribution of the delays (see Figure 3) and the risks for large delays that shippers like COOP face. The standard deviation (including early arrivals) for the transport time of COOP:s shuttle train is lower (on average 63,3 minutes) than the standard deviation for freight trains in Sweden (76,5 minutes, see 2.3). The skewness is -4.65 going south and -4.94 going north (the normal distribution has skewness of zero) and the kurtosis is 39.55 going south and 39.46 going north (the normal distribution has a kurtosis of 3).

Figure 3: Histogram for COOP’s shuttle train transportation time deviations

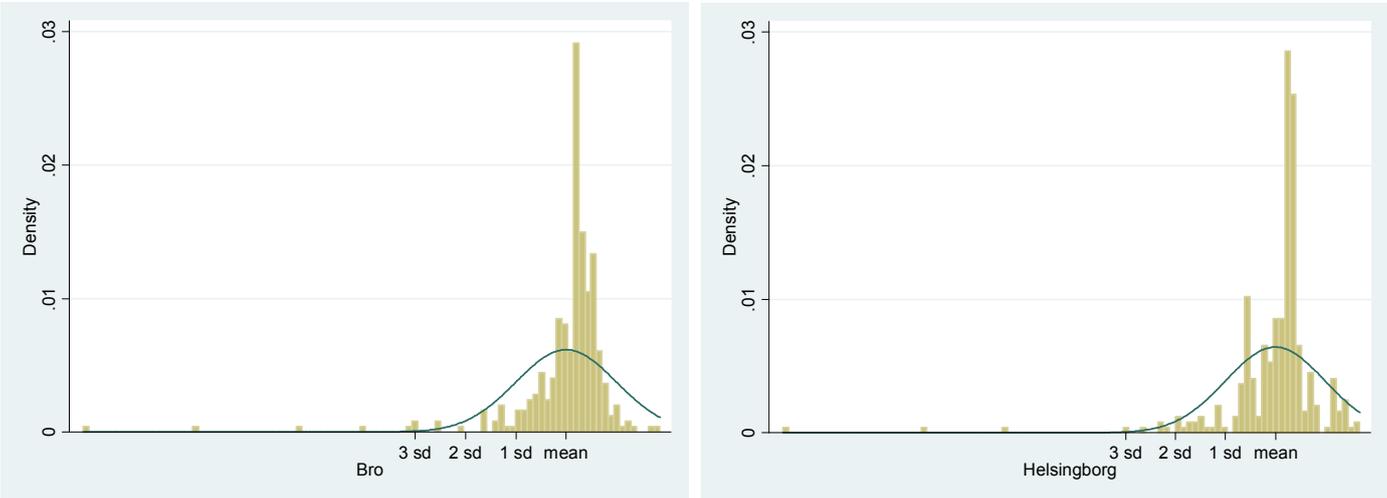


Table 2 show that the average delay in the 90th percentile was around 2.5 hours and that the 10 percent worst delays contributed to more than half of the total delay going south and to almost two thirds of total delays going south. In a sense, Table 2 summarizes the transportation risk for COOP in numbers. If we take transportation south to Helsingborg as an example, in 5 % of all transportations, the delay is almost 2 hours (118.9 minutes) or more according to the percentile value. That is, once per 20 transportations a delay of this size or larger is expected to happen southwards and since there are 2 trains per day, it means once every 10 days or 3 times per month. The average delay in each percentile is the conditional tail expectation (CTE). For example, going north to Bro there is a 10 % probability of a delay of at least one hour. The average within the percentile is close to 2.5 hours, so that with a 10% probability (once every 10 transportations or once every 5 days) COOP has to expect a average delay of 2.5 hours. Hence, the CTE gives a better estimate of how big the risk is in the tail of the delay distribution.²¹ Whereas the percentile values mark the best outcome of any given worst case scenario (e g 10 % worst cases), the CTE gives the average for a given worst case scenario.

Table 2: Percentile share in percentage of total delay minutes for final stations in Helsingborg and Bro

Helsingborg					Bro				
Percentile	Centile	Observations	Avg. delay	% of total	Percentile	Centile	Observations	Avg. delay	% of total
80 th	53	71	101.5	72.6	80 th	29	67	96.1	85.4
90 th	66.9	34	148.1	50.7	90 th	60.2	33	149.7	65.5
95 th	118.9	17	208.7	35.7	95 th	99.2	16	224.8	47.7
99 th	299.45	3	488.0	14.8	99 th	332.12	3	496	19.7

According to information from COOP that is based on the operator's coding for the period 1 June 2011 to 29 February 2012, most of the delays were caused by the infrastructure holder (54 per cent), the rail operator was responsible for 16 per cent of the delays and COOP as shipper for ten per cent of the delays. 20 per cent of the late arrivals were due to other reasons (i.e. weather conditions).

²¹ The centile corresponds to Value-at-risk and the conditional tail expectation corresponds to the Expected shortfall, both risk measures frequently used in finance, where the latter has gained attention since the financial crisis of 2008.

3.3 Costs to manage transportation time variability

Below we try to identify and estimate COOP's costs associated with controlling the intermodal concept that comprises the shuttle trains (that was introduced in September 2009). We present also the costs that were paid as consequence of the delayed and cancelled trains during the period June 2011 – September 2012. The calculations are based on information from COOP and own assumptions.

COOP has taken *precautionary measures* in order to cope (in advance) with the shuttle trains' reliability problems:

- Three transport planners out of four in total are due to the fact that the transfer of the trailers from rail to road, the road transports and consequent effects on the turnaround have to be managed in case of major delays of the shuttle train. We roughly estimate these additional staff costs to around $25000 \cdot 12 \cdot 1.5 = 1.35$ MSEK plus office space 0.15 MSEK. In total ca. MSEK 1.5 per year and ca. 2.0 MSEK for the 16-month-period we study.
- COOP owns in total 72 trailers²² that can easily be transferred from the shuttle train to a truck in case of a major delays. The investment costs are around SEK 200 000 higher for these transferable trailers than for standard trailers. Based on an assumed life-span of ten years we estimate the additional costs to be around $(72 \cdot 200000) / 10 = 1.44$ MSEK per year and ca. MSEK 1.9 for the 16-month-period.
- Due to the unreliability of the rail transports COOP choose trailers and not swap bodies as cargo units. Swap bodies have a lower net weight (three tonnes, they can accommodate 18 pallets) than trailers (ten tonnes, they can accommodate 33 pallets). For COOP in 90 per cent of all cases the volume of the products (and not the weight) is the limiting factor when filling cargo units. The use of trailers with a higher net weight than swap bodies leads tentatively to higher energy costs per tonne-kilometre; these extra costs are negligible for COOP.
- Since October 2011 COOP engages a rail operator for the shuttle train that uses stronger locomotives. This is expected to reduce the delays caused by leaves on the track (TransportNytt, 2011).
- COOP did not extend the safety stocks in the shops when the combined road-rail concept with the shuttle trains was introduced. The shelves in the shops (that serve as safety stocks) were assessed to be sufficiently deep and the company does not have additional ware house costs.

²² 36 trailers for each train

COOP paid also *operative costs* during the sixteen-month-period that we study:

- COOP had additional costs in the extreme delay occasions when trailers were transferred from rail to road. If we assume that this happened in the 24 occasions when the delays were over two hours and that the road transport costs for a trailer were SEK 4 500 (half of the costs for a trailer transport by truck from Bro to Helsingborg) we calculate extra road transport costs of MSEK 3.1. According to COOP no other costs fall away.
- The southbound train that serves the shops is more time sensitive than the northbound train that goes (directly) to the warehouse in Bro. If the train arrives too late at the railway stations in Alvesta or Helsingborg, the trucks reach the shops probably also too late. According to COOP, this means that the employees that pick up the products in the shops have to wait and that working hours are lost.²³ COOP assumes staff costs of about SEK 350 per hour and that it takes about one hour to pick up one rolling pallet. This implies that the staff costs were $350 \cdot 20 \cdot 29 \cdot 24 = 4.8$ MSEK if 20 rolling pallets per trailer, 29 trailer per train and 24 occasions are anticipated. If the delays exceed 24 hours COOP expects lost revenues because customers demand products that are not available in the shops. This type of costs is not relevant as the maximum delay in the period we look at is eleven hours.²⁴
- The additional staff costs in the ware house in Bro due to late arrivals of the shuttle train in 75 occasions are assumed to be much lower than for the southbound train. We assume MSEK 0.1 million for simplicity reasons.
- For the two cancelled trains COOP estimates direct costs in form of increased transport costs of around SEK 1 700 000.

The cost figures in Table 3 are based on rough assumptions and have to be taken with caution. However, they show COOP: s willingness to invest in precautionary measures and the need to pay additional costs when the trains are much delayed or cancelled. There is a trade-off between the precautionary and the operative costs; the additional road transport costs in case of the major delays would i.e. be higher without the transferable trailers (that COOP has invested in). It has also to be taken into account that the combined road/rail concept is quite new and that there are probably learning curves for all stake holders.

²³ It can also be necessary to pay overtime costs; these are not taken into account below.

²⁴ Costs due to the loss of the value of the products are assumed for transports of fresh fruit and vegetables that are late more than 24 hours.

Table 3: Rough estimates of COOP's precautionary and operative costs related to the combined road/rail concept Helsingborg – Bro for period 2011-06-01 to 2012-09-30

	MSEK	Source
Precautionary costs		
Staff costs to manage the shuttle train	Ca 2.0	COOP (including own assumptions)
Extra costs for transferable trailers	Ca 1.9	COOP, assumption life span 10 years
Operative costs		
Road transport costs (in case of transfers)	Ca. 3,1	COOP, assumption 24 occasions
Staff costs due to late arrivals in shops	Ca. 4.8	COOP, assumption 24 occasions
Staff costs due to late arrivals in ware house	Ca. 0,1	Assumption
Costs due to two cancelled trains	Ca. 1,7	COOP

Based on the information that COOP experienced ca. 400 delay hours during the period 1 June 2011 to 30 September 2012, and assuming 29 trailers per train and ten tonnes per trailer, we can calculate the following cargo-related VTTV:s:

- SEK 34 per tonne-hour delay (lower bound) based on precautionary costs
- SEK 69 per tonne-hour delay based only on operative costs for delays
- SEK 15 per tonne-hour delay based on operative costs for cancellations
- SEK 84 per tonne-hour delay based on operative costs for delays and cancellations
- SEK 117 per tonne-hour delay (“upper bound”) based on precautionary costs and operative costs for delays and cancellations

Alternatively we can arrive at a VTTV-estimate based on revealed preference data for the specific COOP case by dividing the precautionary costs by the standard deviation of the transport time. If we assume costs for precautionary measures of ca. SEK 3.9 MSEK (see Table 3), a standard deviation of around 63 minutes (incl. early arrivals, (see Table 1) or 77 minutes (excl. early arrivals, see Table 1) and that 200 000 tonnes are transported (see section 1) we compute a VTTV of SEK 18 per tonne-hour based on the standard deviation resp. SEK 15 per tonne-hour.

3.5 Mode choice

For COOP transport costs are the most important factor for the mode choice of incoming products in general. However, for (outgoing) transports to the shops also transport time and frequency are important mode choice criteria. Reliability is seen as closely related to the frequency. Low reliability is manageable if frequency is high (Vierth, 2012). The environmental aspect is also important for COOP as company.

Generally rail transports costs are estimated to be around five to ten per cent lower than road transport costs. However, it is regarded as a disadvantage that deliveries by train (containing a maximum of 36 trailers) are concentrated at one point of time and not - like deliveries by truck - spread over time. Normally damages in intermodal transport chains are more costly than in pure truck transports, since it is often not clear where in the transport chain the damage occurred.

There is a trade-off between the exploitation of economies of scale that requires high volumes (and low frequencies) and the number of departures. Due to the relatively low frequency of the train (compared to trucks) major delays and cancellations typically lead to major consequences. According to COOP a remedy would be to jointly use shuttle trains together with other grocery companies. Many food producers are located in the South of Sweden, whereas several grocery companies have their ware houses close to Stockholm. This would reduce the transport costs per unit (economies of scale can be exploited) and the frequency could be increased (if several firms start shuttle trains) and the whole (larger) system would probably be less vulnerable.

As mentioned above COOP has given up using the intermodal road-rail concept for the transports of time sensitive fresh fruit and vegetables. This means that the reliability problems in the rail system prevent the grocery company from using the cheapest mode for a share of their transports. However, the rail transport costs do probably increase when their reliability is improved.

4 Market based approach for estimating VTTV – stock market reactions on delays

4.1 Efficient Market hypothesis and informational efficiency

The efficient market hypothesis (EHM) says that all information is accurately reflected in the actual stock price. The price is unpredictable, since there is no way to use past stock price patterns to predict future stock prices. Not only information on price patterns but also all public available information is incorporated (the strong version of EHM states that even inside information is incorporated). The only relevant information is the current stock price. The best predictor is the current price:

$$E[S_{t+1}] = S_t \quad (23)$$

The only thing that moves stock prices is the steady stream of new information. Every bit of information, even very small, moves the price when transaction costs are sufficiently small. Such new information can be the sudden materialization of delays. Even if freight train delays are expected, every delay occurring is a confirmation of a prior belief and can therefore possibly move stock prices. However, unexpected delays, especially very large ones, should have an effect on the stock price, *ceteris paribus*. In the next section we set out to estimate the impact on company market value of delays (per hour) and of cases of very large delays due to transport system vulnerability.

4.2 Delay data and stock returns

We have data for daily train delays in Sweden measured in total hours of delays per day. The data span over 2008 and 2009. This leaves us with 731 observations. Since the Swedish stock market is closed on weekends and holidays we only have 509 observations with stock prices.

We do not know when and where the transportation is used. A long delay in south of Sweden might not have any impact on firms who do not transport in south of Sweden but in the north and vice versa. Large firms who rely on railroads to transport goods can also have backup plans for when there are large disturbances on the railroad making train delays irrelevant. One concern might be that firms who are dependent on getting their goods to the destination on time might use alternative, more precise, transportation already and therefore only firms who are not time sensitive are using the railroads. However, some firms use railroad since there no alternative, mainly for large heavy volumes without sea-access.

Since the train delays data are measured in 24 hours intervals we have to decide how we want to examine the stock prices. We have three main choices for measurement:

- The intraday difference: $\ln(\text{Closing } S_t) - \ln(\text{Opening } S_t)$

- The day to day difference: $\ln(\text{Closing } S_t) - \ln(\text{Closing } S_{t-1})$
- Day-after day-before difference: $\ln(\text{Closing } S_{t+1}) - \ln(\text{Closing } S_{t-1})$

Since the stock market is only open during 9 am to 5.30 pm but as the trains keep going during the whole day and we do not discern when exactly the train delays happened, we have a problem. The data would preferably be measured at least hourly intervals instead of using day to day measures. If this was the case we could more precisely be able to measure the reaction of stock prices with respect to train delays. Mondays should pose the biggest problem, since when the stock market opens Monday morning there are several days of information that the stock market has to assess.

4.3 Empirical method and results

In a first step we compute the raw correlations between stock prices and delay. We examine all major traded stocks with an exposure to the utilities and transportation sector (thus excluding, media, financial and IT stocks). We do not know the exposure of each company to rail freight transportation, but since the economy is interconnected, non-obvious links might exist. The correlation between delay and the three stock price measurements for each firm are presented in Table 4.

Table 4: Correlation stock prices and delays

Closing(t)-Opening(t)		Closing(t)-Closing(t-1)		Closing(t+1)-Closing(t-1)	
Delay t		Delay t		Delay	
Volvo B	-0,0361	Volvo B	-0,0224	Volvo B	-0,0703
Trelleborg	-0,1262	Trelleborg	-0,03	Trelleborg	-0,0751
SSAB B	-0,0532	SSAB B	-0,0258	SSAB B	-0,0661
SKF B	-0,0276	SKF B	-0,0203	SKF B	-0,0954
SCANIA B	-0,0482	SCANIA B	-0,0632	SCANIA B	-0,0936
SCA B	-0,0781	SCA B	-0,0759	SCA B	-0,1133
SANDVIK	-0,046	SANDVIK	0,0146	SANDVIK	-0,0142
OMX STHLM	0,0167	OMX STHLM	-0,0376	OMX STHLM	-0,077
ICA	0,0067	ICA	-0,041	ICA	-0,094
HUSQVARNA B	0,0733	HUSQVARNA B	-0,0082	HUSQVARNA B	-0,0728
HOLMEN B	-0,0836	HOLMEN B	-0,0928	HOLMEN B	-0,1164
STORA ENSO B	-0,104	STORA ENSO	-0,0794	STORA ENSO R	-0,1106
BOLIDEN	-0,0162	BOLIDEN	-0,0383	BOLIDEN	-0,0526
BILLERUD	-0,0373	BILLERUD	0,0064	BILLERUD	-0,025
AXFOOD	-0,0027	AXFOOD	-0,0827	AXFOOD	-0,1081
AUTOLIV	-0,0728	AUTOLIV	-0,0076	AUTOLIV	-0,061
ATLAS COPCO B	-0,0128	ATLAS COPCO B	-0,044	ATLAS COPCO B	-0,0736
ASSA ABLOY	-0,0379	ASSA ABLOY	-0,0462	ASSA ABLOY	-0,1007
ALFA LAVAL	-0,0091	ALFA LAVAL	0,0541	ALFA LAVAL	0,0328
ABB	-0,0197	ABB	-0,0002	ABB	-0,0382

As can be seen, nearly all coefficients have the expected negative sign. Hence, as delays in the transport system increase, companies with exposure to rail transportation loose market value. In order to get point estimates for the size of the per hour delay effect, we estimate by means of ordinary least squares the following relationship:

$$S_t = a + b \cdot \text{DELAY}_t + c \cdot \text{OMX}_t + d \cdot D_t \quad (24)$$

where the stock price is one of our three measures describe earlier, delay measured daily for the whole network and the Stockholm OMX30 stock index²⁵ return is measured in the same way as the corresponding stock price return. The inclusion of the general market return allows us to discern between idiosyncratic events from systematic events. D_t is a dummy variable, taking the value one if a vulnerability event occurs on day t , defined as a daily delay in the top-percentile. D_t allows us thus to examine if there is an additional effect of the very largest delays in addition to the pure cost of delay time. Table 5 shows the regression results (the result shown is based on $S_{t+1} - S_{t-1}$, the two other measures show similar results).

Table 5: Regression output

STOCKS	DELAY	OMX STHLM	D(80)	CONSTANT	R2
VOLVO B	-1.76e-05	1.485***	0.00283	0.000176	0.745
TRELLEBORG	-4.65e-05	1.228***	0.000737	0.00182	0.268
SSAB B	-3.90e-05	1.485***	0.00619	0.00255	0.584
SKF B	-4.80e-05*	1.124***	0.00446	0.00547**	0.634
SCANIA B	-2.62e-05	1.183***	-0.000332	0.00171	0.577
SCA B	-4.59e-05*	0.874***	0.00203	0.00424	0.521
SANDVIK	5.48e-05	0.00522	-0.0136	-0.00292	0.004
ICA	-2.07e-05	0.550***	-0.00179	0.00220	0.251
HUSQVARNA B	2.83e-05	1.079***	-0.00818	-0.00154	0.491
HOLMEN B	-9.45e-05***	0.743***	0.0108**	0.00702**	0.403
STORA ENSO R	-0.000111***	0.927***	0.0121**	0.00705*	0.371
BOLIDEN	2.95e-05	1.594***	-0.00499	-0.000423	0.482
BILLERUD	-6.08e-05	0.839***	0.0136*	0.00319	0.181
AXFOOD	-4.67e-05*	0.339***	0.00197	0.00385*	0.154
AUTOLIV	-2.58e-05	0.851***	0.00271	0.00260	0.340
ATLAS COPCO B	-2.51e-05	1.356***	0.00316	0.00334	0.677
ASSA ABLOY	-3.37e-05	0.979***	0.000668	0.00444	0.541
ALFA LAVAL	0.000126	1.220***	-0.00460	-0.0162	0.126
ABB	-1.74e-05	1.145***	0.00657	0.000228	0.596

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

As can be seen, only some coefficients for delay are significant. In the next section, we analyze the economic significance by computing the effect of a one hour delay on the market value of the

²⁵ The Stockholm OMX30 consists of the 30 most traded stocks.

companies traded. The dummy for the worst case (the 20% of days with worst outcome) do not have any additional effect except the one pure delay time component. The results are the same when we use only the 10% worst days.

4.4 Implications for VTTV

Based on the results in Section 4.3, we compute the change in company value, which is equal to the unforeseen delay costs. Table 6 shows the monetary values.

As can be seen, each hour of additional delay, causes company value to decrease by between 2,000-28,000 SEK (just taken into consideration the negative values). The positive values show the importance of just considering the significant effects. They are SKF, SCA, HOLMEN, STORA ENSO and AXFOOD. It is however difficult to interpret these values since we do not know the amount in tonne that are transported by each company on a daily basis. The mean effect of the five significant stocks is 15,400 SEK per hour delay. For Sweden as a whole, every hour roughly 2850 tonne are transported on rail. If we assume that the above companies are in some way or another exposed to 50% of all tonnes transported, the results would imply 2.7 SEK per tonne and delay hour.

Naturally this effect might be biased downwards because of several reasons, for example measurement error, mainly because of geographical dispersion between delay and production/retail. However, the method proposed here is well suitable to examine the question if there are nonlinearities in the valuations of delay. If existent, it would imply that the values for small delays and the very large delays would differ on a per hour basis. The results here indicate that this is not the case, and hence, in societal evaluations of delays we do not see any necessity to add cost of vulnerability to the pure time cost of reliability.

Table 6: Monetary effect per delay hour

STOCKS	ΔVALUE
VOLVO B	-28072
TRELLEBORG	-11280
SSAB B	-3244
SKF B	-19957
SCANIA B	-10480
SCA B	-28267
SANDVIK	68740
ICA	-2465
HUSQVARNA B	12704
HOLMEN B	-5872
STORA ENSO R	-20502
BOLIDEN	8069
BILLERUD	-12660
AXFOOD	-2450
AUTOLIV	-1773
ATLAS COPCO B	-9794
ASSA ABLOY	-11852
ALFA LAVAL	52851
ABB	-10127

5 Conclusions

This paper suggests and demonstrates three different approaches to calculate estimates of VTTV for Swedish freight transports. The VTTV are not directly comparable, since they are based on different definitions of VTTV.

The *precautionary cost approach* uses the hypothesis that the company reacts to a stochastic delivery time by holding a safety stock, which is a function of standard deviation in transport times: $s=f(\sigma)$ ceteris paribus. The cost for holding a safety stock and hence the cost of variability in transport times is the cost of physical storage of the goods and the capital costs of the goods stored. Hence we can compute the societal cost of variability under certain simplifying conditions. To the best of our knowledge the safety stock approach, although proposed at several occasions, has not been implemented to calculate the VTTV. We show i) that the marginal precautionary costs measure marginal VTTV and that ii) a precautionary stock approach can in principle be made operational by aggregating all companies with freight transport exposure, computing a virtual safety stock and using key aggregate figures about transport time variance, inventory costs, freight flows and required service level. Required service levels are obtainable from companies since they are key figures used in practice.

The *case study approach* identifies the amount and type of the additional costs that shippers face due to the variability of the rail transport time with help of company cases. As a first case, we study the shuttle train run by the grocery company COOP. We measure the degree of variability that the company faces, and identify and estimate the precautionary costs COOP were willing to accept to manage costs of transport time variability. We show i) that by doing a case study it is possible to get an estimate of VTTV valid for a specific company and that ii) in conclusion, assuming an high degree of transport market concentration with regard to shippers and carriers, just a few case studies for key companies in the market might be sufficient to get a surprisingly representative VTTV-measure.

The *market based approach* is based on the hypothesis that publicly traded company stocks accurately reflect the steady stream of information and hence that unexpected delays should have an effect on stock prices. We show i) that stock prices for companies with rail transport exposure in fact react on delays, ii) that changes in company market value per hour delay can be used as a VTTV-estimate given figures about freight flows for a certain company and that iii) the method can be used to discern between costs for relatively small delays and the very large delays. In other words, the method has the potential to identify costs of transport system vulnerability not covered by VTTV.

We demonstrate the use of each approach using crude figures as an input. We advocate therefore further research on getting more realistic inputs. Moreover, the theoretical and empirical methods should be developed. Also, more research should be done on how to incorporate the extremeness of empirical delays in models and definitions of VTTV. Last but not least, new methods based on micro-level data on company stocks and the standard SP-method should be used in combination with the approaches suggested here, in order to validate VTTV-estimates.

References

- Bruzelius, N. (1996). *Företagens MA-kostnader och företagens kapitalkostnader för fordon, Två uppsatser om samhällsekonomiska effekter av vägåtgärder*. Borlänge: Statens Vägverk.
- Hansmann, K.-W. (2006), *Industrielles Management*, 8th ed, Oldenbourg.
- Krüger, N.A.; Vierth, I; Fakhraei Roudsari, F. (2013). *Spatial, Temporal and Size Distribution of Freight Train Delays: Evidence from Sweden*, unpublished working paper.
- Minken, H. (1997). *Næringslivets nytte av raskere og mer pålitlig godstransport*, TØI-rapport 347/1997, Oslo.
- Haug, E.G. and Taleb, N.N. (2011): *Option traders use (very) sophisticated heuristics, never the Black-Scholes-Merton formula*, 77(2-3), 97-111, Journal of Economic Behavior and Organization.
- Trafikanalys. (2010). *Varuflödesundersökningen 2009*. Östersund: Trafikanalys (Statistik 2010:16).
- Vierth, I., Lord, N., & Mc Daniel, J. (2009). Representation of the Swedish transport and logistics system (Logistics model version 2.0). VTI Notat 17A-009.
- Vierth, I. (2012). *Värdering av tidsvinster och högre tillförlitlighet för godstransporter*. Stockholm: VTI (VTI notat 24-2012).