

Can repeated surveys reveal the variation of the value of travel time over time?

VTI Working Paper 2021:7

Authors

Maria Börjesson^{1,2}, Marco Kouwenhoven³, Gerard de Jong^{3,4}, Andrew Daly⁴

¹Transport Economics, VTI, Swedish National Road and Transport Research Institute

²Linköping University

³Significance

⁴ITS Leeds

Abstract

This paper studies intertemporal changes in the value of travel time (VTT) and investigates whether the change of VTT over time can be studied based on national VTT data, collected at two points in time. We use repeated national VTT data from the Netherlands and Sweden, collected 13 and 14 years apart. The results show mostly a declining VTT for a given income level. The results show also a large within-country heterogeneity across modes and purposes, in the cross-sectional income elasticity of the VTT, and in its development over time. The explanation most consistent with our results and those of others is that the VTT has in fact increased due to income increases, but that the repeated stated choice data cannot detect this given the data, methodology and population changes. In particular, it seems that the response rate has dropped considerably in the later surveys partly due to a higher share of (busy) respondents declining to be recruited. The main contribution of this paper is to document the differences between the studies carried out in different years, indicating the reasons why it is difficult to identify temporal changes in the VTT.

Keywords

Value of Time; Stated Preference; Income elasticity; Cost-benefit Analysis; Repeated Studies; Data Collection.

JEL Codes

D61, H54, R41, R43, R48



1 INTRODUCTION

A forecast of the value of travel time (VTT) is a key parameter in the appraisal of infrastructure investment, since time saving is the main benefit of many transport investments. Some countries take the intertemporal income elasticity to be the elasticity estimated on cross-sectional data. Sweden for instance applies the cross-sectional income elasticity to the cost parameter in the national forecast model. Other countries, such as the UK and The Netherlands, use data collected at different points in time to make assumptions on the intertemporal income elasticity, either through meta-studies (Wardman et al., 2016) or repeated SP studies (Gunn, 2001; Gunn et al., 1999); also Sweden applies this method for valuing transport benefits.

However, there are few such studies based on data collected at different points in time to study the development of the VTT over time, and they give puzzling results (see below). This paper contributes to this literature by analysing data from national VTT surveys collected at two or three points in time, in the Netherlands (1997 vs 2009/2011) and Sweden (1994 vs 2007 vs 2008). The aim of this paper is twofold. First, we analyse how the VTT has changed over time by estimating the VTT on the two data sets for each year by applying the same econometric model inspired by Hess et al. (2017) and De Borger and Fosgerau (2008). Second, we analyse the comparability of the surveys. Even if the data from the different years were state-of-practice national stated choice surveys when they were conducted, they are subject to differences regarding sampling and recruitment methods, response rates, experimental design, and survey design. We therefore aim at documenting these differences and exploring to what extent they impact the comparability of the different studies.

The VTT depends on three parameters (DeSerpa, 1971) that all may vary over time for different reasons

$$VTT = \frac{\mu}{\lambda} - \frac{u_t - u_w}{\lambda},$$

where μ is the opportunity utility of time, λ is the marginal utility of money and $u_t - u_w$ is the direct utility of travel time, i.e. the differences between the utility of time during travel (u_t) and the utility of time spent working (u_w).¹ At the individual level, λ should reduce with income. However, the relationship between λ and income found in the cross-section might not apply over time at the population level, for instance if the income distribution or population composition changes. The opportunity utility of time reflects the value of time as a resource and equals the utility that could be attained if the travel time could be used for some other activity. This utility may therefore change if people become more (or less) productive or busy at the origin or destination location. The direct utility of travel time (compared to time spent working or on some other reference

¹ This formula is valid if working hours can be chosen freely. If not, it reduces to $VTT = \frac{\mu}{\lambda} - \frac{u_t}{\lambda}$.

activity) depends on the comfort of the travel time, but also on how productive or enjoyable the travel time is. It may change over time due to new technology (e.g. laptops and smartphones), crowding and congestion etc..²

The average VTT in the population can also change due to changes in the composition of the population (e.g. in terms of employment, or age). The gross income might also have different impact on the VTT if the tax systems change. This paper does not, however, attempt to answer how the average VTT has changed over time in the national populations, but rather how the VTT has changed for similar individuals in the two samples.

The implications of moving away from the assumption that the VTT grows over time proportionally with wage rate or income because the marginal utility of travel time declines, would reduce the future year time benefits in the appraisal of transport projects. On the other hand, if the time parameter in forecast models decreases, more or longer trips would be forecast, increasing the benefits in the appraisal.

We know of four previous studies that have used repeated stated choice data collected roughly 10 years or more apart to study the temporal change in the VTT. The first was based on data conducted in the Netherlands in 1988 and 1997 (Gunn, 2001; Gunn et al., 1999). The second study used data collected in Britain 1985 and 1994 (Wardman, 2001). The third study used data collected in Britain 1994 and 2006 (Tapley et al., 2007), and the fourth used data collected for drivers in Sweden 1994 and 2007 (Börjesson et al., 2012a). The first study found that the VTT had decreased within each income group, but the income increase was large enough to cancel out the trend decline at each real income level, such that the real average VTT remained largely unchanged. The second and third studies found a slight trend decline in the VTT, in spite of an income increase. Gunn (2001) suggests that the non-increasing VTT could be explained by a declining direct disutility of travel time.

The fourth repeated study by Börjesson et al. (2012a), using log willingness to pay with random variation, found no significant difference between the cross-sectional and intertemporal income elasticity. Using the same data, Börjesson (2014) found that the travel time coefficient had remained constant over time but that the travel cost parameter had declined in real terms – which is a central result for implementation in forecasting models. Using RP data Fox (2015) came to the same conclusion (see section 2). Hence, of all the repeated SP surveys, only the Swedish repeated car study consistently indicated increased VTT over time. This is also the only study where exactly the same questionnaire, experimental design, and survey method were used in the second wave; the sampling of drivers was also identical, including the season, time of day and locations; a reanalysis of this data for the current paper confirms these earlier results. This indicates that differences in recruitment method, interview method, response rate, season, questionnaire and experimental design can become important reasons for

² There was an ITF-OECD Roundtable in 2018 on the theme of 'Zero value of time' (ITF-OECD, 2019), where most experts agreed the VTT will probably not increase in line with income, may very well go down in time because of new ICT and automated vehicles, but will remain positive.

differences in the estimated VTT. As far as possible, we control for such differences in this paper, attempting to do this more thoroughly than in the previous repeat studies. Still our results suggest that the differences that we cannot control for have such a large impact that the resulting VTT cannot be meaningfully compared. In particular it seems likely that the trend decline in response rates, partly due to changes in recruitment method, leads to a substantial downward bias in the VTT in the later year.

A fifth repeat study is Weis et al. (2021), covering only five years (2010 and 2015), where both waves were designed as a repeated follow up SP experiment to the Mobility and Transport Microcensus (MTMC). The authors state that the survey method, recruitment strategy and experimental design remained similar, even if some attributes were slightly revised. They also found that the VTT was stable between the samples. Another study, Flügel et al. (2020), compares the Norwegian VTT from 1997, 2009 and 2018 published in the official reports (i.e. without specific modelling), and finds that they had mostly increased from 1997 to 2009 (except for air trips) and from 2009 to 2018. However, for car, the VTT had declined from 2009 to 2018, in spite of income increases.

Section 2 reviews evidence from cross-sectional models, meta studies and repeated RP studies. Section 3 briefly describes the data used and Section 4 discusses the model specification. Section 5 then summarises the results and discusses the findings; Section 6 concludes. Appendices describe the data more fully and present the detailed model results.

2 PREVIOUS STUDIES

This section reviews earlier studies on the intertemporal and cross-sectional income elasticity in addition to the repeated SP studies reviewed above. It shows that here is a considerable disagreement about VTT elasticity and how to forecast VTT when income changes, which is the main justification for more research such as the present study.

Stated preference cross-sectional data yield income elasticities in a wide span of 0.07 – 0.9, with a consensus value of 0.3-0.5 (see Table 1). Daly and Fox (2012) find a cross-sectional income elasticity in the middle of this consensus span but point out that income elasticities estimated on cross-sectional models cannot necessarily be applied when forecasting the VTT over time because they could be biased by unobserved variables influencing the VTT and correlating with income in the cross-section. To avoid such bias, the data should include longitudinal variation.

Using RP cross-sectional data collected at different points in time in Toronto, Fox (2015) finds that the time parameter remained constant and that the cost parameter declined with income. Swärdh (2008) estimated the income elasticity on the VTT for commuting trips based on revealed choices on income and commuting time for workers changing jobs. Swärdh used Swedish register data from 1983, 1990 and 1993 and found that the intertemporal income elasticity did not significantly differ from unity.

Assuming that the marginal utility of money is inversely proportional to income, implying that the VTT would increase with income, would all else equal also imply reducing price elasticities for travel. That is, if the importance of cost declined over time because income increases, this would not only increase the VTT but it would also reduce the price elasticities for travel. However, the empirical evidence is mixed; Neither Hanley et al. (2002) nor Wardman (2014) find any decline in price elasticity over time. But Bastian et al. (2016) show that income elasticities of demand for driving have declined over time while fuel price elasticities increased over time, in several western countries.

Meta studies are regressions on the outcomes of many cross-sectional VTT studies, which are explained from attributes of the study area, survey period and the method used. Meta-analyses comparing studies based on data collected in different countries at one point in time would estimate a cross-sectional elasticity. Meta-analysis comparing studies based on data from one single country collected at different points in time would estimate an intertemporal elasticity. However, other meta-analyses use studies based on data from many countries and from different points in time, and it is therefore difficult to quantify the difference between intertemporal and cross-sectional elasticity. Also, meta studies should be interpreted with caution – precisely because the quality of the underlying studies is often unclear.

Table 1. Cross-sectional income elasticity of the VTT from SP studies. Note that income elasticities obtained from using banded income are usually subject to more uncertainty than those directly estimated as part of the model specification.

Reference	Study area and data	Year of data collection	Income elasticity of VTT
Gunn et al. (1999) ³	Netherlands	1988	0.28-0.48 depending on purpose (non-business) and mode. From income segmentation of the VTT
Gunn et al. (1999) ³	Netherlands	1997	0.33-0.65 depending on purpose (non-business) and mode. From income segmentation of the VTT
Daly and Fox (2012)	Review of 12 SP studies, including most studies mentioned in this table	1987-2009	Average of studies: 0.24 (0.40 for personal income)

³ This reference does not explicitly state income elasticities of the VTT, but we have calculated these from the VTT estimated by income segment using detailed historic income distribution information from Statistics Netherlands (CBS). Note that these VTTs are calculated after weighting / expansion, so that these also include other effects that influence the VTT and that are correlated with income. Hence the resulting elasticities are not pure income elasticities with all other effects being equal.

Dillén and Algers (1998)	Sweden	1994	All purposes and modes ⁴ 0.46 From income segmentation of the VTT. Personal income.
Börjesson and Eliasson (2014)	Sweden	2007/08	All purposes: Car: 0.5; PT: 0.17; Long distance PT: 0.20. (personal income). Income elasticity estimated in the model.
Wardman (2001)	UK	1985	Less than unity. From income segmentation of the VTT.
Gunn et al. (1996)	UK	1994	0.65 (commuting), 0.45 (business) and 0.35 (other) Income elasticity “derived from the models”.
Batley et al. (2017)	UK	2014	0.3–0.4 for rail; 0.3-0.7 for car. Income elasticity estimated in the model. Mix of personal and household incomes.
Ramjerdi (2010)	Norway	2009	Car short distance: 0.432 PT: 0.617 Long distance bus 0.318 Train: 0.454 Car long distance: 0.5698 Air: 0.248 (personal income) Income elasticity estimated in the model.
Ramjerdi et al. (1997)	Norway	1995	Less than unity. From income segmentation of the VTT. Personal income.
Flügel et al. (2020)	Norway	2018	Car driver: 0.216 (short distance), 0.474 (long) Train: 0.475 (short) All elasticities are for personal income. Income elasticity estimated in the model.
Fosgerau (2005)	Denmark	2004	Car all purposes: 0.90 (personal income) Income elasticity estimated in the model.

**Elasticities refers to income at the household level, unless otherwise reported.*

Wardman published several meta-studies on UK VTT data. As the period and the number of studies were extended, the income elasticity increased: from 0.075 ± 0.029 in Wardman (1998) to 0.510 ± 0.300 in (2001), 0.723 ± 0.043 in (2004) and to 0.900 ± 0.035 in Abrantes and Wardman (2011), the final study applying material from 1960-2008. The elasticity in the latest study is probably most accurate, as it was based on the largest data set, but also to a larger extent

⁴ Dillén and Algers write: “The lowest income category is disregarded, and if the rest of the categories are grouped in two halves, then the household income elasticity of the VoT is 0.46 for single person households and ranges from 0.07 to 0.24 for 2 person households with and without children respectively”.

reflecting an intertemporal elasticity due to a wider time span. If the latter factor played a role, it indicates that the intertemporal income elasticity is higher than the cross-sectional.

Shires and de Jong (2009) used VTT studies from 30 (mainly OECD) countries almost exclusively for the period 1990-2003 for their meta-analysis. They found the income elasticity to be 0.47 for business, 0.67 for commuting and 0.52 for other purposes. Since they use studies from many countries and from several years, they estimate a mixture of cross-sectional and intertemporal elasticities. However, it is possible that the cross-sectional variation is dominant in the estimation results, since variation in the VTT between countries is much larger than the variation over time within countries. In a study for the European Investment Bank (Wardman et al., 2012), the data on the UK were extended again, but also combined with data for many other European countries over a long period. The income elasticity of the VTT then increased to 0.72, possibly due to a larger data set with larger variation.⁵

3 DATA

In this paper, we analyse data from five national VTT stated choice studies, two in the Netherlands and three in Sweden. The Dutch studies denoted NL97 and NL11 were collected in 1997 and 2011, respectively. The Swedish studies, denoted SE94, SE07 and SE08 were collected in 1994, 2007 and 2008, respectively. The SE07 study was conducted in Sweden for the specific purpose of studying the how the VTT had changed over time. Therefore, an exact replication of the SE94 study was carried out, for car drivers only. In this replication care was taken to use exactly the same questionnaire, design (but the cost differences were inflated by 40 percent corresponding to inflation and GDP growth) and sampling of drivers, including the season, time of day and locations, as in the SE94 study (even the mistakes were repeated on purpose). The same market research firm was used, receiving the same instructions, for the recruitment and interviews and the same computer software. The VTT over time using the SE94 and the SE07 study was studied in Börjesson et al., (2012a) and Börjesson (2014). But in the present study we analyse them together with the SE08 car data.

A third Dutch VTT study was conducted in 2009, NL09, where respondents were recruited from a commercial internet panel. However, the analysis of this data showed an implausibly low VTT. This was attributed to the composition of this internet panel: mostly people that have a lot of time to fill in these internet surveys and even sometimes supplement their incomes with the rewards from completing a large number of surveys. Therefore, it was decided to recruit another set of respondents in 2011 in (almost) the same way as was done in 1997. In the final analysis, and in the present paper, the absolute level of the VTT was based on the 2011 data set only, but the socio-economic interaction coefficients

⁵ More recent analysis on this data set, using GDP in purchasing power parity (PPP) instead of market exchange rates, gave an income elasticity of around 1 (Wardman, et al, 2016).

were based on the joint 2009/2011-datasets. This is not likely to have any major impact on the resulting VTT relevant for the present study.

Table 2 summarises the main characteristics of the four surveys. Further description of the surveys is provided in Appendix A. We conjecture that there are five major differences between the data sets, that cannot be corrected for in retrospect. These are described in the lower part of the table. The first is the number of attributes, implying that the model specifications cannot be identical for the Dutch and the Swedish data. Four surveys included stated choice experiments with only a time and a cost attribute, but the experiment for public transport (PT) travellers in the Swedish SE94 study also included attributes for frequency and the number of transfers, giving four attributes in total.

Table 2. Main characteristics of the VTT surveys included in this paper.

	NL97	NL11	SE94	SE07	SE08
Area	Netherlands	Netherlands	Sweden	Sweden	Sweden
Period of data collection	Mar.-Apr. 1997	Nov. 2009 & Mar.-May 2011	Oct.-Nov. 1994	Oct.-Nov. 2007	Oct.-Nov. 2008
Modes	car, public transport	car, public transport, air, recreational boating	car, bus, train, long distance bus, long distance train	car	car, bus, train, long distance bus, long distance train
Purposes	commute, business, other	commute, business, other	commute, other	commute, other	commute, other
Income question	Gross monthly household income	Net annual household income	Gross monthly personal income ⁶	Gross monthly personal income ⁶	Gross monthly personal income
Survey method	Paper + mail-back	Web-based	Paper + call-back	Paper + call-back	Web-based (or call-back, if preferred)
<i>Major differences</i>					
1. Number of attributes in SP experiment	2: time, cost	2: time, cost	PT 4: time, cost, frequency, transfers Car 2: time, cost	2: time, cost	2: time, cost
2. Presentation format	time and cost difference with respect to reference	absolute times and costs, pivoted around reference levels	absolute times and costs, pivoted around reference levels	absolute times and costs, pivoted around reference levels	absolute times and costs, pivoted around reference levels

⁶ Household income was also stated but explained less of VTT. Joint taxation does not apply in Sweden.

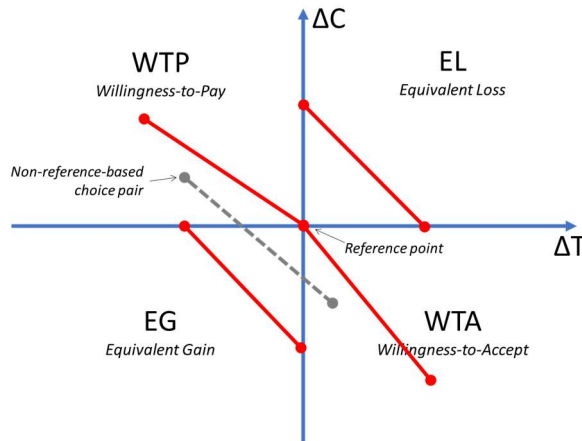
3. Choice types	5 reference-based + 6 non-reference based, Same time/cost difference ratios for all respondents	All reference-based-type questions WTP/WTa/EL/EL More variation in time/cost difference ratios	All reference-based-type questions WTP/WTa	All reference-based-type questions WTP/WTa	All reference-based-type questions WTP/WTa/EL/EL Wide variation in time/cost difference ratios to cover the tail of the VTT distribution
4. Time and cost levels in the experiment and the reference trips	See Table 7	See Table 7	See Table 8	See Table 8	See Table 8
5a. Recruitment method and location	en-route at railway stations, bus stops, petrol stations, parking garages	2009: panel 2011: en-route at railway stations, bus stops, petrol stations, parking garages	en-route, bus, train: on-board the vehicle, while travelling car: number plate registration	en-route, number plate registration	bus, train: en-route at railway stations and bus stops car: random samples from the national population register
5b. Recruitment response rate	81%	Lower than in 1997	PT: > 95% car: 100%	car: 100%	PT: 53-76% car: 100%
5c. Survey response rate, <i>when recruited</i>	59%	48%	PT: 91-95% car: 65%	car: 55%	PT: 69-74% car: 36%
5d. Reward	Prize draw	2009: € 1,50 2011: € 500 prize draw	None	None	Lottery ticket

The second major difference is the presentation format of the attributes. While two surveys presented the time and cost levels in an absolute way, the NL97, SE94 and SE07 studies presented the difference of the travel time and cost compared to the time and cost of the reference trip, e.g. “Travel time: 10 minutes shorter than now” or “Travel cost: same as now”. The NL97 questionnaire did ask the respondents to state the travel time of their reference trip, but not the travel cost. Therefore, the absolute cost levels in the choice experiment cannot be reconstructed.

The third major difference concerns choice types with respect to a reference trip. In a reference-based design, each choice matches one of the following four types (see Figure 1): Willingness-To-Pay (WTP): a choice between the reference trip and a faster but more expensive trip; Willingness-To-Accept (WTA): a choice between the reference trip and a slower but cheaper trip; Equivalent Gain (EG): a choice between a trip that is faster than the reference trip and has the reference trip cost and a trip that is cheaper than the reference trip and has the reference travel time; or Equivalent Loss (EL): a choice between a trip that is slower than the reference trip and has the reference trip cost and a trip that is more expensive

than the reference trip and has the reference travel time. The NL97 survey was partly reference-based and partly non-reference-based, the other four surveys were fully reference-based.⁷ However, the SE94 and the SE07 study included only WTP and WTA type choices.

Figure 1 Four quadrants of reference-based choice pairs and an example of a non-reference-based choice pair.



A fourth difference is that the time and cost of the reference trip differ between the years, but also the average time and cost differences between the alternatives differ between the years. Table 7 in Appendix A shows that, for the Dutch car data, the average travel time is lower in the later years, due to sampling differences (fewer drivers were recruited at fuel stations with longer travel distances in the later year). Table 8 shows that the travel time and cost differences are larger in the later Swedish study, because the design covered a larger range of trade-off values of time to uncover the tail of the VTT distribution.

A fifth major difference that cannot be corrected for relates to the response rate and recruitment method. The rates of respondents willing to participate in the surveys were much higher in the nineteen-nineties than in the more recent surveys. In addition, in the Swedish PT surveys, the recruitment method shifted from on board the bus/train, where respondents had little chance of avoiding being approached, to recruitment on the platforms. Recruitment on platforms makes it easier for busy travellers, with high VTT (possibly in a hurry to catch their train or bus), to escape recruitment or to decline. In the Swedish car study, number plate registration was used in SE94 and SE07, but the SE08 study used a random population sample of individuals having made a car trip during a pre-defined survey. The recruitment method in the NL97 and NL11 surveys was identical, but fewer respondents declined recruitment in the earlier year. Hence, it seems likely that the drop out of busier travellers with high VTT was larger in the later survey. And after being recruited, the rates of respondents that completed the survey also dropped in more recent years. This pattern of reduced

⁷ Stated choices from non-reference-based trips are usually much more difficult to analyse, since not only the difference in travel time between the two alternatives enters the utility equations, but also the difference between the travel time of each alternative and the reference travel time (and the same for cost), as was done in the recent UK-VTT study (Batley et al. 2017). However, data of this type can give more insight into preferences, as it contains more variation, while any preference effect favouring the current reference trip is eliminated.

response rates is observed in both the Netherlands and Sweden. Moreover, only the SE08 study gave a lottery ticket to all recruited to the study.

The major differences one to three relate to differences in the stated choice experiment, while the fifth arises in the recruitment process. The fourth difference is partly a result of differences in the design, but arises also partly in the recruitment process. None of these differences can be entirely controlled for in the estimation. However, differences in experimental design, in terms of size and sign effects, are controlled for as much as possible by including these variables in the model estimation. Moreover, the absolute levels of the time and cost of the reference trip are also controlled for by including them in the models (for the NL data sets, only time could be included since no information on the reference cost could be included).

Regarding the issues with self-selection, differences in the socio-economic sample composition can be controlled for by including such variables in the estimation if they have a significant impact. Hence, differences in the VTT that we find should not depend on the socio-economic composition of the samples. In this way we can control for differences in self-selection effects due to lower response rates (in the recruitment and in the survey) to the extent that they can be captured by observed sample characteristics. Note, however, that we do not attempt to answer how the average VTT has changed over time in the national populations, so we do not reweight the samples to make them representative of the population at the two points in time.⁸ We only address how the VTT has changed for individuals in the two samples, when controlling for the socio-economic characteristics that impact the VTT. Moreover, since we do not know how the recruitment method impacts the sampling probability, we refrain from weighting observations in the estimation.

However, we cannot control for self-selection effects that arise from unobserved effects. Such effects can be large, not least because the VTT can vary also within the same person for different trips. In fact, we cannot control for changes in any factor impacting the VTT that was not included in the experiments. For instance, if reliability, congestion levels, crowding levels, comfort, average travel distance/travel time per day etc. impact the VTT and have changed over time we cannot control for this. Indeed, as stated in the introduction such factors might be one reason why the VTT would increase faster or slower than the income.

All incomes in the NL97 survey were converted to net annual incomes.⁹ All incomes and costs were inflated in line with the growth of the Dutch consumer

⁸ This was investigated in the Dutch 2009 survey (internet panel) generating very low VTTs. However, it was found that correcting for the socio-demographic and economic changes in the population did not help much: the low VTT occurred across all groups recruited by an internet panel. Note also that standard choice modelling theory (e.g. Lerman and Manki (1977) quoted by Ben-Akiva/Lerman (1985)) says that weighting in estimation to deal with exogenous sampling is not required.

⁹ Statistics Netherlands (CBS) provides average gross and net household incomes for each income decile. The resulting gross to net income ratios were interpreted as the ratios for the midpoint of each decile. From these, the gross to net income ratios for the boundaries of the income bands in

price index 1997 – 2011. Similarly, all incomes and cost from the 2009 data collection phases of the NL11 survey were inflated in line with the growth of the Dutch consumer price index 2009 – 2011 so that all values for the Netherlands in the rest of this paper are given in 2011 prices. In Sweden, gross monthly incomes are used. However, the tax system did not change to any large extent between these years. All incomes and costs in the SE94 survey were inflated in line with the growth of the Swedish consumer price index 1994 – 2008 so that all values for Sweden in the rest of this paper are given in 2008 prices.

4 METHODOLOGY

4.1 Choice of model

The central problem that arises in attempting to compare VTT findings over time and/or between countries is dealing with the differences in the data designs, discussed in the previous section. Ideally, we would wish to have the same model formulation for both countries and for all years. However, because we cannot reconstruct the absolute cost levels in the choice experiment in the NL97 data, we cannot use the model applied for the Swedish data to the Dutch data. And, because the SE94 PT data includes many more attributes, we cannot use the model applied for the Dutch data to the Swedish data. Most important is to maintain consistency within each country, so that intertemporal VTT comparisons are less affected by data changes. We are also restricted to formulations that give reasonable estimates of VTT.

The log value specification, which has been found to give the best results in a number of studies (De Borger and Fosgerau, 2008; Börjesson and Eliasson, 2014; Hess et al., 2017) cannot be used for the SE94 PT data because it has more than two attributes. The log value approach works by comparing a postulated random VTT for each respondent with a limiting value calculated from the data. With more than two attributes, a single limiting value cannot be calculated, as the critical value for the respondent depends on multiple features of the data. For the Dutch data, we were able to use the log value approach, while for the Swedish PT data we used a ‘multiplicative’ formulation that has some features of the log value approach and was used, for example, in the multi-attribute experiments of the recent UK study (Hess et al, 2017). For the Swedish car data, we used the log value specification with a mixing distribution to make the results comparable with Börjesson et al. (2012a), who used this model for SE94 and SE07. Utility functions with additive random errors were tried but yielded a considerably worse fit (such models for Swedish long-distance trips are shown in Table 14 and Table 15 in Appendix B).

We do not aim at the most sophisticated model. We did not use mixed logit models (except for the Swedish car data). However, limited tests were made with a model estimating a log-normally distributed VTT for the Swedish long-distance

the survey were calculated. For each income band in the survey the average net income was calculated by using detailed information on the national distribution of incomes within each survey band which is also provided by Statistics Netherlands (CBS).

data (Table 14 and Table 15 in Appendix B). The results did not change much, though the VTT increased somewhat, which is not surprising given that the lognormal distribution has a flat tail. The main reason why we do not apply the mixed models in our main models is that when estimating the VTT distribution, in our case a log-normal distribution, also outside the range supported by the data, the results may depend on the parts of the distribution that are extrapolated outside the data range. The impact on the estimated mean that this has is in general larger, the larger the part of the tail that is not supported by the data. This means that a comparison of the VTT resulting from mixed models estimated on data from different years could be seriously biased if the range not supported by the data is large in any of the samples, in particular if the VTT distribution has different support of the data in the two samples. This might be the case since the experimental designs differ.

The specification is applied separately to the two years of sample data in each country, so that only the cross-sectional income elasticity is estimated. However, for the Swedish car data, the SE94, SE07 and SE08 were pooled in one model to make the results comparable with Börjesson et al. (2012a). Moreover, the recruitment method changed from en-route sampling in the SE94 and SE07 to an exogenous random sample in SE08. The selection probability increases with trip length in the earlier two but not in the last survey. Since trip length is positively associated with VTT, the observations must be weighted in the application to be comparable. In estimation however, we account for the difference in trip length by including travel distance in the pooled data model. For this reason, we only compare the difference in the VTT by yearly dummy variables in the estimation in the Swedish car data. Joint estimation data from both survey years was also applied for Swedish PT data but did not provide any further insights.

4.2 Netherlands

The common element in the SP experiments in the NL97 and NL11 data is a binary choice experiment with time and cost as the only attributes. In this paper, we use data from these experiments only. As in many econometric studies, the fit of all the models estimated in this paper increases with a multiplicative error formulation relative to an additive error formulation. For the time-cost experiments in the Dutch studies the utility specification is based on the logarithmic specification, i.e. the multiplicative error structure, used by de Borger and Fosgerau (2008)

$$\begin{aligned} V_1 &= \mu \cdot \log \left(-\frac{\Delta c}{\Delta t} \right) + \varepsilon \\ V_2 &= \mu \cdot \log (W), \end{aligned} \tag{1}$$

where W is the value of time and ε is a standard logistic error, so that a logit model results; μ is a scale parameter. V_1 is the quicker and more expensive alternative and V_2 is the slower and cheaper alternative: this definition leads to a negative expected value for μ . Δc and Δt are the absolute differences between the travel costs and travel times, respectively, in the binary choices in the experiment.

This specification was already used by Börjesson et al. (2012a) and more recently by Hess et al. (2017) and performed well in those studies.

It is well known that the VTT depends on the design variables, amongst other reasons through size and sign effects, and that reference travel time and travel cost also have considerable impact on the estimated VTT. Hence, the models must take these design and trip characteristics into account for a fair comparison across surveys.¹⁰

In the Swedish model, dummy variables were included to control for type choices (see Eq. (8) in section 8.2.2). However, in the NL97 survey almost half of the choices were not of the type WTP, WTA, EG or EL. As a result, dummy variables cannot be used to find a reference-free VTT. Instead, the size and sign effects of the cost and time differences are modelled explicitly. This is done in a similar way as in the recent UK value of travel time study (Batley et al., 2017; Hess et al., 2017). Hence Eq. (1) is rewritten as

$$\begin{aligned} V_1 &= \mu \cdot \log \left(-\frac{v(dc_1) - v(dc_2)}{\theta \cdot v(dt_1) - \theta \cdot v(dt_2)} \right) + \varepsilon \\ V_2 &= 0, \end{aligned} \quad (2)$$

where v is the “value” function of the change ($c_1 - c_0$), in which c_1 is the cost of the leftmost alternative, c_0 is the cost of the reference alternative. The value of the cost difference relative to the reference alternative is

$$v(dc_1) = \text{sgn}(c_1 - c_0) \cdot \exp(\eta_c \cdot \text{sgn}(c_1 - c_0)) \cdot |c_1 - c_0|^{1-\beta_c}, \quad (3)$$

where sgn is the sign function. In Eq. (2), θ denotes the reference-free VTT. In Eq. (3), η_c and β_c allow for the reference-dependent value to be different for gains and losses and non-linearly impacted by the size of the gains or losses relative to the reference, in the cost dimension; analogous parameters adjust the time value function. Hess et al. (2017) also allowed for non-symmetric marginal valuations of gains and losses in the value function. Due to data limitations in the present study, however, Eq. (3) assumes that gains and losses have the same functional form.

The parameter θ is defined as a function of several socio-economic and trip variables

$$\begin{aligned} \theta &= (\theta_{Car} \cdot \delta_{Car} + \theta_{Train} \cdot \delta_{Train} + \theta_{LocalPT} \cdot \delta_{LocalPT}) \cdot \\ &\quad (1 + fac_{Fem} \cdot \delta_{Fem}) \cdot (1 + fac_{HH1} \cdot \delta_{HH1}) \cdot \\ &\quad (1 + fac_{Age3650} \cdot \delta_{Age3650} + fac_{Age51+} \cdot \delta_{Age51+}) \cdot \\ &\quad \left(1 + fac_{Inc} \cdot \left(\frac{Income - 35000}{10000} \right) \right) \cdot \\ &\quad \left(\frac{TravTime}{60} \right)^\zeta. \end{aligned} \quad (4)$$

¹⁰ This increases the complexity of the model, and it can be questioned how much of this can be justified by increases in model fit. However, in our case, we are not aiming for a good prediction model but rather to be able to compare the VTT in two different data sets. It is then necessary to control as much as possible for all design and trip related variables that we know impact the VTT. This holds even if this has a limited impact on the prediction power of behaviour.

In the functions (2-4), the variables and estimated parameters are defined as follows.

μ	estimated scale
$TravTime$	travel time of the reference trip
ζ	travel time elasticity: separate values are estimated for each mode or groups of modes
$Income$	annual household income after taxes at price level 2011
fac_{Inc}	income factor: separate values are estimated for each mode
$\theta_{Car}, \theta_{Train}, \theta_{LocalPT}$	estimated reference VTT for the mode given by the 0/1 δ indicator
β_t, β_c	estimated size effects for time and cost
η_t, η_c	estimated sign effects for time and cost
fac_{xx}	estimated parameters estimated for 0/1 variables as follows
δ_{Fem}	indicator for women
δ_{HH1}	indicator 1-person household
$\delta_{Age3650}$	indicator for age between 36 and 50
δ_{Age51+}	indicator for age over 50

In the model estimation, the two parameters in each value function (η and β) and the parameters of the θ function are determined, as well as the scale μ . Hess et al (2017) and De Borger and Fosgerau (2008) then show that, since the θ is outside the value function in Eq. (2), η disappears from the calculation and the VTT W can be computed as

$$W = \theta \cdot |t - t_0|^{\kappa-1}, \quad (5)$$

where

$$\kappa = \frac{1-\beta_t}{1-\beta_c}.$$

We expect the size effect to be larger in the time dimension, i.e. $\beta_t < \beta_c$, so that the VTT reduces with smaller time savings.

Unfortunately, the delta method of calculating the variance of the estimated mean VTTs that was applied for the Swedish estimation could not be used for the Dutch analysis because of the complexity of the utility functions. Therefore, we had to revert to sample enumeration to derive the t-ratios for The Netherlands as described in section 5.

4.3 Sweden

For the SE94 PT data the logarithmic difference specification with multiplicative error structure cannot be used because there are more than two attributes. For this reason, we apply a multiplicative error specification to the Swedish PT data, taking the logs of the observed parts of the utility specification separately

$$\begin{aligned} V_1 &= \mu \cdot \log(c_1 + W \cdot t_1 + \dots) + \varepsilon, \\ V_2 &= \mu \cdot \log(c_2 + W \cdot t_2 + \dots), \end{aligned} \quad (6)$$

where W is the VTT, c_1 and c_2 are the absolute travel costs in the two alternatives of each binary choice and t_1 and t_2 are the absolute travel times in the two alternatives. The specification of Eq. (6) cannot, however, be applied to the Dutch data, because for the 1997 respondents the absolute travel cost levels are unknown.

A central result from De Borger and Fosgerau (2008) is that the underlying reference-free value VTT (W_{rf}) can be obtained as a geometrical mean of the VTT for the choice types WTA and WTP, and the geometrical mean of the VTT for the choice types EL and EG

$$W_{rf} = (W_{WTP} \cdot W_{WTA})^{1/2} = (W_{EG} \cdot W_{EL})^{1/2}. \quad (7)$$

This specification holds assuming that the time and cost differences are symmetric, i.e. that the gains are under-weighted as much as losses are over-weighted. If Eq. (7) does not hold this is evidence for asymmetry of the size effect. Under the assumptions of De Borger and Fosgerau the sign effects can be implemented as multiplicative factors for EL, EG, and WTP choices, assuming the base value of time corresponds to WTA choices. The VTT is therefore parametrised as

$$W = \exp(\beta + \beta_{\Delta t} \log(\Delta t) + \beta_{\Delta c} \log(\Delta c) + \beta_t \log(t) + \beta_{EL} \theta_{EL} + \beta_{EG} \theta_{EG} + \beta_{WTP} \theta_{WTP} + \beta_I \log(I) + \beta_{Imiss} \theta_{Imiss}), \quad (8)$$

where income is I and θ_{Imiss} is a dummy variable indicating missing income (approx. 20% of the sample in each year) and θ_{EL} , θ_{EG} and θ_{WTP} are dummy variables indicating choice types. This formulation ensures that VTT is positive, while the ranges of β are unrestricted. Insignificant variables were removed to obtain the final model presented in tables 11 and 12.

In the model estimation, the parameters of the W function are determined. To compute the variance of the estimated mean VTTs, to be able to compute the t-statistics for the difference in the estimated mean VTT between the years (two unknown means), the delta method giving the Cramér-Rao lower bound (Daly, Hess, and de Jong, 2012) was used on the Swedish data. The variance of the estimated coefficients is then

$$\text{var}(W) = W'^T \Omega^* W', \quad (9)$$

where Ω^* is the robust covariance matrix simulated by the Biogeme software (Bierlaire, 2003). If there are L parameters in the VTT function, W , the vector W' has L elements, namely the derivatives of W with respect to each of the L elements in β . In our final model, the VTT function W includes 7 parameters so that

$$w'^T = \left[\frac{\partial W}{\partial \beta}, \frac{\partial W}{\partial \beta_{\Delta t}}, \frac{\partial W}{\partial \beta_t}, \frac{\partial W}{\partial \beta_c}, \frac{\partial W}{\partial \beta_{WTP}}, \frac{\partial W}{\partial \beta_I}, \frac{\partial W}{\partial \beta_{Imiss}} \right]$$

$$= [W, W \log(\Delta t), W \log(t), W \log(\Delta c), W \log(I), W \theta_{Imiss}, W \theta_{WTP}]$$

The covariance matrix Ω^* has dimension 7 by 7.

In the Swedish car data, the specification of Eq. (1) is used. However, in this model, data from all years (SE94, SE07 and SE08) and purposes were pooled to increase the number of observations. It extends Eq. (1) by allowing the intercept β to follow a normal distribution, assuming it to be constant for each individual (an additional parameter β_σ measures the standard deviation of β), following Börjesson et al. (2012a).. W is parametrised as

$$W = \exp(\beta + \beta_{WTP} + \beta_1 \log \Delta t + \beta_2 \log d + \beta_3 \log t + \beta_I \min(\log I, \log I_{50}) + \beta_{I50} (\min(\log I, \log I_{75}) - \log I_{50}) \cdot 1\{\log I > \log I_{50}\} + \beta_{I75} (\log I - \log I_{75}) \cdot 1\{\log I > \log I_{75}\} + \beta_{employed} \theta_{employed} + \beta_{commute} \theta_{commute} + \beta_{recreation} \theta_{recreation} + \beta_{school} \theta_{school} + \beta_{service} \theta_{service}),$$

The variables and parameters used in these equations are defined as follows.

μ	estimated scale
β	the base value for VTT
β_{xx}	parameters adjusting the VTT relative to the base β for the following variables
Δt	travel time difference from the reference trip
Δc	travel cost difference from the reference trip
$\theta_{EL}, \theta_{EG}, \theta_{WTP}$	VTT differences for choice types: Equivalent Loss, Equivalent Gain and Willingness to Pay
I	Annual net income, price level 2008
θ_{Imiss}	0/1 indicator for missing income
I_{50}, I_{75}	median and 75% quantile of net annual income
t	travel time
d	travel distance
$\theta_{commute}, \theta_{recreation}, \theta_{school}, \theta_{service}$	dummy variables identifying employed people and trip purpose

The controls include dummy variables for year and trip purposes, using 2007 and other trips as base year and trip purpose. Other dummy variables identify employed persons, flexible working hours, gender etc. Insignificant socio-economic terms were removed in the final model presented in table 13.

5 RESULTS

In this section we present and discuss the resulting VTT.

5.1 VTT by survey year

The models estimated for the two years and countries are presented in Tables 9-13 in Appendix B. For each country (except the Swedish car data), the VTT was then derived for all observations, using sample enumeration, in the pooled (combined) sample of respondents from the two years. By pooling the sample,

the differences in the VTT found between the years will not be impacted by changes in design variables or socio-economic composition of the samples, if these effects are perfectly modelled. The sample enumeration was carried out as follows. The VTT was computed for each observation in the pooled sample, based on the estimated model (and the socio-economic statistics and design variables for the observation). Then the average of the resulting VTT distribution was computed and presented in Table 3 and 4, for NL and SE, respectively.

Two versions of the pooled samples were derived for each country, only differing regarding the income variable. Take the NL data as example. In the first version of the pooled sample, all 1997 respondents kept their income as reported, while the incomes of the 2011 respondents were deflated by the ratio $\frac{\text{sample mean income 1997}}{\text{sample mean income 2011}}$ to correct for the income growth between 1997 and 2011. The second version of the pooled sample mirrored this procedure, but all 2011 respondents kept their income as reported, while the incomes of the 1997 respondents were multiplied by the ratio $\frac{\text{sample mean income 2011}}{\text{sample mean income 1997}}$.

For each country, the VTT was derived from sample enumeration, applying three different combinations of model and sample version (shown in column one, two and six). Column one shows the average VTT resulting from the sample enumeration applying the NL97 (SE94) model on the first version of the pooled sample (reflecting income levels -97 and -94). Column two shows the average VTT resulting from the sample enumeration applying the NL11 (SE08) model on the second version of the pooled sample (reflecting income levels -11 and -08).

In the Swedish data the t-statistics are computed applying the Delta method as explained in Section 4.3. In the Dutch data, the t-values were computed by varying the estimated coefficients in Eq. (5) between +/- 1.96 times their standard deviation, taking correlations between parameters into account and again applying a sample enumeration, resulting in a distribution of the average VTT. The Swedish sample sizes are larger than the Dutch, implying higher significance levels.

However, in the main model we did not account for the panel structure of the data (repeated choices from the respondents). Limited tests were made with the Swedish long-distance data to account for the panel structure of the data. They indicate that the recognition of the panel nature of the data increases the error estimates, in particular in the 2008 data. For these reasons the significance levels are overstated, which needs to be considered when interpreting our results.

The VTT has increased between the years for some modes and purposes and decreased for others in the Dutch data, despite the higher incomes in the 2011 sample. In the Swedish data, the average VTT is lower in the later survey for all modes and purposes (but since standard errors are overstated the change might not be significant) except for long distance bus. The decline is five to ten percent.

Column six derives the VTT by sample enumeration, applying the NL97 (or SE94) model to the second version of the pooled sample (reflecting the income levels of

NL11 and SE08). This column can be interpreted as the VTT for NL11 or SE08 that would have been forecast in the nineties, given that the income growth had been known. By comparing column six with column two, reflecting the real VTT outcome for the later year, we get an indication of the changes in the VTT that cannot be explained by the increase in income, assuming that the cross-sectional income elasticity estimated on the sample from the nineties applies over time, and that the experimental design and trip variable are correctly modelled. For Sweden, we find a 14 percent decline for rail trips (for long distance and for the two purposes of regional trips) and a 7 to 18 percent decline for bus. For the NL the results vary much more from an increase of 41 percent for business rail trips to a decline by 41 percent for local public transport business trips. It is hard to find any clear pattern or plausible explanation for this variation.

Table 13 shows a model estimated on the pooled samples SE94, SE07 and SE08 for car. As shown by Börjesson et al. (2012a) there are no significant difference in the VTT for SE94 and SE07 when accounting for differences in income, design, distance, purpose etc. However, the dummy for SE08 is -0.361 and significantly different from zero, implying that the VTT for SE08 is 70 percent of the VTT for SE07 ($\exp(-0.361)$).

Table 3 The average simulated VTT, The Netherlands, euro of 2011

	Number of respondents			Mean VTT 1997 model 1997 income		Mean VTT 2011 model 2011 income		VTT change (incl. income effect)		Income change		Inter-temporal income elasticity		Mean VTT 1997 model 2011 income		VTT change (constant income)	
	(1997)	(2009)	(2011)	1		2		3		4		5		6		7	
				Value	t-ratio	Value	t-ratio	Value	t-ratio			Value	t-ratio	Value	t-ratio	Value	t-ratio
Commute																	
Car	908	1008	150	12.10	8.2	10.90	13.0	-10%	-0.9	29%	6.8	-0.34	-0.9	14.85	8.2	-27%	-2.9
Train	416	699	105	7.90	7.7	7.94	10.2	1%	0.0	15%	2.8	0.03	0.0	8.72	7.7	-9%	-0.8
Local PT	293	469	97	6.61	7.6	4.30	10.3	-35%	-4.6	21%	3.9	-1.66	-2.9	7.05	7.6	-39%	-5.4
Business																	
Car	866	349	236	13.03	6.2	10.33	11.1	-21%	-2.3	30%	8.4	-0.69	-2.2	16.20	6.1	-36%	-4.9
Train	226	235	41	6.59	7.1	11.39	8.8	73%	2.8	28%	2.9	2.61	2.0	8.08	7.0	41%	1.9
Local PT	69	61	11	4.72	4.9	5.32	5.0	13%	0.6	84%	4.3	0.15	0.6	9.00	4.7	-41%	-3.4
Other																	
Car	461	538	93	7.23	7.1	7.48	12.2	3%	0.3	23%	4.7	0.15	0.3	8.18	7.0	-9%	-0.9
Train	542	249	79	5.26	7.4	4.32	7.8	-18%	-2.0	-3%	0.5	5.98	0.5	5.20	7.4	-17%	-1.9
Local PT	327	136	70	5.50	7.3	3.95	9.0	-28%	-3.7	-2%	0.3	14.02	0.3	5.47	7.3	-28%	-3.6

Table 4 The average simulated VTT, Sweden, euro of 2008

	Number of respondents		VTT 1994 model 1994 income		VTT 2008 model 2008 income		VTT change (incl. income effect)		Income change		Inter-temporal income elasticity		VTT 1994 model 2008 income		VTT change (c income	
	(1994)	(2008)	1		2		3		4		5		6		7	
			Value	t-ratio	Value	t-ratio	Value	t-ratio	Value	t-ratio	Value	t-ratio	Value	t-ratio	Value	t-ratio
Commute																
	Train		1295	2830	6.85	186	6.06	149	-11%	-14.30	-0.21	-8.62	7.07	186	-14%	-
Bus			600	2991	4.90	110	4.63	118	-6%	-4.58	-0.12	-5.77	5.02	110	-8%	-
Other																
	Train		1568	2598	3.75	32.34	3.51	153	-6%	-2.06	-0.09	-1.85	4.07	32.3	-14%	-
Bus			1800	2734	2.89	27.05	2.56	122	-11%	-2.99	-0.10	-1.92	3.11	27.1	-18%	-
Long distance																
	Train		4531	3416	9.39	157	8.69	156	-8%	-8.68	-0.37	-19.59	10.13	157	-14%	-
Bus			1631	4744	6.40	95.6	6.80	191	6%	5.24	0.20	6.75	7.34	95.6	-7%	-

5.2 Income elasticity

In this section we first derive the income elasticity calculated based on the change in the VTT between the two years and the income change. Note that this can only be interpreted as an intertemporal income elasticity *under the assumption that all changes in VTT can be attributed to income changes once differences in the trip and design variables have been controlled for*. This assumption might, however, not hold, for instance if the marginal utility of time changes due to changes in travel comfort or productivity, as suggested by the discussion in the introduction. We still believe that these figures are relevant as a benchmark, because the problem of determining how the VTT changes between the base year and future years in CBA is often handled by only applying an income elasticity for expected income changes after the base year.

Column three in Table 3 and 4, shows the percentage change in the mean VTT computed from column one and two, and also the t-values of the change. Column four reports the percentage change in the mean income of the samples. The income increase is substantial in most cases, except for other trips by train and public transport in the Dutch data (which is likely to be related to the small sample size in the 2011 survey). Column five reports the implied income elasticity. The t-statistics for the income elasticity are computed based on the standard errors of the change in the VTT and in the income.

Column five shows that only one of the income elasticities for private trips is significantly different from zero in the Dutch data, and that this is negative, and that almost all are negative in the Swedish data (negative income elasticities are not micro-economically consistent).

In practical applications, an inter-temporal income elasticity is often taken to be an estimated cross-sectional elasticity even if there is no strong reason why they should be equal. For comparison, Table 5 therefore reveals the cross-sectional elasticities in the Dutch and the Swedish samples. In the Swedish data, the income elasticity equals the estimated income parameter since we take the log of the utility function in the estimation. In the Dutch sample the elasticity was calculated by increasing all incomes by 1 percent in the same sample enumeration tool with which the average VTT was calculated. The Swedish numbers for public transport show a tendency towards a higher cross-sectional income elasticity in the later year. However, in the Dutch samples we see the opposite and more varied results: lower cross-sectional elasticities in the later year (with one exception). There is thus no conclusive evidence regarding the pattern of the cross-sectional elasticities for the two countries. The SE94 elasticities were relatively low compared to the international evidence (Table 1). The NL97 elasticities were relatively high. So, in both countries the more recent elasticities moved towards the international mean elasticities.

Table 5 also show the income elasticities estimated for the Swedish pooled car samples in Table 13. In this model the cross-sectional income elasticity increased with income (as also found in Börjesson et al. (2012a)). This is consistent with

the higher cross-sectional income elasticity in the later Swedish sample for public transport, where income is higher.

Table 5 Cross-sectional income elasticities by year and country.

Purpose and mode	NL97	NL11	Purpose and mode	SE94	SE07	SE94 SE07+SE 08 car
Commute Train Local PT	0.69 0.32	0.65 0.48	Commute Train & Bus	0.05	0.40	
Other Train Local PT	0.37 0.23	0.25 <i>Undef</i>	Other Train & Bus	0.10	0.30	
Business Train Local PT	0.81 1.08	0.74 0.78	Long distance Train & Bus	0.22	0.29	
Car Commute Other Business	0.78 0.57 0.81	0.40 0.36 0.78	Car Commute and Others			0.0;1:25; 1.45

For the Swedish car data different cross-sectional elasticities were estimated for different sections of the income distributions, but these were not significantly different for the two years.

5.3 Discussion

We have observed an unexplained reduction in the VTT in all cases but the Swedish exact replication SE94-SE07. For the NL data there is also an unclear pattern of increases and decreases which is puzzling. For the Swedish PT modes, the decline is more consistent over modes and purposes. We also see a 30 percent lower VTT in the SE08 data than in the SE07 data for car.

There are two potential explanations for the lower VTT in many of the later surveys. The first is that the VTT has in many cases declined over time despite income increases. The second is that VTT is not in fact reduced, and that such findings are an artefact of the survey or experimental method and design or survey conduct. The first explanation could be valid if the travel time has become more comfortable or productive as explained in the introduction. It could also be an effect of changes in other factors that we cannot control for. If for instance reliability has increased over time, this might have reduced the VTT, since a small time reduction might be interpreted as pointless if the travel time normally varies more than the time reduction.¹¹ On the other hand, if crowding or congestion have increased, this would work in the opposite direction and increase the VTT. There is however no evidence that reliability or crowding has increased over time for the bus and train services in the NL or in Sweden (long-distance and regional).

¹¹ However, analysis of the NL11 survey checked whether there was evidence of confounding between the VTT and the value of travel time reliability but did not find such evidence (Significance et al., 2013).

The first explanation could also be valid if the decline in the VTT is related to changes in the income distribution in the samples. Our models account for the income dependency of VTT, but if the functional form of this dependency is different from the one that we have assumed, a shift in the shape of income distribution could result in a lower estimate of the VTT for the same income level in later years. The income distribution has become slightly more skewed to the right over time in the two countries. Still, the shifts are modest. The Gini coefficient increases slightly in the two countries from 0.276 (1997) to 0.282 (2011) in The Netherlands and from 0.227 (1994) to 0.289 (2008) in Sweden. Moreover, in the Swedish sample, the income among the sixth of respondents with the lowest income has not increased between the years.¹² This could be one contributing factor for a possible decline in the VTT, but even if the income distribution has become slightly more skewed over time, this would probably not explain the large decline in the VTT that we find.

There is no reason to believe that differences in reliability, congestion or income distribution would be found for car drivers between 2007 and 2008, explaining the 30 percent lower VTT in the latter. However, the Swedish 2008 data was collected during the early part of the financial crisis. It cannot be ruled out that this impacted the results. Still, the crisis did not immediately impact the economy, and we have controlled for the respondents' incomes in the models.

Moving to the second explanation, we identified in section 3 five differences between the yearly samples that we cannot entirely control for in the analysis. The first is that the number of attributes differs. The Swedish data for bus and train had four attributes in the 1994 experiment and two attributes in the 2008 experiment. Hess et al. (2020) suggest that the two-attribute experiment (the simple time-money trade off) tends to result in lower VTT than experiments with more attributes.

As discussed in section 3, the difference in the presentation of the reference time and cost in the experiment, differences in the type of choice questions, differences in the level of the time and costs of the reference trips, the time and cost differences between the alternatives, and differences in socio-economic composition might have impacted the results even though we did our best to control for the differences in the model estimation. Furthermore, there is a possibility of larger self-selection of travellers with low VTT caused by unobserved factors in the later surveys, due to the fall in response rate in the later surveys, and in Sweden also the change in the recruitment method (making it easier for busy PT travellers with high VTT to escape recruitment and even avoid being approached).

The second explanation seems more consistent with the Swedish replication car data and the RP studies (i.e. Swärdh (2012) and Daly and Fox (2012)), all showing that the VTT increases with income. In the Swedish replication study, none of the differences regarding stated choice experiment or recruitment process between

¹² In the Dutch data the eight income classes were differently defined so this is difficult to compare.

the yearly samples were present. Moreover, it is unlikely that self-selection into the sample of drivers with low VTT would be much higher in the later wave of the replication study. This is because recruitment was identical in the two waves and conducted by number plate registration while the cars were moving, so busy drivers could not avoid being recorded during the recruitment. The response rates after recruitment of all *approached* travellers were still lower in the later wave, but the fall is modest in comparison to the other studies presented in this paper (65 percent in 1994 and 55 percent in 2007).

In the SE08 car data (with 30 percent lower VTT than in the SE07 data), the response rate had dropped to 36 percent and the recruitment method changed to a exogenous random population sample. Exogenous population samples are also normally applied when collecting NTS and other travel survey data, for which a general decline in response rates over the past decades has become a major issue (Prelipean et al., 2018). Hence, selection bias could be a problem also in travel survey data.

The home interview RP data used by Daly and Fox also avoided differences arising from changes in experimental design and changes in the recruitment process. Rather the recruitment was conducted using an exogenous sample of individuals as in the SE08 car data. There can still be self-selection present due to low response rates in this form of data collection as well as in studies applying number plate registration (as in the Swedish car study). However, such recruitment still avoids a self-selection in the recruitment process because busy travellers can avoid being approached or decline recruitment on platforms and at gas stations (and probably in web-panels).

In the Weis et al. (2021) study, also finding stable valuations, the time span between the first and second wave was only five years, and both were designed as follow up surveys to the Mobility and Transport Microcensus (MTMC). This made the recruitment process, experimental design and survey design similar across the survey waves. The reported response rate for the later 2015 sample was also high, 76.9 percent, but this was from a sample already recruited to the MTMC, and the response rate of that MTMC is not reported. The response rates for the 2010 sample are not reported in the paper, but it seems probable that the response rate remained similar over the five years.

A relevant question is then how meta-analysis studies, such as Abrantes and Wardman (2011), can result in income elasticities of the VTT close to unity, while our repeated studies do not. A possibility is that the recruitment is more consistent across the included studies than in our surveys. However, since recruitment method or response rate is usually not included as an explanatory factor in the meta-regressions (or even reported), this is unknown. There are also comparisons across countries/areas with different income levels as well as several RP studies in the meta-analysis.

So, what does the analysis in this paper imply regarding the handling of VTT over time? Either we trust the results from the study and conclude that the VTT has declined over time for the public transport modes in Sweden and that the results

for the NL data are mixed in terms of changes over time. Those results are similar to earlier repeated VTT studies, but are not consistent with results from meta-studies, the Swedish exact replication study and RP studies. The other option is not to trust the results from our analysis, or the other repeated SP studies that are not *exact* replications, because differences in the data collection methods and survey conduct have a bigger impact on the VTT than the real change in the VTT over time.

In the latter case, the remaining issue is how to analyse the VTT over time, if the repeated SP studies cannot be trusted. Or precisely, what data sources can be trusted for an analysis of the VTT and the VTT over time. Since this study does not produce exact results in terms of what has been the source of the bias, Table 6 lists three possible sources. The table gives an overview of the presence of these three sources in five different data types. Repeated SP surveys have errors from all three sources. If the key source of the error is the sensitivity to stated choice survey design, we must turn to RP data. However, RP data has other problems, primarily measurement errors in input variables. This is often larger in the cost variable, which tend to attenuate the cost parameters and lead to under-estimation of the VTT in RP data if this is not addressed (Varela et al., 2018).

However, if the key source of the error stems from self-selection during recruitment on platforms and gas stations, SP studies recruited from a random sample or number plate registration would be reliable. If the main source is sample selection due to low and declining response rates, NTS data or any form of data collected with low response rates in any of the waves might also produce biased results. If that cannot be controlled for by weighting etc., the remaining options are data stemming from passive data collection, such as mobile phone data, road pricing data or ticket sales data. The best option might be to combine SP and RP data, but then it is still essential to understand the strengths and weaknesses of each type of data.

Table 6 Sources of errors of the analysis of the changed in VTT over time

	SP- recruitment at platforms and gas stations	SP- recruitment by internet panels	SP- recruitment by number plate registration or from random sample	RP NTS - recruitment from random sample	Passive data mobile phone data, data from HOT lanes, congestion charges, tickets sales, Register data
Stated choice design	yes	yes	yes	no	no
Sample selection due to low response rate	yes	yes	yes	yes	no

Self-selection in the recruitment process	yes	yes	no	no	no
---	-----	-----	----	----	----

6 CONCLUSION

This paper explores the intertemporal income elasticity of the VTT using data from the Netherlands and Sweden, collected at two points in time, thirteen and fourteen years apart, respectively. We do not attempt to answer how the average VTT has changed over time in the national populations, but analysed only how the VTT has changed for similar individuals in the two samples. The results show mostly a declining VTT for a given income level. The results show also a large within-country heterogeneity across modes and purposes, in the cross-sectional income elasticity of the VTT, and in its development over time. This is essentially the same result as was produced by three of the four earlier studies comparing stated choice data at two points in time to study temporal change in the VTT. The main contribution of this paper is to document the differences between the studies collected at different years, indicating the reasons why it is difficult to uncover temporal changes in the VTT. We have concluded that either the VTT has declined for given income levels for many modes and purposes, or the replication study can for some reason not be trusted.

If the decline in the VTT can be attributed to changes in the survey method and experimental design, formulation of the questionnaire and choice questions, survey mode (paper/computer-aided etc, telephone), this would raise fundamental questions regarding the legitimacy of the SP method for VTT research as such. If the outcome is critically sensitive to survey design, it would also imply that the estimated VTT could not be compared across surveys since survey techniques change over time, due to improvements in experimental design and econometric techniques and in survey technology.

However, another likely explanation is the fall in response rate over time, and that the recruitment method (on platforms and at fuel stations) has made it easier for busy travellers to escape recruitment in the later years. If the key explanation is that the response rate has dropped or that busy travellers can escape being recruited, this would not necessarily invalidate the stated choice method if recruitment methods could be improved to the extent that a representative sample is recruited. Recruiting representative samples of respondents is a challenge for all sorts of studies, including travel surveys (NTS data). Declining response rates in surveys has since long been well documented (De Heer and De Leeuw, 2002). In the literature on election polls the issues of sample selection and representativeness are recognized as a key problem (Chen et al., 2019; Conduit and Akbarzadeh, 2020), but this has possibly not received appropriate attention in the value of time literature. A possible way forward could be to develop and estimate some sort of model of response probability in further research.

Indeed, the response rate also including people approached but declining to be recruited to the survey is often not even reported. If it is not possible to recruit representative samples, the remaining options are data from passive data collection, such as mobile phone data, road pricing data or ticket sales data (Daly et al., 2017) for studying the VTT. We recommend more consideration of methods to make it more difficult for busy respondents to avoid being approached and recruited in future research.

7 DECLARATIONS

Funding Maria Börjesson acknowledges financial support by the Swedish governmental agency VINNOVA. Marco Kouwenhoven did most of this work while affiliated to Delft University of Technology.

Conflict of interest statement: On behalf of all authors, the corresponding author states that there is no conflict of interest.

Author Statement: All authors have contributed to the different parts of the manuscript. Marco Kouwenhoven and Maria Börjesson estimated the models.

Availability of data and material: The authors do not have the right to deposit data

Code availability: We have applied Biogeme for model estimation (Bierlaire, 2003)

8 REFERENCES

- Abrantes, P.A.L., Wardman, M.R., 2011. Meta-analysis of UK values of travel time: An update. *Transportation Research Part A: Policy and Practice* 45, 1–17. <https://doi.org/10.1016/j.tra.2010.08.003>
- Bastian, A., Börjesson, M., Eliasson, J., 2016. Explaining “peak car” with economic variables. *Transportation Research Part A: Policy and Practice* 88, 236–250. <https://doi.org/10.1016/j.tra.2016.04.005>
- Batley, R., Bates, J., Bliemer, M., Börjesson, M., Bourdon, J., Cabral, M.O., Chintakayala, P.K., Choudhury, C., Daly, A., Dekker, T., Drivyla, E., Fowkes, T., Hess, S., Heywood, C., Johnson, D., Laird, J., Mackie, P., Parkin, J., Sanders, S., Sheldon, R., Wardman, M., Worsley, T., 2017. New appraisal values of travel time saving and reliability in Great Britain. *Transportation* 1–39. <https://doi.org/10.1007/s11116-017-9798-7>
- Ben-Akiva, M.E., Lerman, S.R., 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.
- Bierlaire, M., 2003. BIOGEME: A free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

- Börjesson, M., 2014. Inter-temporal variation in the travel time and travel cost parameters of transport models. *Transportation* 41, 377–396. <https://doi.org/10.1007/s11116-013-9493-2>
- Börjesson, M., Eliasson, J., 2014. Experiences from the Swedish Value of Time study. *Transportation Research Part A* 59, 144–158.
- Börjesson, M., Fosgerau, M., Algers, S., 2012a. On the income elasticity of the value of travel time. *Transportation Research Part A: Policy and Practice* 46, 368–377. <https://doi.org/10.1016/j.tra.2011.10.007>
- Börjesson, M., Fosgerau, M., Algers, S., 2012b. Catching the tail: Empirical identification of the distribution of the value of travel time. *Transportation Research Part A: Policy and Practice* 46, 378–391. <https://doi.org/10.1016/j.tra.2011.10.006>
- Chen, J.K.T., Valliant, R.L., Elliott, M.R., 2019. Calibrating non-probability surveys to estimated control totals using LASSO, with an application to political polling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 68, 657–681. <https://doi.org/10.1111/rssc.12327>
- Conduit, D., Akbarzadeh, S., 2020. Pre-election polling and the democratic veneer in a hybrid regime. *Democratization* 27, 737–757. <https://doi.org/10.1080/13510347.2020.1725477>
- Daly, A., Fox, J., 2012. Forecasting mode and destination choice responses to income change. Presented at the IATBR, Toronto.
- Daly, A., Hess, S., de Jong, G., 2012. Calculating errors for measures derived from choice modelling estimates. *Transportation Research Part B: Methodological, Emerging and Innovative Directions in Choice Modeling* 46, 333–341. <https://doi.org/10.1016/j.trb.2011.10.008>
- Daly, A., Sanko, N., Wardman, M., 2017. Cost and time damping: evidence from aggregate rail direct demand models. *Transportation* 44, 1499–1517. <https://doi.org/10.1007/s11116-016-9711-9>
- De Borger, B., Fosgerau, M., 2008. The trade-off between money and travel time: A test of the theory of reference-dependent preferences. *Journal of Urban Economics* 64, 101–115. <https://doi.org/10.1016/j.jue.2007.09.001>
- De Heer, W., De Leeuw, E., 2002. Trends in household survey nonresponse: A longitudinal and international comparison. *Survey nonresponse*.
- Dillén, J., Algers, S., 1998. Further research on the national Swedish value of time study. *Selected Proceedings of the 8th World Conference on Transport Research* 3, 135–148.
- Flügel, S., Halse, A.H., Hulleberg, N., Jordbakke, G.N., Veisten, K., Sundfør, H.B., Kouwenhoven, M., 2020. Verdsetting av reisetid og tidsavhengige faktorer. Dokumentasjonsrapport til Verdsettingsstudien 2018-2019 (No. 1762/2020). Transportøkonomisk institutt, Oslo, Norway.
- Fosgerau, M., 2005. Unit income elasticity of the value of travel time savings, in: *Proceedings of the European Transport Conference*.
- Fox, J.B., 2015. Temporal transferability of mode-destination choice models (phd). University of Leeds.
- Gunn, H., 2001. Spatial and temporal transferability of relationships between travel demand, trip cost and travel time. *Transportation Research Part E: Logistics and Transportation Review, Advances in the Valuation of Travel Time Savings* 37, 163–189. [https://doi.org/10.1016/S1366-5545\(00\)00023-5](https://doi.org/10.1016/S1366-5545(00)00023-5)

- Gunn, H.F., Bradley, M.A., Rohr, C., 1996. The 1994 national value of travel time savings study of road traffic in England. Paper for the Course and Seminar on Value of Travel Time Savings. Easthampstead, Berkshire, England.
- Gunn, H.F., Tuinenga, J.G., Cheung, Y., Kleijn, H.J., 1999. VALUE OF DUTCH TRAVEL TIME SAVINGS IN 1997. Presented at the World Transport Research: Selected Proceedings of the 8th World Conference on Transport ResearchWorld Conference on Transport Research Society.
- Hague Consulting Group, 1998. The second Netherlands' value of time study: final report. Report 6098-1 for AVV. HCG, Den Haag.
- Hanly, M., Dargay, J., Goodwin, P., 2002. Review of Income and Price Elasticities in the Demand for Road Traffic.
- Hess, S., Daly, A., Börjesson, M., 2020. A critical appraisal of the use of simple time-money trade-offs for appraisal value of travel time measures. *Transportation*. <https://doi.org/10.1007/s11116-020-10097-w>
- Hess, S., Daly, A., Dekker, T., Cabral, M.O., Batley, R., 2017. A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research. *Transportation Research Part B: Methodological* 96, 126–149. <https://doi.org/10.1016/j.trb.2016.11.002>
- Kouwenhoven, M., de Jong, G.C., Koster, P., van den Berg, V.A.C., Verhoef, E.T., Bates, J., Warffemius, P.M.J., 2014. New values of time and reliability in passenger transport in The Netherlands. *Research in Transportation Economics, Appraisal in Transport* 47, 37–49. <https://doi.org/10.1016/j.retrec.2014.09.017>
- Manski, C.F., Lerman, S.R., 1977. The estimation of choice probabilities from choice based samples. *Econometrica: Journal of the Econometric Society* 1977–1988.
- Prelipcean, A.C., Susilo, Y.O., Gidófalvi, G., 2018. Collecting travel diaries: Current state of the art, best practices, and future research directions. *Transportation Research Procedia* 32, 155–166.
- Ramjerdi, F., Flügel, S., Samstad, H., Killi, M., 2010. Value of time, safety and environment in passenger transport (No. Transportøkonomisk institutt).
- Ramjerdi, F., Rand, L., Sætermo, I.-A.F., Saelensminde, K., 1997. The Norwegian Value of Time Study Part I (No. 379/1997). TØI report, Oslo, Norway.
- Shires, J.D., de Jong, G.C., 2009. An international meta-analysis of values of travel time savings. *Evaluation and Program Planning, Evaluating the Impact of Transport Projects: Lessons for Other Disciplines* 32, 315–325. <https://doi.org/10.1016/j.evalprogplan.2009.06.010>
- Significance, VU University, John Bates Services, TNO, NEA, TNS NIPO and PanelClix, 2013. 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.
- Swärdh, J.-E., 2008. Is the intertemporal income elasticity of the value of travel time unity? Working Paper 2008:3. Statens väg- och transportforskningsinstitut.
- Tapley, N., Wardman, M., Gunn, H., Hyman, G., 2007. Inter-temporal variations in values of time in Great-Britain. Presented at the Paper presented at the European Transport Conference.

- Varela, J.M.L., Börjesson, M., Daly, A., 2018. Quantifying errors in travel time and cost by latent variables (No. 2018:3), Working papers in Transport Economics. CTS - Centre for Transport Studies Stockholm (KTH and VTI).
- Wardman, M., 2014. Price Elasticities of Surface Travel Demand A Meta-analysis of UK Evidence. *Journal of Transport Economics and Policy (JTEP)* 48, 367–384.
- Wardman, M., 2004. Public transport values of time. *Transport Policy* 11, 363–377. <https://doi.org/10.1016/j.tranpol.2004.05.001>
- Wardman, M., 2001. Inter-temporal variations in the value of time. ITS Working Paper 566, December 2001.
- Wardman, M., 1998. The Value of Travel Time: A Review of British Evidence. *Journal of Transport Economics and Policy* 32, 285–316.
- Wardman, M., Chintakayala, P., de Jong, G.C., Ferrer, D., 2012. European wide meta-analysis of values of travel time. Paper prepared for EIB, ITS Leeds.
- Wardman, M., Chintakayala, V.P.K., de Jong, G., 2016. Values of travel time in Europe: Review and meta-analysis. *Transportation Research Part A: Policy and Practice* 94, 93–111. <https://doi.org/10.1016/j.tra.2016.08.019>
- Weis, C., Kowald, M., Danalet, A., Schmid, B., Vrtic, M., Axhausen, K.W., Mathys, N., 2021. Surveying and analysing mode and route choices in Switzerland 2010–2015. *Travel Behaviour and Society* 22, 10–21. <https://doi.org/10.1016/j.tbs.2020.08.001>

9 APPENDIX

9.1 Appendix A: survey descriptions

9.1.1 Netherlands, 1997-survey: NL97

In 1997, a national VTT study was conducted in the Netherlands to determine the VTT in passenger transport (Hague Consulting Group, 1998). The respondents were recruited *en route* (e.g. at petrol stations, parking garages, railway stations and bus stops). 10815 people were approached, of whom 81 percent indicated that they were willing to participate and provided the interviewer with their addresses. Of the 8738 persons that were sent a questionnaire, 5157 responded, i.e. a response rate of 59%.

The SP experiment was a simple binary time/cost experiment with explicit reference to a recently made trip (Figure 2).

	A	(1)	B
Travel time	SAME as now		Travel time 20 MINUTES LONGER than now
Travel cost	SAME as now		Travel cost £ 2,00 LOWER than now

Figure 2 Example of an SP choice in the 1997 survey

Since this survey was a paper-and-pencil survey, only four versions were created: one for each of four travel time classes. The time and cost levels of the eleven binary choices in each version were only different by a scale factor (1, 2, 3 or 4), so that the implicit boundary values of time in these choices were always the same for all respondents.

9.1.2 Netherlands, 2009-2011-survey: NL09 and NL11

In 2013, results from a new national value of travel time study were officially released (Kouwenhoven et al., 2014; Significance et al., 2013). Stated choice data were collected in 2009 (using an existing commercial online panel) and 2011 (using *en-route* recruitment of travellers as in 1997).

In the 2011 surveys, each respondent was asked to answer three SP experiments. Experiment 1 was a simple time-cost trade-off experiment, with choice screens appearing as in Figure 3.

Trip A	Trip B
<i>Usual travel time:</i> 60 min.	<i>Usual travel time:</i> 45 min.
<i>Cost:</i> € 2.80	<i>Cost:</i> € 3.60

Figure 3 Example of SP question of experiment 1 for car respondents in 2011 survey.

In addition to this stated choice experiment, second and third experiments to determine the value of travel time reliability were added. The 2011 recruitment locations were mostly the same as in 1997; however, several new locations were added mainly to recruit respondents for public transport.

The sample statistics by mode and purpose are given in Table 7. For all trip purposes, travel time in the 2011-survey is lower for car and local PT but higher for train compared to the 1997-survey (see Table 7). The travel time and cost changes, ΔT and ΔC , are smaller in the later study for all modes and purposes except for train - business. The differences in travel time of the reference trip, and the differences in ΔT and ΔC between the surveys are related to differences in the recruitment method and in the experimental design.

Table 7 Sample statistics of the 1997 and 2011 Dutch data: mean values. All costs are given in prices of 2011.

Purpose	Mode	Year	Time min.	Cost €	ΔT min.	ΔC €	Gender (Male=1; Female=0)	Age	Net annual household income €/year	Nobs
Work	Car	1997	61.4	-	16.8	1.59	0.75	37.1	40881	908
		2011	46.5	3.97	7.2	1.11	0.85	44.2	52932	150
	Train	1997	73.4	-	19.2	1.83	0.60	36.2	33059	416
		2011	76.1	6.72	10.7	1.59	0.59	38.8	40781	105
	Local PT	1997	59.0	-	16.6	1.57	0.57	38.3	29777	293
		2011	54.1	3.23	8.0	1.20	0.60	38.6	31795	97
Business	Car	1997	77.5	-	19.9	1.89	0.87	37.3	37785	866
		2011	60.3	6.19	8.9	1.99	0.94	44.8	54997	236
	Train	1997	96.4	-	23.0	2.18	0.62	37.0	30691	226
		2011	114.7	15.01	14.3	3.31	0.73	39.5	47947	41
	Local PT	1997	70.6	-	18.5	1.76	0.45	35.7	26105	69
		2011	65.1	2.45	8.8	1.86	0.73	41.8	50103	11
Other	Car	1997	74.1	-	18.7	1.77	0.63	40.4	32725	461
		2011	60.5	6.60	9.0	1.34	0.76	48.2	41534	93
	Train	1997	99.1	-	23.2	2.20	0.43	32.1	23472	542
		2011	100.6	6.54	13.1	1.89	0.61	32.3	27255	79
	Local PT	1997	64.2	-	17.2	1.63	0.36	31.3	21467	327
		2011	48.5	2.80	7.3	1.06	0.65	39.8	28795	70

9.1.3 Sweden, 1994 and 2008-surveys: SE94 and SE08

Two national VTT studies for all modes of passenger transport have been conducted in Sweden, the first in 1994 (Dillén and Algers, 1998) and the second in 2007-2008 (Börjesson et al., 2012a; Börjesson and Eliasson, 2014); the references include details of the experimental design. The sampling of respondents and stated choice design experiments were not identical due to methodological development. The sample statistics of the data are included in Table 8. The travel time differences ΔT were similar in the surveys. However, the travel cost differences ΔC are substantially larger in the 2008 survey. The main reason for this is that the bid range was extended to better capture the tail of the VTT distribution (Börjesson et al., 2012b), which was not considered in the 1994 survey.

The recruitment of respondents for the public transport studies was conducted on board while travelling, in 1994. If they accepted the invitation, the 1994 subjects were mailed the survey questionnaire on paper, where details of the observed trip could be filled in as well as sheets for the stated choice interview. They were later contacted by telephone, repeatedly until reached, but at most seven times. When reached, a computer aided telephone interview was then undertaken on an agreed day, collecting all information on the mailed survey questionnaire and the stated choices. (Hence the paper questionnaire was never handed in but was only used by respondents to support their memory.) In 2008 the travellers were recruited on public transport platforms (hence it might have been easier for busy travellers to avoid recruitment). If they accepted

participating in the study, the respondents' addresses and telephone numbers were collected. They then received a link to a web-based questionnaire; however, they could choose to respond to the questionnaire by a call-back telephone interview instead to reduce potential selection bias. Few respondents were interviewed over the phone. The questionnaires used in the 1994 and in 2008 were of a similar length. However, since the 1994 survey was responded to over the phone, we can assume it took more time than in 2008.

Table 8 reports monthly gross income. Sweden has a progressive tax system, but the system remained relatively similar between the two years.

Table 8 Sample statistics of the 1994 and 2008 data: mean values. All costs are given in prices of 2008.

Purpose	Mode	Year	Time min	Cost €	ΔT min	ΔC €	Gender (Male =1; Female=0)	Age	Gross income €/month	Nobs
Work	Bus	1994	36.7	3.2	8.0	0.7	0.50	38.2	1660	600
		2008	41.1	3.8	8.3	1.7	0.40	41.6	2455	2991
	Train	1994	37.5	4.2	7.9	0.7	0.47	36.9	1808	1295
		2008	37.9	3.4	7.7	1.7	0.40	41.7	2811	2830
Other	Bus	1994	52.4	5.5	10.5	0.9	0.39	33.4	749	1568
		2008	42.1	4.0	8.5	1.8	0.32	40.6	1596	3244
	Train	1994	46.9	5.9	9.8	0.8	0.36	36.0	963	1800
		2008	43.8	4.7	8.9	2.0	0.32	39.8	1697	2975
Long distance	Bus	1994	237	19.1	36.8	3.3	0.36	40.4	1266	4531
		2008	226	19.0	45.4	10.3	0.32	41.0	1651	4744
	Train	1994	253	40.3	37.0	3.0	0.43	37.5	1665	1631
		2008	214	30.3	43.0	10.1	0.31	39.6	2000	3416
				Distance km						
All purposes	Car	1994	106	124.8	17.1	15.0	0.60	46.5	1265	3947
	Car	2007	113	131.9	18.3	19.4	0.64	51.9	1744	3301
	Car	2008	80.0	92.8	16.2	42.8	0.59	48.7	1609	8523

In both surveys, the instructions to the respondents and the choice context were virtually the same. The context of the SP experiment was that of a recently made trip and the attribute levels were pivoted around the observed levels (the 'reference' values). However, the presentations of the alternatives differed between the surveys. In the first Swedish survey, only the alternative with the reference trip was described with absolute attribute levels (on the reference card, see Figure 4). The attribute levels on of alternative trips were expressed in terms of difference in relation to the levels of the reference trip (i.e., the travel time was 10 minutes longer than the reference trip, the travel cost was 1 euro cheaper than the reference trip). However, the departure time and number of transfers were given in absolute levels. In the 2008 survey, the absolute time and cost levels were presented for all alternatives (see Figure 5).

Interviewer reads over the phone "You should now consider different bus trips. Suppose that you, before the trip, could choose among the bus trips that I present to you. The trips will differ with respect to travel time, travel cost, headway of the first service, the number of transfers and transfer time. The trips will be presented in pairs, and I will ask you that for every pair of trips you state which of the two you prefer."

TRIP A	TRIP B
Total travel time <u>45</u> min	The travel time is <u>5</u> min longer
Travel cost <u>50</u> SEK	The travel cost is <u>10</u> SEK lower
First departure every <u>15</u> min	First departure every <u>5</u> min
<u>0</u> transfers	<u>1</u> transfers of total <u>10</u> min

Figure 4: Example of an SP choice and instructions in the 1994 survey. Remember that the interview was designed as a telephone interview and that subjects were asked to, during the interview, write down the attribute levels on the choice cards sent to them in advance.

Instructions appearing on the screen prior to making the choices: Timetables can change and travel costs can go up or down. This means that your travel times and travel costs could be different. We ask you now consider different bus trips. Suppose that, before making the trip, you could have chosen among different bus trips, differing with respect to travel time and travel cost. The trips will now be presented to you in pairs. For every pair we ask you to state which of the two trips you prefer.

Which alternative would you choose?	
Alternative A	Alternative B
Total travel time 45 min	The total travel time is 5 min more
Travel cost 50 SEK	The trip costs 10 SEK less

Figure 5: Example of an SP choice and instructions in the 2008 survey. The survey was conducted online so the binary choices were presented to the respondents on the computer screen.

9.2 Appendix B: Final model estimates

Table 9: 1997 Dutch model results

	Work		Business		Other	
Number of estimated parameters:	12		12		12	
Number of observations:	17369		12485		13629	
Final log-likelihood:	-8532.7		-6264.2		-7007.6	
Rho-square:	0.268		0.215		0.256	
Name	Value	t-test	Value	t-test	Value	t-test
Scale	-1.267	-64.4	-1.095	-49.3	-1.371	-54.5
Reference VTT for car	14.85	33.3	19.52	32.7	7.113	30.9
... for train	12.18	25.3	12.91	17.9	6.444	25.5
... for local PT	10.27	20.7	12.05	9.7	5.806	19.6
Size effect for time	-0.9115	FIXED	-0.9603	FIXED	-0.6178	FIXED
Sign effect for cost	0.2104	FIXED	0.08498	FIXED	0.08496	FIXED
Sign effect for time	0.2338	FIXED	0.2063	FIXED	0.1981	FIXED
Gender = female	-0.1798	-6.6	-0.1456	-3.2	-0.2455	-10.4
1-person household	0.3403	6.2	0.1761	2.7	0.2208	4.5
Age = 36 to 50	-0.1102	-3.7				
Age = 51 and older	-0.3356	-9.7	-0.2244	-4.8	-0.4265	-17.3
Income factor for car	0.2177	16.1	0.2286	18.1	0.1698	10.3
... for train	0.204	14.7	0.2387	16.7	0.1357	9.8
... for local PT	0.104	4.4	0.3014	23.8	0.09338	4.6
Travel Time elasticity						
... for car / train			-0.317	-7.2		
... for car / local PT					-0.119	-3.0
... for train					-0.2988	-6.9
... for train / loc.PT	-0.172	-3.6				
... for local PT			-0.7454	-4.9		

Note: β_T , η_C and η_T are constrained to the 2011-estimated values. Also note that all costs and incomes are in 2011 prices.

Table 10: 2011 Dutch model results

	Work		Business		Other	
Number of estimated parameters:	18		12		16	
Number of observations:	15168		5598		6990	
Final log-likelihood:	-7024.8		-2585.6		-2874.6	
Rho-square:	0.233		0.274		0.262	
Name	Value	t-test	Value	t-test	Value	t-test
Scale	-0.9205	-24.2	-0.9867	-17.5	-1.216	-19.3
Reference VTT for car	13.25	7.0	11.19	7.4	5.827	6.6
... for train	10.56	7.3	14.64	7.6	3.955	8.7
... for local PT	6.652	7.5	5.812	5.5	3.225	6.7
Size effect for time	-0.9116	-10.9	-0.9603	-11.2	-0.6177	-7.1
Sign effect for cost	0.2104	6.3	0.08499	1.9	0.08495	2.3
Sign effect for time	0.2338	9.4	0.2063	5.5	0.1981	6.4
Gender = female	-0.1446	-3.4			0.2262	3.1
1-person household	0.1359	1.9	0.3331	2.4	-0.2121	-3.2
Age = 36 to 50	-0.1456	-3.0			-0.2107	-3.3
Age = 51 and older	-0.3476	-8.5	-0.2972	-5.4	-0.4896	-11.5
2009 respondents	-0.5736	-17.6	-0.5287	-12.0	-0.3634	-6.6
Income factor			0.2067	11.0		
... for car	0.08859	3.5				
... for train	0.1814	9.3				
... for car / train					0.09499	3.8
... for local PT	0.1389	4.0				
Travel Time elasticity						
... for car					0.327	4.4
... for car / train	0.1570	2.5				
... for local PT	-0.4973	-3.5			-0.5931	-3.7
Scale for 2011 data	0.9011	17.5	0.9539	15.1	0.7782	13.0

Table 11: 1994 Swedish model results. Other socio-economic variables did not significantly impact WTP once income was controlled for

	Work		Other		Long Distance	
Number of estimated parameters:	17		17		17	
Number of observations:	3368		1895		6162	
Final log-likelihood:	-1508.80		-765.49		-2998.09	
Rho-squared:	0.354		0.417		0.298	
Name	Value	t-test	Value	t-test	Value	t-test
Constant	3.47	3.70	3.55	2.27	1.39	3.68
WTP dummy, rel. to other types	-0.89	-4.68	-0.804	-4.95	-0.66	-12.45
Cost c	1		1		1	
Time t	-0.713	-2.45	0.091	0.44	0.067	1.10
Time difference Δt	1.01	2.68	-0.026	-0.09	0.21	3.27
Train dummy	0.238	1.62	0.339	2.03	0.31	4.11
Income	0.102	1.03	0.0493	0.32	0.22	6.40
Income, missing	0.881	1	0.526	0.35	1.97	6.09
Transfer penalty for train	142	6.90	137	3.42	202	8.02
Transfer penalty for bus	138	6.20	85.4	7.33	138	0
Wait Bus 0-10 min	15.5	4.97	12.9	6.11	119	3.68
Wait Bus 11-30 min	7.53	4.82	9.61	6.62	6.08	1.87
Wait Bus > 30 min	6.12	3.64	9.56	4.31	4.54	2.73
Wait Rail 0-10 min	25.9	7.01	26.7	3.7	80.6	4.28
Wait Rail 11-30 min	17.3	6.66	12.6	3.4	10.2	3.77
Wait Rail > 30 min	7.62	3.17	12.2	2.57	12.5	5.29
Scale μ_1 for train	10.3	9.7	11.9	5.76	24.5	14.07
Scale μ_2 for bus	7.83	8.37	11.8	10.39	25.0	14.39

Table 12: 2008 Swedish model results. Other socio-economic variables did not significantly impact WTP once income was controlled for

	Work		Other		Long Distance	
Number of estimated parameters:	11		11		11	
Number of observations:	6220		5821		8160	
Final log-likelihood:	-2296.54		-2473.34		-3603.23	
Rho-squared:	0.467		0.387		0.363	
Name	Value	t-test	Value	t-test	Value	t-test
Constant	0.942	2.11	-0.311	-0.48	1.85	5.92
WTP dummy, rel. to WTA	-0.476	-5.29	-0.497	-6.54	-0.446	-8.12
EG dummy	-0.082	-1.30	-0.119	-2.16	-0.198	-4.52
EL dummy	-0.273	-3.34	-0.346	-4.81	-0.289	-5.43
Time t	-0.405	-3.75	-0.254	-2.56	-0.529	-6.84
Time difference Δt	0.540	5.39	0.54	5.91	0.675	9.42
Train dummy	0.303	4.37	0.296	5.12	0.226	5.64
Income	0.304	7.89	0.401	6.53	0.29	11.24
Income, missing	3.01	7.88	4.13	6.59	3.36	12.83
Scale μ_1 for train	-9.86	-17.99	-9.49	-17.86	-12.1	-25.34
Scale μ_2 for bus	-11.5	-18.23	-11.4	-18.96	-13.7	-23.48

Table 13: Model results for the pooled Swedish 1994, 2007 and 2008 data, all purposes.

	Car	
Number of estimated parameters:	20	
Number of observations:	15771	
Number of individuals:	2226	
Final log-likelihood:	-6666.507	
Rho-squared:	0.39	
Name	Value	t-test
Year 2008 dummy	-0.361	-4.24
Year 1994 dummy	0.0739	0.79
Mean of base log VTT β	0.576	0.98
Standard deviation of log VTT β_σ	1.15	35.63
WTP dummy, relative to WTA	-0.74	-19.27
EG dummy	-0.478	-8.09
EL dummy	-0.314	-5.32
Log time	-0.04	-0.37
Log time difference	0.303	6.07
Log distance	0.217	2.6
Log (income < 50) β_l	-0.135	-1.09
Log (income 50) β_{l50}	1.25	3.79
Log (income 75) β_{l75}	1.45	6.04
Purpose dummy for Recreation	-0.17	-1.53
Purpose dummy for School	0.119	0.38
Purpose dummy for Service	-0.249	-2.33
Indicator for employed people	0.418	5.33
Purpose dummy for Commute	-0.0291	-0.26
Scale for SE08 data	1.64	13.23
Scale for SE94 and SE07 data	1.44	10.06

Table 14: 1994 Swedish model sensitivity results. The first model assumes additive errors instead of multiplicative as in our main model, but is otherwise identical to the corresponding main model. The first mixed logit model (MXL 1) adds a normally distributed error term, constant within individuals, to the Gumbel distributed error term in the utility function. The second mixed logit model (MXL 2) adds a normal distributed error term to W in Eq. (8). The mixed logit models are otherwise identical to the corresponding main model.

	Additive errors		MXL 1		MXL 2	
Number of estimated parameters:	17		18		18	
Number of observations:	6162		6162		6162	
Number of individuals	6162		853		853	
# Random draws	-		1000		1000	
Final log-likelihood:	-3035.75		-2969.01		-2888.26	
Rho-squared:	0.289		0.305		0.324	
Name	Value	t-test	Value	t-test	Value	t-test
Constant	3.07	8.41	1.37	3.77	1.28	2.54
WTP dummy, rel. to other types	0.175	2.98	0.190	3.02	0.237	3.42
Cost c	-0.194	-3.7	0.092	1.55	0.057	0.82
Time t	0.187	3.04	0.297	4.06	0.350	4.07
Time difference Δt	-0.647	-12.11	-0.663	-11.74	-0.710	-12.97
Train dummy	0.213	6.57	0.217	6.48	0.224	4.71
Income	1.91	6.22	1.93	6.17	1.98	4.45
Income, missing	1		1		1.00	
Transfer penalty for train	210	9.9	198	8.04	200	7.78
Transfer penalty for bus	138	0	138	0	138	0.00
Wait Bus 0-10 min	15.5	0	112	3.72	179	3.84
Wait Bus 11-30 min	8.66	1.64	6.39	1.81	5.93	2.00
Wait Bus > 30 min	9.14	3.49	4.35	2.40	4.42	2.92
Wait Rail 0-10 min	42.1	2.09	80.5	4.29	84.3	4.20
Wait Rail 11-30 min	22.1	5.52	10.2	3.75	9.15	3.77
Wait Rail > 30 min	13	5.2	12.7	5.34	12.3	5.39
Scale μ_1 for train	0.308	-26.55	25.4	13.19	27.3	12.91
Scale μ_2 for bus	0.354	-45.75	26.9	14.80	35.0	11.53
Standard Dev. Mixing distribution					0.580	16.58
Standard Dev. Mixing distribution for train			0.405	2.41		
Standard Dev. Mixing distribution for train			0.684	10.63		
Mean VTT train* €/h					11.4	
Mean VTT train* €/h					7.55	

**Average VTT simulated with 1994 model, 1994 income and design variables, corresponding to column (1) Table 4*

Table 15: 2008 Swedish model sensitivity results. The first model assumes additive errors instead of multiplicative as in our main model, but is otherwise identical to the corresponding main model. The first mixed logit model (MXL 1) adds a normally distributed error term, constant within individuals, to the Gumbel distributed error term in the utility function. The second mixed logit model (MXL 2) adds a normal distributed error term to W in Eq. (8). The mixed logit models are otherwise identical to the corresponding main model.

	Additive errors		MXL 1		MXL 2	
Number of estimated parameters:	11		12		12	
Number of observations:	8160		8160		8160	
Number of individuals:	-		1132		1132	
# Random draws	-		1000		1000	
Final log-likelihood:	-3913.01		-3602.33		-3075.86	
Rho-squared:	0.467		0.363		0.456	
Name	Value	t-test	Value	t-test	Value	t-test
Constant	1.000	2.4	1.86	5.94	1.17	2.53
WTP dummy, rel. to WTA	0.369	4.66	0.671	9.37	0.345	6.31
EG dummy	-0.382	-4.54	-0.526	-6.81	-0.183	-2.49
EL dummy	0.285	7.17	0.225	5.60	0.277	4.83
Time t	-0.107	-2.05	-0.290	-5.40	-0.207	-4.87
Time difference Δt	-0.278	-4.86	-0.448	-8.09	-0.413	-9.56
Train dummy	0.409	11.8	0.290	11.26	0.287	7.34
Income	4.50	12.9	3.36	12.86	3.29	8.33
Income, missing	-0.255	-4.82	-0.199	-4.51	-0.246	-6.62
Scale μ_1 for train	-0.256	-28.25	-12.2	-24.90	-21.8	-24.85
Scale μ_2 for bus	-0.228	-23.49	-13.8	-23.17	-29.4	-21.73
Standard Dev. Mixing distribution			0.237	2.48	0.775	29.37
Mean VTT train* €/h					11.7	
Mean VTT train* €/h					91.7	

**Average VTT simulated with 2008 model, 2008 income and 1994 design variables, corresponding to column (2) Table 4*