# Master Thesis

## The Impact of Travel Cost Reimbursement on the Value of Travel Time

Levi Mulder

Delft University of Technology

TU Delft

# Master Thesis

## The Impact of Travel Cost Reimbursement on the Value of Travel Time

by


Levi Mulder


to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday August 26, 2024 at 13:00.

| | |
|---|---|
| Supervisors: | Dr. E.J.E. Molin, Prof.dr. O. (Oded) Cats, Dr. M. Kouwenhoven, Ir. J Muller |
| Project Duration: | February, 2024 - August, 2024 |
| Faculty: | Faculty of Civil Engineering and Geo-sciences<br>& Faculty of Technology Policy and Management, Delft |

Cover: Rolex Watch Closeup by Martijn Mulder (Modified)



TU Delft

# Voorwoord

Beste lezer, voor u ligt mijn afstudeerscriptie, ik hoop dat u kan interesseren met mijn schrijven en dat u er wellicht wat wijzer van wordt. Ikzelf heb in elk geval een hoop geleerd tijdens het doen van dit onderzoek. Dit voorwoord schrijf ik ongeveer drie weken voordat ik mijn bevindingen mag presenteren, hier zou ik graag nog een aantal mensen willen bedanken waarvoor de tijd wellicht tekort schiet tijdens de presentatie. Want nu het moment is aangebroken waarop mijn afstuderen, met al zijn pieken en dalen, tot een slotakkoord komt, realiseer ik mij weer eens dat je niks bent zonder de mensen om je heen.

Allereerst wil ik mijn ouders bedanken die mij een sterke emotionele basis hebben meegeven waardoor ik in staat ben geweest tegenvallers op te vangen en successen te vieren. Verder wil ik ook bedanken; Roelof Akkermans die mij enthousiasmeerde voor Techniek en Natuurkunde, Bert Rijpkema voor zijn support gedurende mijn eindexamenjaar op het Linde College zodat ik nog last-minute wiskunde-B kon toevoegen aan mijn eindlijst en kon beginnen aan de bachelor werktuigbouwkunde aan de TU Delft, Julie voor haar steun en liefde wat mij hielp om lange dagen te maken, Mijn huisgenoten die voor en tijdens mijn afstuderen altijd wilden luisteren naar mijn overpeinzingen en gedachten, en kritisch doch constructief met mij meedachten, Mijn vrienden voor het vrienden blijven ondanks het weinige contact wat wij eventueel hadden tijdens mijn afstuderen, Natuurlijk mijn afstudeerbegeleiders vanuit de TU Delft; Dr. Eric Molin en Prof. Oded Cats voor hun begeleiding, positieve insteek en constructieve feedback, Jeroen Muller en Marco Kouwenhoven vanuit Significance voor het bieden van een geweldige afstudeerplek met een overvloed aan expertise, Jeroen in het bijzonder voor zijn immer kritische feedback.

Als de confetti van het afstuderen is neergedaald ga ik drie maanden op reis met Julie in Europa. Daarna zal ik mij storten in het burgerleven en hopelijk iets positiefs bijdragen aan onze samenleving.

*Levi Mulder*
*Delft, August 2024*

*This page is intentionally left blank.*

# Preface

Before you lies my graduation thesis. I hope to engage you with my writing and perhaps even impart some knowledge. I have certainly learned a great deal throughout this research process.

I am writing this foreword approximately three weeks before I present my findings, and I would like to take this opportunity to thank several people, since there might not be enough time during the presentation. As my graduation, with all its highs and lows, draws to a close, I am reminded once again that we are nothing without the people around us.

First and foremost, I would like to thank my parents, who have given me a strong foundation, enabling me to cope with setbacks and celebrate successes. Additionally, I would like to thank Roelof Akkermans for sparking my enthusiasm for Technology and Physics, Bert Rijpkema for his support during my final year at the Linde College, allowing me to add Mathematics B to my final list at the last minute and start my Bachelor's in Mechanical Engineering at TU Delft, Julie for her support and love, which helped me endure long days, my housemates, who were always willing to listen to my musings and thoughts and provided critical yet constructive feedback, my friends for remaining friends despite the limited contact we may have had during my graduation, of course, my graduation supervisors from TU Delft: Dr. Eric Molin and Prof. Oded Cats, for their guidance, positive approach, and constructive feedback, Jeroen Muller and Marco Kouwenhoven from Significance, for providing a fantastic graduation spot with an abundance of expertise, and Jeroen in particular for his consistently critical feedback.

When the confetti from graduation has settled, I will travel through Europe for three months with Julie. After that, I will immerse myself in civilian life and hopefully positively contribute to our society.

*Levi Mulder*
*Delft, Augustus 2024*

# Summary

To determine whether an investment in an infrastructure project, e.g. a new train track, is worthwhile often a cost-benefit analysis is used. This cost-benefit analysis (CBA), as conceived by Dupuit (1844) is used to determine whether the benefits of a particular investment outweigh the costs. In the context of infrastructure projects, the largest benefits are often the travel time savings as a result of a shorter travel time for the (new) users of that piece of infrastructure. To trade off the monetary costs required to build e.g. the railroad and the travel time savings one needs a ”conversion factor”. This factor is referred to in practice as the value of travel time ($VTT$). Since governments have limited resources for infrastructure projects, accurately determining the $VTT$ helps prioritize projects that deliver the greatest societal benefits.

In practice, this $VTT$ is often determined through discrete choice experiments (DCE). In a DCE a respondents are asked to make trade-offs between time, money and other attributes. Through the estimation of a mathematical model, one can extract the $VTT$. However one can imagine that making this tradeoff is somewhat different for respondents who are normally reimbursed for their travel cost. These respondents normally do not have to consider costs, or part of them, in their day-to-day travel decisions. Therefore their value of travel time might be different. If such a difference exists one would want to take this into account when calculating national $VTT$ averages. Thus the research question is:

*To what extent does a travel cost reimbursement affect the value of travel time?*

We investigate the impact of travel cost reimbursement (TCR) on the value of travel time ($VTT$) using both the NL2022 (Netherlands) and UK2014 (United Kingdom) datasets. We hypothesise three mechanisms which could cause a $VTT$ difference: whether TCR influences the $VTT$ directly, if respondents with a TCR make less consistent choices, and if these respondents exhibit a higher willingness to accept (WTA) compared to those without TCR. Utilizing Multinomial Logit (MNL) and Mixed Logit (ML) models, we analyse the effect of receiving TCR on the $VTT$ while allowing for discrete scale heterogeneity. Furthermore, by considering the respondents' reference situation, we seperate a TCR interaction on the $VTT$ into willingness to accept (WTA) and willingness to pay (WTP) components. The models developed for this research account for socio-economic interactions and ensure that the findings are robust across different experimental setups.

The results indicate that commuter respondents, in the NL2022 dataset, with a full Travel Cost Reimbursement have a $VTT$ which was approximately 29% - 45% higher compared to respondents who do not receive a reimbursement. The same type of respondents in the UK2014 dataset exhibit a $VTT$ which was approximately 58% - 143% higher. This effect is significant and consistent for experiments in which only time and cost were presented as attributes of the alternatives. As the number of attributes in the choice task increases the difference in $VTT$ is no longer significant. Furthermore, we are unable to discover any consistent patterns regarding scale heterogeneity between the $VTT$ for nTCR and TCR respondents. Therefore we cannot accept our hypothesis that respondents who receive a travel cost reimbursement make less consistent choices. Moreover, the interaction with the $VTT$ was more predominant in the $WTA$ domain compared to the $WTP$ domain. However, the model which took this into account did not perform statistically better in terms of model fit compared to a model which only allowed for equal TCR interaction on both the $WTA$ and the $WTP$ quadrant.

We conclude that the $VTT$ is significantly different for commuting travellers when receiving and not receiving a travel cost reimbursement. However, these changes cannot be explained through scale heterogeneity or a WTA, WTP disparity. We argue that national $VTT$ values should take travel cost reimbursement into account as national $VTT$ values are in practice based on the 2-attribute time-cost experiments.

# Contents

# Nomenclature

## Abbreviations

| Abbreviation | Definition |
|---|---|
| *TCR* | Travel Cost Reimbursement (Refers to *any* travel cost reimbursement) |
| *fTCR* | full Travel Cost Reimbursement |
| *pTCR* | partial Travel Cost Reimbursement |
| *nTCR* | no Travel Cost Reimbursement |
| *NL2022* | Dutch National Stated Preference Dataset 2022 (Kouwenhoven et al., 2022) |
| *UK2014* | UK National Stated Preference Dataset 2014 (ARUP, 2015) |
| *ODIN 2023* | Dutch National Travel Survey (CBS, 2024) |
| *WTP* | Willingness To Pay |
| *WTA* | Willingness To Accept |
| *CBA* | Cost-Benefit Analysis |
| *SCBA* | Societal Cost-Benefit Analysis |
| *SP* | Stated Preference |
| *RP* | Revealed Preference |
| *DCE* | Discrete Choice Experiment |

## Symbols

| Symbol | Definition | Unit |
|---|---|---|
| $U$ | Utility | - |
| $V$ | Systematic Utility | - |
| $VTT$ | Value of Travel Time | [€ /h] |
| $\overline{VTT}$ | Weighted Value of Travel Time | [€ /h] |
| $\mu$ | Scale Parameter | - |
| $\varepsilon$ | Error Term | - |

# 1 Introduction

To determine whether the investment in an infrastructure project, e.g. a new train track, is worthwhile, governments often use a cost-benefit analysis. This cost-benefit analysis (CBA), as conceived by Dupuit (1844) is used to determine whether the benefits of a particular investment outweigh the costs. In the context of infrastructure projects, the largest benefits are often the travel time savings as a result of a shorter travel time for the (new) users of that piece of infrastructure (Romijn & Renes, 2013). To balance the monetary costs of e.g. building a railroad, against the benefits of travel time savings, a 'conversion factor' is used to quantify the value of those time savings. This factor, known as the value of travel time (VTT), is expressed in euros per hour. By multiplying the travel time benefits with this VTT one can associate a monetary value with the travel time benefits. Since governments have limited resources for infrastructure projects, accurately determining the VTT helps prioritize projects that deliver the greatest societal benefits.

The trivial way to determine this VTT would be to ask people directly, *"What are you willing to pay to shorten your journey by 1 hour?"*. However, people are not "programmed" to explicate the way we make choices. This is also empirically proven in for example location research into refugee camps (Liebe et al., 2018). When people are asked whether to explicate their choices they tend to give socially desirable answers. In this example by Liebe et al. (2018), respondents initially *stated* their willingness to welcome refugees. However, when given the opportunity to *choose* between different policies varying the number of refugees near their homes, their actual choices revealed a far lower willingness than previously expressed. On top of that, economic theory (McFadden, 1973) is based on what people do, thus what choices they make, not on their judgements. Therefore, in practice and academics, there are two methods to determine the VTT. Revealed preference (RP) as conceived by Samuelson (1938) and stated preference (SP) as introduced by Louviere and Hensher (1983) and Louviere and Woodworth (1983).

RP uses real-world travel behaviour to determine the VTT. The underlying principle is to observe the travel routes people take and construct hypothetical alternative travel routes for each person. This creates a choice which can be used to determine how people value travel time. The advantage is that the analyst knows a lot about the context in which a choice is made. Context such as; weather, congestion, time of day etc. can easily be observed in revealed preference studies. However, the attributes, travel time, travel cost, reliability, comfort etc which the individual uses to determine their choices are not known to the analyst and must therefore be assumed. These variables can be measured but as these often correlate heavily with one another, called multicollinearity, it is difficult to attribute a choice to a specific variable. For example, in real life, a route choice is often quicker but also cheaper. Therefore it is in practice difficult to know exactly whether the respondent made his route choice based on the journey being cheap or on being quick. Furthermore, all of the options, thus the number of possible routes in the choice set are also not known to the analyst.

To overcome these drawbacks researchers can turn to stated preference (SP). Here, respondents (people who answer the questionnaire) are presented with hypothetical choice situations which can be fully controlled by the researcher. One example of such a choice set is shown in Figure 1.1. Here respondents are asked to make a tradeoff between travel cost and travel time. One of the options is cheaper but slower, the other one is quicker but more expensive. Furthermore, contrary to RP, in SP the alternatives (choice options) and the number of attributes (characteristics) can be fully controlled by design. It is thus possible to present respondents with choices which they normally are not able to make in real life. These characteristics make SP a popular method to determine the VTT and are one of the reasons this method is used in the Netherlands. (Kouwenhoven et al., 2022)

| Rit A | Rit B |
|---|---|
| Reistijd:<br>**45 min.** | Reistijd:<br>**60 min.** |
| Kosten:<br>**€ 7.80** | Kosten:<br>**€ 5.10** |

**Figure 1.1:** Example of a stated preference choice set from Kouwenhoven et al. (2022)

To calculate the VTT, from an SP study, one needs to present respondents with variations of a choice task as in Figure 1.1. This is done by changing the travel time and the travel costs. Thereafter, the researcher can calculate the value of travel time by analysing the choices using a mathematical model. The most popular in practice and acclaimed in science is the multinomial logit model (MNL) as conceived by McFadden (1973). It is rooted in the economic framework of utility maximisation. Utility, in economics, refers to the satisfaction or benefit that consumers derive from consuming goods or services. This framework postulates that consumers aim to maximise utility by choosing the goods or services that provide the most utility to them, taking into account their individual preferences and constraints. The exact theory needed to calculate this VTT is covered in chapter 2.

One can imagine that making this tradeoff could potentially be somewhat distanced for respondents who receive a travel cost reimbursement (TCR). These respondents suddenly have to take costs into account, which they normally might not or only partially. Travel costs might not show up in their bank account and it often does not impact their monthly expenses. Therefore, if such a respondent bases their choices on their usual day-to-day travel decisions, their VTT could be significantly higher as compared to respondents who have to pay for the travel costs themselves. As an impressive 80% of employees in the Netherlands receive some form of travel cost reimbursement, the number of people in the population to whom a different VTT might apply is extensive (Bachaus, 2023). It would therefore be relevant to evaluate whether the VTT for respondents with and without a travel cost reimbursement differs indeed, as the VTT is one of the most important parameters in allocating money to projects which yield the greatest societal benefits. This relationship is, to the author's knowledge, not yet covered in the scientific literature, as will be elaborated on in section 1.1.

## 1.1. Research Gap

As was stipulated above the focus of this study is to investigate whether respondent travel cost reimbursement impacts the value of travel time. In this section, this aim is analysed by synthesising the body of scientific knowledge. We therefore aim to pinpoint the ”edges” of the research gap. Do note that the full literature review is saved for chapter 2. Only the literature strictly necessary to understand the hypotheses is discussed here.

To postulate hypotheses we first briefly examine the literature on travel cost reimbursement and travel behaviour. Many different forms of reimbursement have been examined. Some are incentivizing travel through a financial bonus, (Baldassare et al., 1998; Bueno et al., 2017) others penalizing some form of travel through a financial penalty. (Baldassare et al., 1998; Evangelinos et al., 2018; Raux et al., 2015) All aim to quantify the effects of some policy on respondent preferences. A good example is provided by Bueno et al. (2017) who investigated the relationship between commuter benefits and transportation mode choice. They found that benefit programs designed to encourage walking, cycling, and using public transport were effective in increasing the use of these modes of transportation. One interesting finding is that all studies focus on some form of reimbursement scheme and mode-choice behaviour. None of these articles review to what extent a travel cost reimbursement could impact the VTT.

We must therefore turn to the behavioural sciences to establish a line of reasoning. We argued earlier that respondents who normally are reimbursed for their travel costs might base their choices partially or fully on their reimbursed reference situation. If respondents with a TCR base their choices on their reference situation we would thus expect them to more frequently choose the more expensive option because cost is of less, or no, importance in their reference situation, therefore increasing their VTT. The size of the potential difference will thus be dependent on the extent to which these respondents base their choices on their “normal reimbursed” situation. Why respondents would base their decisions on their reimbursed situation is discussed in chapter 2.

This culminates in the **first hypothesis**; We hypothesise that respondents with a travel cost reimbursement will have a significantly higher VTT as compared to respondents who do not receive a travel cost reimbursement.

Additionally, there might be other behavioural mechanisms which could cause respondents with a travel cost reimbursement to alter their choice behaviour and therefore indirectly affect the VTT. Respondents with a travel cost reimbursement might have a lower choice task motivation because their choices normally have no direct consequences in real life. This is called lack of payment consequentiality (Zawojska et al., 2019) or lack of budget constraints (Ding et al., 2005). From the point of view of the respondent, this in essence means that they feel like; ”I don’t really care about these choices as my travel costs are reimbursement anyhow.” Therefore they could exert less cognitive effort when making choices. It has been shown that less cognitive effort and choice task motivation could lead to a lower choice consistency (Jang et al., 2022; Swait & Adamowicz, 2001). Choice consistency in its turn can bias model parameters (Chen et al., 2016). The reasoning is that a higher degree of randomness in the choices can be captured as a lesser degree of importance of the model parameters, therefore lowering the VTT. As such, a lower degree of choice consistency for respondents with a travel cost reimbursement could indirectly lower the VTT, whilst in essence, the VTT might actually be higher. This effect could therefore also dampen the direct effect postulated in Hypothesis 1. The exact workings of how choice consistency can bias model parameters is left to chapter 2. Furthermore, other potential behavioural mechanisms are also discussed in chapter 2.

This culminates in our **second hypothesis**; We hypothesise that respondents with a travel cost reimbursement make less consistent choices as compared to respondents without a travel cost reimbursement.

There is a third way, a third hypothesis, through which travel cost reimbursement could impact the VTT. It might be possible that respondents with a travel cost reimbursement reasonably assume that cost levels up to their reference costs are reimbursed and costs above their reference level are not. This assumption is not unreasonable when one considers that there are many forms of travel cost reimbursement which are fixed at a certain limit. If this assumption echoes through in their choices, they will opt more often for the more faster but expensive route if the costs are below their reference, and they will more often choose the cheaper but slower route whenever the costs are above their reference costs. This would thus imply two different values for the VTT. We provide a visual interpretation in Figure 1.2.

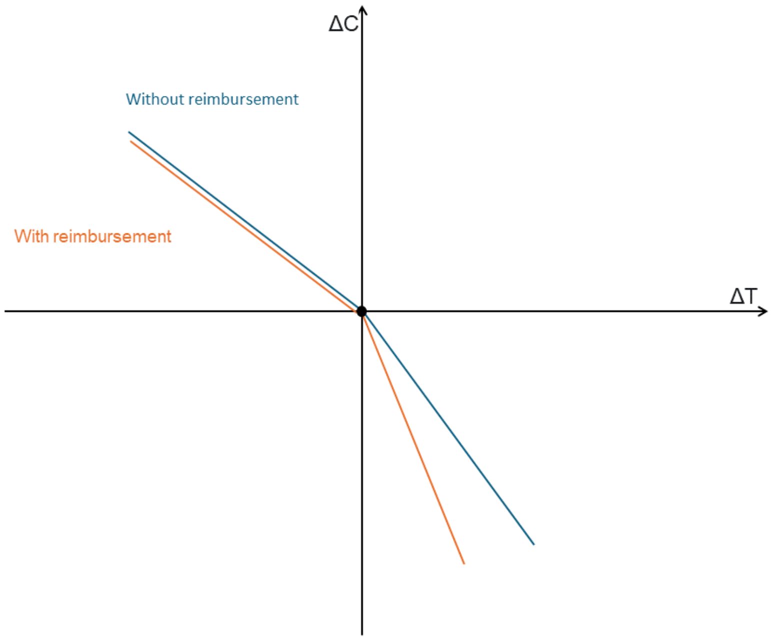



**Figure 1.2:** Visual representation of Hypotheses 3. Indifference curves for a two-attribute time, money trade-off. The graph is centred on the reference travel time and travel costs of the respondents. $\Delta T$ and $\Delta C$ respectively denote the time and the cost difference relative to the respondent's reference time and cost levels.

The upper left quadrant in Figure 1.2 provides information on the willingness to pay (WTP) to save a certain amount of travel time. The lower right quadrant provides information on the willingness to accept (WTA) a certain deterioration in travel time. Whenever the WTA and the WTP are equal the logical result is indeed a straight indifference curve. An indifference curve is sometimes also referred to as an iso-utility curve. Every point on this curve thus has the same utility. This implies that a respondent is equally likely to choose between any two options which are on the indifference curve. Whenever modelling results are discussed one can "translate" this discontinuity into a difference between the "willingness to pay" (WTP) and the "willingness to accept" (WTA).

Do observe that the discontinuity in Figure 1.2 as described above is also present for respondents without a reimbursement. This discontinuity at $\Delta T = 0, \Delta C = 0$ is referred to as a *sign effect* in the choice modelling literature(Hess et al., 2007). It is often explained through the behavioural phenomena known as loss aversion(Tversky & Kahneman, 1991). This sign-effect is well documented and it appears in both scientific (Hess & Train, 2017a) and applied work (Kouwenhoven et al., 2022). However, as also visualised Figure 1.2 we argue this well-known discontinuity might very well be more prominent for respondents with a reimbursement as compared to respondents without.

To summarise our **third hypothesis**; We hypothesise that respondents with a travel cost reimbursement have a significantly higher willingness to accept (WTA) as compared to respondents without a travel cost reimbursement.

It is important to state that these hypotheses are all stated under the condition that all other variables are equal, *ceteris paribus*. This naturally poses a challenge for model development to ensure other effects are accounted for. Therefore the mathematical formalisation of the hypotheses is preserved for chapter 2.

Let's for a moment *assume* that it does turn out that respondents with a travel cost reimbursement have a significantly higher VTT. Would that imply that these respondents have a value of travel time which is ”wrong”? Not necessarily, a stated choice experiment is designed to reveal respondents’ preferences. It might very well be the case that respondents who receive a travel cost reimbursement do have different preferences as compared to respondents who are not reimbursed for their travel costs. Researchers aim to establish a model, based on theory, which explains the observed choices the best. Unless proven otherwise we must assume that a potential difference is a true reflection of respondent preferences. However, that does not mean that the results of this study are irrelevant. On the contrary, whenever the results indicate a significant VTT difference between reimbursed and non-reimbursed respondents, a change in national average VTT metrics could be justified. Therefore potentially changing the outcome of societal cost-benefit analysis.

## 1.2. Goal & Research Question

Thus to further understand the relationship between travel cost reimbursement and the value of travel time the author aims to answer the following research question:

*"To what extent does a travel cost reimbursement affect the value of travel time?"*

The sub-questions are presented below following the previously stated hypotheses. chapter 2 further elaborates on the necessary theory and relevant literature for each of these sub-questions.

1. For whom and for which attribute types does a travel cost reimbursement impact the $VTT$?
2. Do respondents with a travel cost reimbursement make less consistent choices?
3. Does a travel cost reimbursement impact the WTP and the WTA differently?
4. To what extent do sign effects and choice consistency explain an effect of TCR on the $VTT$?
5. What is the impact of the findings on national cost-benefit analysis?

By answering this research question and its subquestions the contribution to the literature is **threefold**: It will potentially provide analysts tasked with calculating value of travel time metrics grips on how to take travel cost reimbursement effects into account. Secondly, it will provide researchers insight into the choice behaviour of respondents with a travel cost reimbursement. Lastly, it could further inform policymakers about the effects of travel cost reimbursement and how it changes people's preferences.

We tested our hypotheses on the Dutch national (Kouwenhoven et al., 2022) value of travel time study and the UK national (ARUP, 2015) value of travel time study. These studies both documented whether or not a respondent received a travel cost reimbursement. As we a priori expected potential effects to be small, the large sample sizes of these studies were deemed to be of use. Secondly, these studies can both be subdivided into multiple sub-studies, which varied the number of attributes. This enabled us to test the hypotheses across these sub-studies and therefore allow for a more varied analysis and generalize the results to a greater extent.

Furthermore, these two datasets enabled us to evaluate the hypotheses for two different types of instructions. Respondents for the NL2022 dataset were instructed to ignore their reimbursement and make choices as if they had to pay out of their own pocket. The UK2014 survey instructed respondents to take a travel cost reimbursement into account if applicable.

To generalize the results as much as possible we aimed to develop both Multinomial- (McFadden, 1973) and Mixed Logit models (McFadden & Train, 2000). Furthermore, non-parametric and parametric analyses had to be performed to ensure that the travel cost reimbursement parameters did not capture any other socioeconomic effects. Testing the statistical significance of the final model parameters, enabled the evaluation of possible reimbursement effects. The specifics of the methods are elaborated in chapter 4.

## 1.3. Structure

This report is structured as follows; In chapter 2 the necessary literature and theory are presented to fully understand the hypotheses which are formalized in chapter 3. Subsequently, in chapter 4 the method to test these hypotheses is discussed. This Method chapter covers the model development process, the data samples used, model estimation and finally model application on a case study. Thereafter chapter 5 discusses the results structured around the hypotheses and ends with the model application results on the case study. The findings are discussed and summarized in chapter 6 and discusses subsequently the; methodological implications, policy implications, limitations and suggestions for further research.

An expert reader could opt to start at the hypotheses in chapter 3 as they are already familiar with the theory discussed in chapter 2. Non-expert readers can also choose this approach and go back to the theory whenever they are missing background information.

# 2

# Theory

This chapter covers all the necessary theories to understand the hypotheses in chapter 3, it starts with the fundamentals, the $VTT$ and Travel Cost Reimbursement. The concepts thereafter build on these fundamentals. Firstly, we provide a brief literature review on travel cost reimbursement and the value of travel time. The aim of the literature review was to identify a literal research gap. Furthermore, as stated in chapter 1, this body of literature is limited; therefore, this section is relatively short. Secondly, we define what the Value of Travel Time is from a microeconomic viewpoint, how it can be retrieved and how it is calculated. Thirdly, we define what travel cost reimbursement entails in the Netherlands and the UK. The goal here is to understand what it practically means when a respondent states that they receive a travel cost reimbursement, aiding our understanding of the behavioural foundations. Fourthly, we discuss the interpretation of the scale parameter and its link with choice consistency. Both scientific literature and a brief synthetic demonstration are used as evidence of its relevance to the value of travel time (VTT). Lastly, we elaborate on the behavioural notion of loss aversion and its empirical counterpart, the sign effect, and why they are relevant to the VTT.

## 2.1. Literature Review

The purpose of the literature review was to identify a research gap, thus to evaluate to what extent a relationship between the value of travel time and travel cost reimbursement had been investigated. The method used to identify relevant papers was as follows; First, based on informal expert talks at Significance relevant concepts were identified. These talks culminated in the search query which was used:

```
   ("stated preference" OR (choice AND experiment)) AND (
(travel AND allowance)
OR (travel AND expenses AND reimbursement)
OR (employer AND paid)
)
```

Variations on this query were also tested but did not yield more relevant results. On the 28th of March, this query resulted in 15 documents on the Scopus scientific search engine. The papers, resulting from the search query, were scanned and irrelevant papers were discarded. The remaining papers were read thoroughly and provided a basis for further snowballing. The final selection of papers is presented in Table 2.1. Do note that the objective of this review was to identify the research gap. The *behavioural literature* to support the hypotheses are covered in the *next* sections.

**Table 2.1:** Summary of studies on financial bonuses/penalties to impact travel behaviour.

| Reference | Retrieved | Type of model | Topic |
|---|---|---|---|
| (Evangelinos et al., 2018) | query | MNL | Pricing workplace parking via cash-out: Effects on modal choice |
| (Baldassare et al., 1998) | query | Regression | Suburban attitudes toward policies aimed at reducing solo driving |
| (Farahmand et al., 2021) | query | ML | MaaS for TDM |
| (Bueno et al., 2017) | query | MNL | The relationship between commuter benefits and mode choice |
| (Bachaus, 2023) | snowballing | - | Policy analysis of tax-free travel cost reimbursement in the Netherlands |

#### Literature Synthesis

The current body of research into travel cost reimbursement logically focuses on means to initiate a modal shift away from the car (Baldassare et al., 1998; Bueno et al., 2017) and not on the impact on the value of travel time. This is of course to be expected as the need to stimulate a model shift is arguably an urgent topic in the context of climate change. Nonetheless, analysis of the effects of financial bonuses or penalties on mode choice behaviour can shed light on the choice behaviour of respondents who receive a travel cost reimbursement.

The first general observation is that different forms of travel cost reimbursement or extra travel cost payments have been the topic of investigation.

Evangelinos et al. (2018) has researched the impact of an increase in travel costs through a workplace parking scheme. One of the main takeaways was that an increase in workplace parking prices has a significant effect on car mode-choice probability. The underlying mechanism which was used in this study was rather clever. Respondents were presented with an option of receiving money if they were to give up their parking place. This ensured that respondents did not interpret the scheme as ”taking away their parking places” but rather as a bonus when they chose not to take the car. These findings are in line with the observations of Baldassare et al. (1998). They noted that a reimbursement (a financial bonus) yields a greater likelihood of changing respondents’ mode choice as compared to a financial penalty.

Bueno et al. (2017) came to similar conclusions when evaluating different forms of travel cost reimbursement on mode shares. He found a significant relationship between commuter benefits and transportation mode choice. Specifically, he found that benefits such as public transport cost reimbursement and mileage reimbursement influenced an individual’s decision on how to commute.

One of the most interesting findings was reported by Farahmand et al. (2021). They investigated the potential role of Mobility-as-a-Service (MaaS) as a transport demand management (TDM) tool to influence commuting mode choice. Farahmand et al. (2021) concluded that price plays an important role in the development of MaaS packages. Furthermore, a substantial number of respondents in his sample received a travel cost reimbursement, 38.56%. They however stated that the importance of the price attribute is important even for respondents who receive a travel cost reimbursement. If this observation would also hold for route-choice experiments, [1] the value of travel might not differ between reimbursement statuses after all. Unfortunately, they did not include the reimbursement status as a variable in the results, therefore we are not able to further disentangle their observations related to travel cost reimbursement.

Business research focused on evaluating the tax-free level of travel cost reimbursement has been performed (Bachaus, 2023). But yet again the focus has been put on congestion and on labour market effects not so much on the value of travel time. As was already stipulated in chapter 1 the impact of a travel cost reimbursement on the value of travel time has not been investigated to the authors' knowledge.

[1] Unlabeled route choice experiments are typically used to determine the VTT, an example of such an unlabeled route choice experiment was given in Figure 1.1

## 2.2. The $VTT$

Transport professionals often use the terms value of time $(VOT)$, value of travel time $(VTT)$, and value of travel time savings $(VTTS)$ interchangeably. These terms refer to the value associated with reductions in travel time, henceforth referred to as $(VTT)$. In this context, $VTT$ does not refer to the general value placed on time but rather to the value of changes in time spent travelling compared to alternative use of that time (ITF, 2019). This is the relevant metric if you want to determine how travellers react to changes in a network as a result of an infrastructure project. Therefore the $VTT$ provides decision-makers with information about the merits of a transport project. Because the value transport users put on reductions in travel time influences their response to a change in the transport network. The response, for example, in terms of the increase in demand or modal shift, to a project which reduces travel time will be greater if the value of reductions in travel time is high.

The $VTT$ can be determined using stated preferences, in which a questionnaire is distributed where people can make choices between alternatives. Such a choice task looks like Table 2.2.

For example, if a respondent has a $VTT$ of €$35,-/h$ he would be more likely to choose the left alternative as opposed to the right alternative. The $VTT$ is determined by assessing the relative importance of the time versus the cost attribute in these choices. This is typically done through a mathematical model, in which we assume that respondents are utility maximisers, thus that they make choices according to microeconomic theory McFadden (1973). This means that respondents choose the alternative which has the highest utility to them. One often describes the utility of an alternative through a utility function, which is a simplification of reality. An example of such a formulation is given in Equation 2.1, where utility for individual $i$ choosing alternative $j$ is specified as follows:

**Table 2.2:** Example choice set with 2 alternatives

| Route A | Route B |
|---|---|
| **Travel time:** 45 min. | **Travel time:** 60 min. |
| **Cost:** €7.80 | **Cost:** € 5.10 |

$$U_{ij} = \beta_{\text{cost}} \cdot \text{cost}_{ij} + \beta_{\text{time}} * \cdot \text{time}_{ij} + \varepsilon_{ij} \tag{2.1}$$

Where $cost_{ij}$ and $time_{ij}$ represent the attributes of the alternatives, $\beta_{cost}$ and $\beta_{time}$ represent the parameters which need to be estimated on the choices, and $\varepsilon_{ij}$ is an error term. We thus assume that we can describe the utility, which a respondent derives from an alternative, as a linear combination of preferences (the $\beta$'s) and attributes (time and cost). To account for a certain degree of randomness, an error term, $\epsilon_{ij}$ is introduced. This is a random number, a random variable, which is drawn from a distribution for every respondent $i$ and for every alternative $j$. Typically an Extreme Value Type I distribution is used, also known as a Gumbel distribution.

After the model is formulated the researcher aims to find the $\beta_{cost}$ and the $\beta_{time}$ which make the observed choices the most likely. Going into the exact methods used to find these $\beta$'s is outside the scope of this study. After the model is estimated we can calculate the value of travel time as the relative importance of the time and the cost parameter as shown in Equation 2.2. Note that further explanations on how to interpret the $VTT$ is given in chapter 3, where the necessary theory to understand the first hypothesis is discussed.

$$VTT = \frac{\beta_{time}}{\beta_{cost}} \tag{2.2}$$

The value of travel time can be thus interpreted as the amount of money one is willing to pay to reduce travel time by a certain amount. Or equivalently the willingness to accept a deterioration in travel time.

### Calculation of the Value of Travel Time

As was mentioned in section 2.2 the $\beta_{cost}$ and $\beta_{time}$ are estimated to make the observed choices as likely as possible. Thereafter one can calculate the $VTT$ with Equation 2.2. One other option is to estimate the value of travel time directly instead of calculating it ex-post. This has some practical advantages in welfare analysis. To do so the utility function in Equation 2.1 ought to be rewritten to the specification in Equation 2.3.

$$U_{ij} = \mu * (\text{cost}_{ij} + VTT \cdot \text{time}_{ij}) + \varepsilon_{ij} \tag{2.3}$$

where $\mu$ and $VTT$ are now the parameters to be estimated. The interpretation in WTP space is different from a utility specification in preference space. As we estimate the $VTT$ directly, scaling of the utility function is done through the $\mu$ parameter. As the random noise, which is represented in the $\varepsilon$ parameter, is fixed, the $\mu$ parameter determines the respondent's sensitivity to changes in utility. If $\mu$ is zero it means respondents are insensitive to changes in utility and do not base their choices based on changing attribute levels. In other words; the model does not explain the data. Whenever the $\mu$ parameter is very high it means respondents are sensitive to changing attribute levels and are therefore "flip" their choices quickly whenever they are presented with slightly different attribute levels. What this interpretation of the scale parameter means exactly in terms of choice probabilities is left for section 2.4.

### Interpretation of the Value of Travel Time

The mathematical interpretation of the $VTT$ can be best visualized by plotting the choice probabilities and utilities for a choice between two alternatives. We fix alternative 2, setting its travel time to 60 minutes and setting its travel cost to 10 euros. We also fixed the costs for alternative one but we varied its travel time. This produces the MNL choice probabilities as in Figure 2.1.

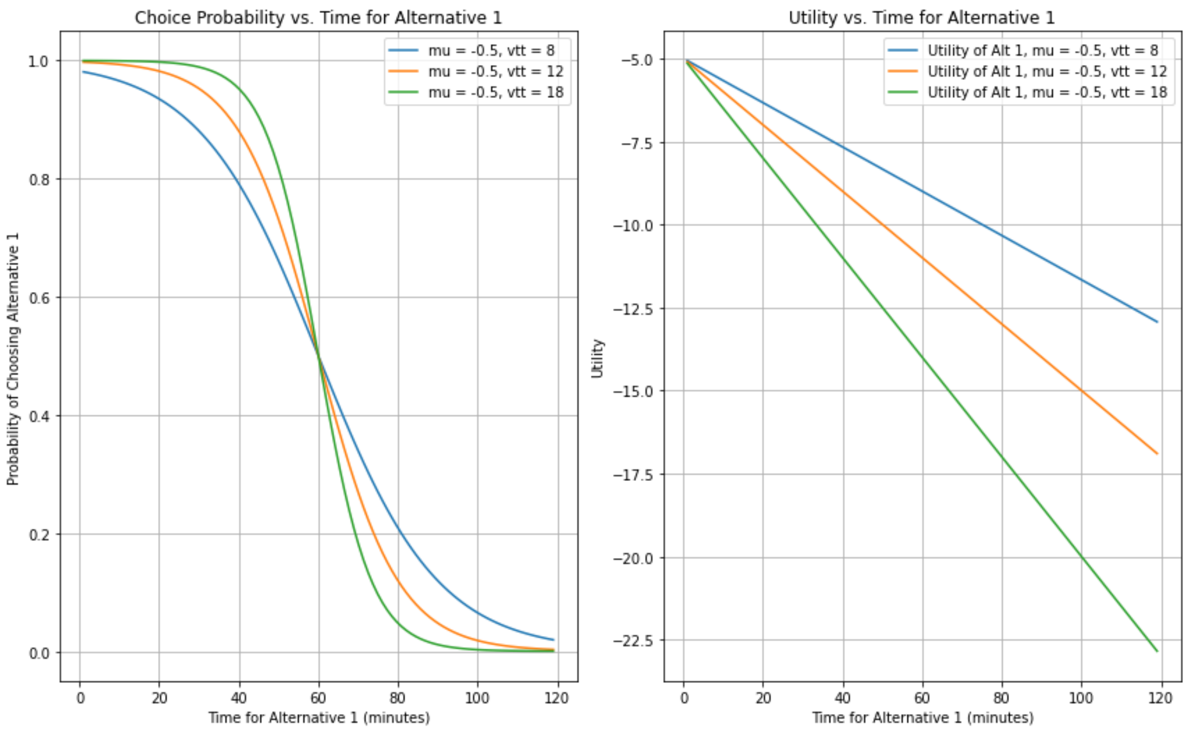



**Figure 2.1:** Choice probabilities and utilities for different $VTT$ values based on choice probabilities of an MNL model. The other alternative is defined with a travel time of 60 minutes and a travel cost of 10 euros.

We highlight a few key points which will aid us in interpreting the VTT. First, observe that at the 60-minute level, the choice probabilities are exactly as one would expect, 0.5. For higher $VTT$ values the choice probabilities change more quickly with respect to the time attribute. This can be related to the steepness of the utility graphs on the right in Figure 2.1 One might expect a similar effect when we vary the cost attribute and fix the time attribute. The $VTT$ is, after all, the ratio of time and cost, right? However the graph in Figure 2.2 tells a different story. When varying the costs, the choice probabilities for different values of travel time are all the same and are plotted on top of one another. This is reflected in the utility graph as all utility curves have the same steepness. Thus the $VTT$ does not impact the relative sensitivity to the costs.

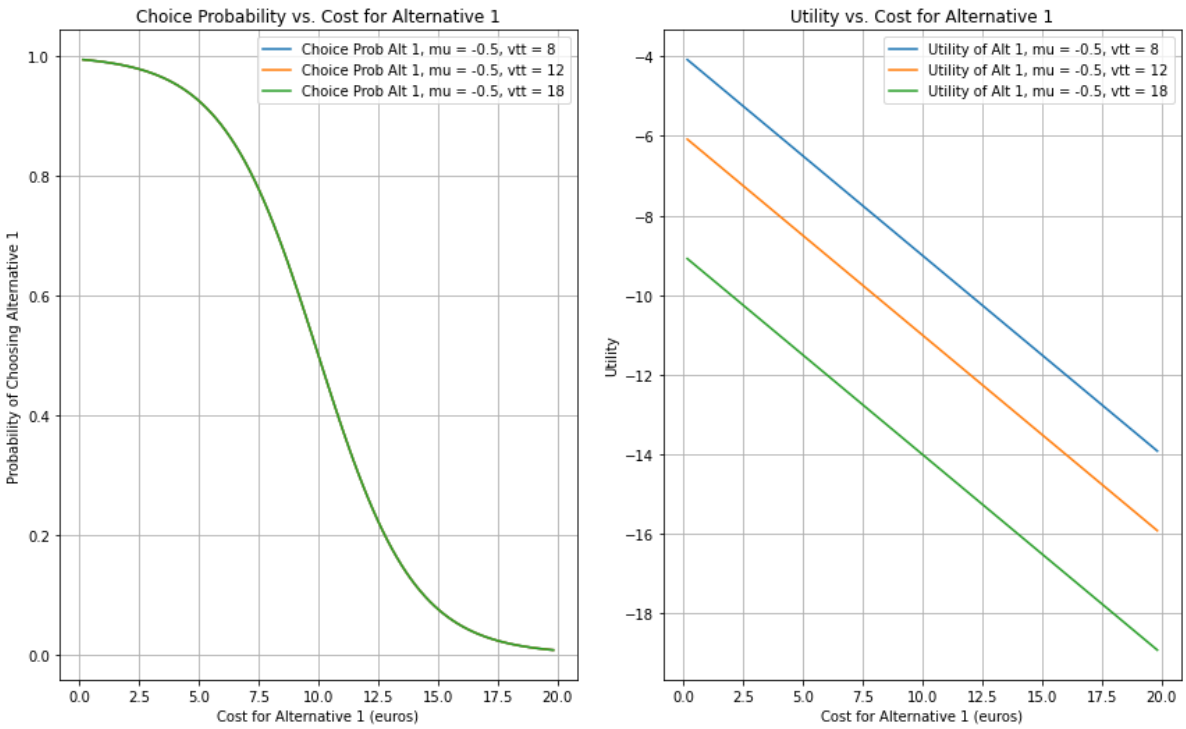



**Figure 2.2:** Choice probabilities and utilities for different $VTT$ values based on an MNL model. The other alternative is defined with a travel time of 60 minutes and a travel cost of 10 euros.

## 2.3. Travel Cost Reimbursement

This section covers what type of travel cost reimbursements exist in the Netherlands and the UK. The aim of this section is thus to establish what it means whenever a respondent states that they received a travel cost reimbursement. These types of reimbursement were identified based on a review of non-scientific sources. Table 2.3 is the result of this review. It is based on information retrieved from Bachaus (2023), Belastingdienst (2024), GOV.UK (2024), and HM Revenue Customs (2024).

**Table 2.3:** Types of Travel Cost Reimbursement in the Netherlands & UK

| Type of Reimbursement | Description | Modes | Country |
|---|---|---|---|
| Kilometer Allowance | Employers can provide an allowance based on the distance travelled | Car, Train, BTM | NL, UK |
| Public Transport Reimbursement | Employers can reimburse the costs of public transportation based on actual costs. | Train, BTM | NL,UK |
| Company Car | Employers can offer a company car. The costs are paid by the employer, and the employee can use this car for private purposes, with a tax addition for private use. | Car | NL,UK |
| NS-Business Card | Employers can offer an NS-Business Card for unlimited travel by public transport. | Train, bus, tram, metro | NL |
| Season Tickets | Reimbursement or subsidy for the cost of season tickets for public transport. | Train, Bus, Tram, Metro | UK |
| Fixed Travel Allowance | Employers may choose to pay a fixed monthly travel allowance, regardless of the actual travel costs. | Any mode of transport | NL, UK |

As companies are in essence free to determine their way of reimbursement employees for their travel expenses this list is not exhaustive. The most important difference between the UK and the Netherlands is not so much in the types of travel cost reimbursement offered by employers but in the tax benefits thereof.

In the Netherlands, employees who commute using their private vehicles can receive up to €0.23 per kilometre from their employers without incurring taxes, regardless of the vehicle type. Reimbursements exceeding this amount are considered taxable income, and employers must withhold payroll taxes on the excess. For those using public transportation, employers can either reimburse the actual travel expenses incurred or provide a reimbursement based on the number of kilometres travelled, again up to a maximum of €0.23 per kilometre (Belastingdienst, 2024).

Such a policy does not exist in the UK, thus there is no financial incentive from the employer's viewpoint to reimburse their employees for their travel expenses as it is taxed on the same grounds as regular salary (HM Revenue Customs, 2024).

### 2.3.1. Types of Reimbursement and their Tehavioural Implications

The different types of travel cost reimbursement listed in Table 2.3 can impact route choice behaviour in real life and perhaps in a choice experiment differently. When we further dissect the types of travel cost reimbursement stated in Table 2.3 we can identify two main categories; (1) types of reimbursement where an incentive to save on travel cost remains and (2) types of travel cost reimbursement where there is no incentive left to reduce one's travel expenses. We denote the first category as a partial Travel Cost Reimbursement (pTCR) and the second category as a full Travel Cost Reimbursement (fTCR) Let us try to categorize them one by one.

Firstly, a distance-based allowance, which is issued by employers in the UK and the Netherlands, usually entails that the employee reports their distance from work to their home and the actual travelled distances are multiplied by a factor which results in the total amount of travel expenses which is reimbursed. This form of reimbursement thus does not motivate to reduce one's total travel distance, however, there is still an incentive to reduce costs as this reimbursement scheme does not take the actual travel costs into account. In a choice experiment, this could mean that respondents are less sensitive to costs whenever they assume that their reimbursement applies, however as in their real-life choices costs are still a factor therefore we would not expect them to disregard costs altogether. Respondents who receive this type of travel cost reimbursement could be categorized as respondents who receive a partial Travel Cost Reimbursement.

Secondly, a Public Transport Reimbursement, not to be confused with the Dutch NS-Business Card, usually entails that the employee submits an overview of their travel expenses to their employer and they reimburse the costs which were made for commuting purposes. In the Netherlands this type of reimbursement is fully tax-deductible, in the UK it is not. With this type of reimbursement travel expenses do still show up in the bank statements of the employee, therefore arguably making them aware of the monetary value that is associated with commuting by public transport. However, an incentive to reduce one's travel expenses does arguably not exist with this type of reimbursement, as the full travel costs are reimbursed anyhow. This reimbursement scheme can be categorized as a full Travel Cost Reimbursement, but with the sidenote that respondents who receive this type of reimbursement are generally aware of their travel costs.

Thirdly, a company car is an interesting option from a financial perspective in the Netherlands but less so in the UK. In the Netherlands driving a company car for commuting purposes means that the company car is seen as an extra benefit, and is therefore taxed accordingly. A percentage of the value of the car is added to the taxable income of the employee and they thus incur taxes on that added value according to their marginal tax rate. A financially attractive option for a company is to pay for the fuel of both business and private trips of the employee as all the costs associated with the company car can be subtracted from the company's income statement, therefore reducing their taxable profit. With such a reimbursement scheme the employee incurs the costs of travel solely through a fixed amount in their yearly tax return. Furthermore, as the fuel costs and other operational costs are fully paid by the employer, incentives to reduce costs from the employee's perspective are non-existent. As the employee refuels the car they can see the amount of money which is paid by the employer, thus they could arguably have a feeling for the fuel costs incurred. Such a reimbursement scheme can be classified as a full Travel Cost Reimbursement as all operational costs incurred through travel are reimbursed. Interestingly, this type of reimbursement could also affect the sensitivity to travel costs incurred by trips without a business or commuting purpose as all operational costs are covered. With a reimbursement scheme as such we expect respondents to be significantly less sensitive to costs. Nonetheless, it is possible that whenever a respondent self-reports whether they receive a pTCR or an fTCR they still opt to choose the pTCR option as they are incurring costs due to travel through their yearly tax return.

A company car in the UK with a reimbursement of all fuel costs is financially less attractive in the UK. This is mainly because when a company in the UK provides fuel for private use in a company car, it is considered a taxable benefit for the employee, known as the fuel benefit charge. This charge is calculated based on a fuel benefit multiplier set by the government, which is then multiplied by the car's CO2 emission percentage, resulting in the taxable value of the fuel benefit provided for personal use. The employee must pay income tax on this benefit, and the employer must pay National Insurance contributions on the value of this benefit. Interestingly, from the employee's perspective, the way that they thus incur the operational costs of the car is through a fixed fuel benefit charge irrespective of the amount of fuel their car consumes in a year. This makes it an interesting option for high-mileage users or when using low-emission vehicles. Behaviourally this scheme arguably incentivizes you to drive faster and consume more fuel as such a driving style ensures that the costs you would otherwise incur are higher compared to the fixed amount of fuel benefit charge you are paying yearly. Such a reimbursement scheme can be classified as a full Travel Cost Reimbursement. For the employer, the costs associated with providing fuel for private use are not allowable business expenses and thus cannot be deducted from the business's taxable income. Instead, these costs are considered personal expenses of the employee, even though the employer covers them. Thus making it a less interesting option from the financial point of view of an employer.

As a Fourth option, used in the Netherlands, employers can offer an NS-Business card to their employees, rules regarding the administrative burden around employers issuing NS subscriptions and cards were recently lifted (https://nos.nl/l/2502239), thus employees are now able to also use their subscription for all sorts of private travel all tax-free. The legislation is in practice the same as before, only the administrative burden was reduced. From a behavioural perspective, similar to employees who receive a company car which can be used privately, there is no incentive left to reduce one's travel costs. This also applies to trips made neither for business or commuting purposes. This type of reimbursement can thus be classified as a full Travel Cost Reimbursement.

A Fifth option used predominantly in the UK, is a so-called season ticket, usually for a specific line or trip which the employee takes in their day-to-day commute. Again in the UK this type of reimbursement is not tax deductable, thus there is no incentive for employers to provide it. A season ticket is usually paid upfront, thus the costs of the season ticket are not directly observed by the employee. Behaviourally this would imply that the feel for the costs associated with travel would decrease, however, as this type of reimbursement is very specific the commuter is in real life not able to choose e.g. another mode of transport or even another route. We argue that for this type of reimbursement, there is no real incentive for the employer to reduce their travel costs, but no actual opportunity to do so. This type of reimbursement can thus be categorized as a full Travel Cost Reimbursement.

Lastly, a fixed travel allowance is a fixed amount of money irrespective of actual travel costs and can be based on numerous other factors which are determined by the company itself, this could include but is not limited to; a travel-time-based reimbursement, a reimbursement based on seniority within the company etc. As the tax-benefits in the Netherlands are based on a tax-free amount per kilometre we expect this kind of travel cost reimbursement to be used sparsly. Employers in the UK might opt for this kind of travel cost reimbursement as there is no financial incentive from the company's perspective to offer another type of reimbursement scheme. Behaviourally this means that there is still an incentive for employees to reduce their actual travel costs, as only a portion of their costs are reimbursed. Thus this type of reimbursement can also be typed as a pTCR.

To conclude this section, a distance-based allowance in the UK and the Netherlands reimburses employees based on reported distances, not actual travel costs, motivating them to save on travel. Similarly, a fixed travel allowance, a set amount regardless of actual costs, encourages employees to minimize travel expenses since the allowance might not fully cover their costs.

Conversely, full travel cost reimbursements arguably remove the incentive to save on travel costs. Public transport reimbursements, where commuting expenses are fully covered, leave no motivation to reduce costs. Company cars in the Netherlands and the UK, with fully paid fuel and operational costs, also fall into this category. In the Netherlands, the cost is incurred through a fixed tax amount, while in the UK, a fixed fuel benefit charge applies. The NS-Business Card in the Netherlands, used for private travel at no cost, and UK season tickets, paid upfront, further eliminate any incentive to reduce travel expenses, as all costs are covered.

What these different types of reimbursement imply in terms of choice behaviour in a stated preference experiment can be diverse. It all depends on the extent to which respondents base their choices on their reimbursed "normal" situation. Furthermore, it will depend on the extent to which respondents adhere to instructions to ignore or to include the fact that they receive a reimbursement.

## 2.4. The Scale Parameter

The scale parameter, which was denoted by the symbol $\mu$ in section 2.2, has a special role in a choice model because it is normally not separately identified, in other words, it is normally not a parameter itself. But before covering the mathematical identification let's observe what this scale parameter does in terms of choice probabilities. Let us observe Figure 2.3 to do so.

Contrary to what we observed for the VTT, the scale parameter interacts with the choice probabilities of both the cost and the time parameter. A smaller value for the scale parameter makes the choice probabilities less sensitive to changes in attribute levels. When looking at Figure 2.3 we can see what happens when the scale parameter approaches zero. The choice probabilities get close to 0.5 across the entire x-axis range. Therefore, the choices are almost entirely dominated by randomness.

### Preference space identification

This scale parameter $\mu$ is normally not identified in a linear in-parameters choice model. To understand why we can observe the utility function of such a specification again; here Equation 2.4 denotes the utility for alternative $j$ which is described by the attribute levels $\text{cost}_{ij}$ and $\text{time}_{ij}$.

$$U_{ij} = \underbrace{\beta_{\text{cost}} \cdot \text{cost}_{ij} + \beta_{\text{time}} * \cdot \text{time}_{ij}}_{\text{Observed Utility}} + \underbrace{\varepsilon_{ij}}_{\text{Noise}} \tag{2.4}$$

We are thus assuming that we can describe the utility of an alternative j by this utility function. We assume that the utility of alternative j is dependent on the travel time, the travel costs and random noise which is denoted by $\varepsilon_{ij}$. This random noise in an MNL model is modelled by independently drawing a random number from the EV Type 1 distribution with a mean of zero and a fixed variance of:

$$\text{Var}(\epsilon_{ij}) = \frac{\pi^2}{6} \qquad \forall i, j \in I, J \tag{2.5}$$

This distribution ensures that the choice probabilities can be analytically calculated and results in Equation 2.6. Which denotes the choice probability of choosing alternative $k$ out of the set of all alternatives $J$ becomes as per the standard MNL model:

$$P(k) = \frac{e^{\boldsymbol{\beta}^\top \mathbf{x}_k}}{\sum_{j=1}^{J} e^{\boldsymbol{\beta}^\top \mathbf{x}_j}} \tag{2.6}$$

Do note that we combined the parameters $\beta_{\text{cost}}$ and $\beta_{\text{time}}$ into one vector denoted by $\boldsymbol{\beta}$. Similarly the vector $\boldsymbol{x}_{ij}$ combines the $\text{cost}_{ij}$ and $\text{time}_{ij}$ attributes. Let us now try to scale the magnitude of the entire utility function by a real number denoted by $\mu$ and plug those in Equation 2.6, this gives:

$$U_{ij} = \underbrace{\mu(\beta_{\text{cost}} \cdot \text{cost}_{ij} + \beta_{\text{time}} * \cdot \text{time}_{ij})}_{\text{Observed Utility}} + \underbrace{\varepsilon_{ij}}_{\text{Noise}} \tag{2.7}$$

Which results in:

$$P(k) = \frac{e^{\mu \boldsymbol{\beta}^\top \mathbf{x}_k}}{\sum_{j=1}^{J} e^{\mu \boldsymbol{\beta}^\top \mathbf{x}_j}} \tag{2.8}$$

Here we can observe that for any given $\mu$, we can find a corresponding set of parameters $\boldsymbol{\beta}$ that will yield the same choice probabilities. This is because the probabilities depend only on the product $\mu\boldsymbol{\beta}$, not on $\mu$ and $\boldsymbol{\beta}$ individually. The key point here is that any change in the scale parameter $\mu$ can be absorbed by the parameters in the deterministic part of the utility. This implies that we cannot separately identify $\mu$ and the parameters of $V_{ij}$. To illustrate; If the vector $\boldsymbol{\beta}$ are the parameters in the deterministic utility function such that $V_{ij} = \beta^T x_{ij}$, then scaling the utility by $\mu$ gives:

$$\mu V_{ij} = \mu(\boldsymbol{\beta}^T x_{ij}) = (\mu\boldsymbol{\beta})^T x_{ij} \tag{2.9}$$

The product $\mu\boldsymbol{\beta}$ are what we estimate from the data, not $\mu$ and $\beta$ separately. Therefore what is usually done is to set $\mu$ equal to 1 to ensure that we can identify the vector of parameters $\boldsymbol{\beta}$.

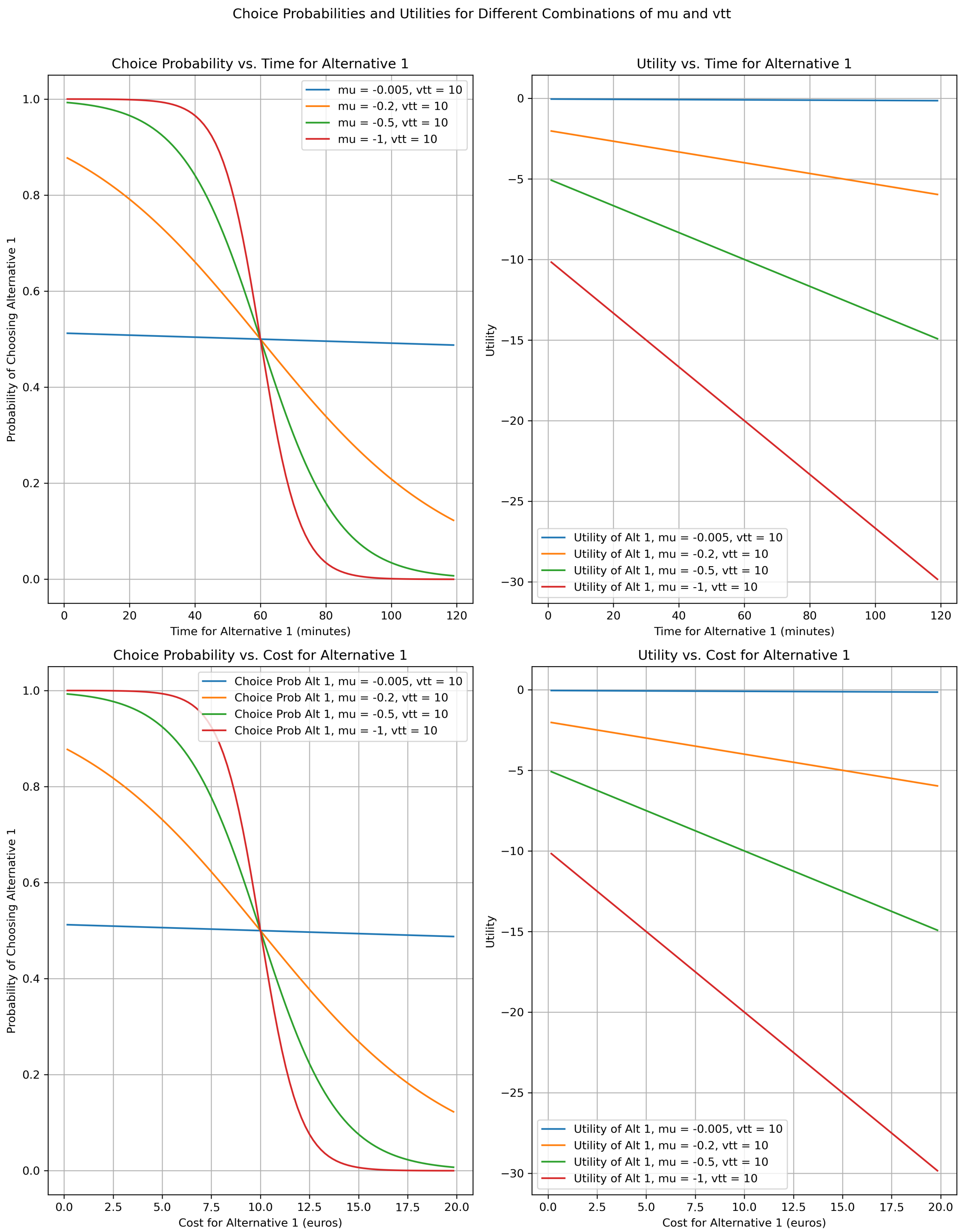



**Figure 2.3:** MNL choice probabilities and iso-utility curves with varying scale parameters

## WTP space identification

Contrary to a utility specification in preference space, a utility specification in WTP space does allow for the unique identification of the scale parameter $\mu$. Stated even more explicitly; a correct model in WTP space must include a separately identifiable scale parameter. Such a model is just a reparametrization of a linear in-parameters model, where the parameters are divided by the parameter which was previously used for the cost attribute.

To express the utility in terms of willingness-to-pay (WTP), we define the WTP for each attribute by dividing it with the cost parameter $\beta_c$, as by doing so we obtain the willingness-to-pay values. Thus for the time parameter this becomes; $\lambda_t = \frac{\beta_t}{\beta_c}$ By substituting $\beta_t = \lambda_t \beta_c$ into the utility specification as written in Equation 2.4 we can rewrite the deterministic part of the utility function as:

$$V_{ij} = \beta_c \cdot \text{cost}_{ij} + \lambda_t \cdot \beta_c \cdot \text{time}_{ij} \tag{2.10}$$

Rearranging results in;

$$V_{ij} = \beta_c \left(\text{cost}_{ij} + \lambda_t \cdot \text{time}_{ij}\right) \tag{2.11}$$

We can see that the cost parameter ($\beta_c$) now multiplies the entire utility function, this also changes its interpretation; it should no longer be seen as just the relative sensitivity to cost but as the relative sensitivity to all attributes in the utility function; in our case cost & time. To avoid ambiguous interpretation we opt to denote a model in WTP space as:

$$U_{ij} = \underbrace{\mu * \left(\text{cost}_{ij} + VTT \cdot \text{time}_{ij}\right)}_{\text{Observed Utility}} + \underbrace{\varepsilon_{ij}}_{\text{Noise}} \tag{2.12}$$

The key point here is that contrary to the model in preference space a change in the scale parameter can *no longer* be *fully* absorbed by altering other parameters in the deterministic part of the utility function. One other way to reason about this is to simply compare the number of parameters to the number of attributes. If we were not to specify a scale parameter in this WTP specification the number of parameters would be less than the number of attributes.

This scale parameter $\mu$ is inversely related to the error variance (Chen et al., 2016) as:

$$\text{Var}(\epsilon_{ij}) = \frac{\pi^2}{6\mu} \qquad \forall i, j \in I, J \tag{2.13}$$

In preference space specification the scale parameter is normalized to 1, which reduces Equation 2.13 to Equation 2.5.

### Choice Consistency and Error Heterogeneity

After observing the choice probabilities in Figure 2.3, one can imagine that the link with choice consistency is quickly established. This was done so by several papers. (Dellaert et al., 1999; Deshazo, 2002; Hensher, 2006; Jang et al., 2022; Swait & Adamowicz, 2001; Train & Weeks, 2015; Zhang & Adamowicz, 2011) all related a some form of choice consistency to the scale parameter. However, scale parameter variations are often directly linked with choice task complexity. Dellaert et al. (1999) is one of the few who formally links error heterogeneity to the notion of choice consistency. This is arguably a better definition as the actual observation in the choice data is a variation in choice consistency which is captured by a heterogeneous error term. In many papers, the link with choice task complexity is the main research objective and the link with error heterogeneity is made without mentioning choice consistency. We therefore adopt the definition of Dellaert et al. (1999), who states that:

> ”Choice consistency is measured as the variance of the random error component in the consumer utility function: the smaller this variance, the higher choice consistency.” (Dellaert et al., 1999)

As established earlier, per Equation 2.5, we can thus relate choice consistency to the magnitude of the scale parameter $\mu$. However, we must stress that we can only talk about choice consistency relative to the observed utility, as per Equation 2.12. Thus conceptually, choice consistency is in essence the degree to which respondents ”follow” the utility function defined by the researcher. In other words, how well the utility function can explain the observed choices.

### Modelling Scale Parameter Variations

There are several options available to model scale parameter variations. These scale parameter variations are sometimes referred to as, error heterogeneity, scale heterogeneity or heteroskedasticty (Dellaert et al., 1999). These terms are used interchangeably in the scientific literature.

Bradly and Daly (1996) were one of the first to provide a model for (discrete) error heterogeneity. It was employed to be able to account for error heterogeneity across datasets to be able to estimate a combined RP, and SP model. It involves a nesting structure where respondents are assigned to nests based on whether they were part of the RP or the SP dataset. They estimated a parameter which denoted the difference in scale across the two datasets.

Fiebig et al. (2010) further developed this to what they call a Generalized Mixed Multinomial Logit Model, to be able to account for both taste and scale heterogeneity. This approach enabled Fiebig et al. (2010) to model a continuous form of heteroskedasticity. He based his approach on the idea that the degree of randomness in choices has a certain distribution in the population.

The work by Fiebig et al. (2010) work was criticised by Hess and Train (2017a). They argue that the G-MNL model is not a generalisation of the Mixed Logit Model but instead a more restrictive version. This argument is based on the fact that mixed-logit when specified properly can accommodate a full covariance matrix between the randomly distributed attributes. In the standard G-MNL specification, this covariance matrix in essence collapses to a single parameter.

As we aim to model heteroskedasticity across, what we assume are homoskedastic subsets, the more discrete approach provided by Bradly and Daly (1996) provides the best framework.

### Relevance Scale Parameter for VTT

We end this section by discussing why heteroskedasticity is relevant for the main research objective; Investigating an effect of travel cost reimbursement on the value of travel time. We aim to do so through a brief literature overview and by providing an empirical example.

Swait and Adamowicz (2001) and Munizaga et al. (2000) found that not accounting for scale heterogeneity can sometimes bias WTP estimates. The reasoning is that if different error variances are not accounted for, the model might misattribute some of the variability in the data to the explanatory variables instead of the error term. This misattribution could therefore bias the model parameters, and therefore the VTT. Depending on the relative sizes of the error variances in the data, this could lead to either an overestimation or an underestimation.

This notion is best illustrated through a synthetic example we will be discussing now. We first define five synthetic respondent sub-samples of 2000 synthetic respondents each. We are thus creating synthetic individuals. The actual $VTT$ for all synthetic respondents is the same, defined at 10 €/h, meaning the true $VTT$ is identical across the entire sample. However, across the sub-samples, the degree of randomness in the choices of the respondents differs as presented in Table 2.4.

**Table 2.4:** Synthetic Subsamples with fixed $VTT$, $mu$ but variable error variances. Note that the fixed value for $\mu$ was arbitrarily chosen, it corresponded to a value which was once estimated on the actual NL2022 data

| Subsample | $\mu$ | VTT [€/h] | $\varepsilon_s$ | Change |
|---|---|---|---|---|
| Subsample 1 | -0.7681 | 10 | $\pi^2/60$ | 10% |
| Subsample 2 | -0.7681 | 10 | $\pi^2/12$ | 50% |
| Subsample 3 | -0.7681 | 10 | $\pi^2/6$ | 100% |
| Subsample 4 | -0.7681 | 10 | $\pi^2/4.5$ | 150% |
| Subsample 5 | -0.7681 | 10 | $\pi^2/3$ | 200% |

To create the questionnaire for the synthetic respondents, we first defined the number of respondents, the number of questions each respondent would answer, and the value of travel time (VTT) which was set at 10 euros. Each synthetic respondent was presented with eight hypothetical travel scenarios, each comprising a pair of alternatives characterized by varying costs and travel times. The left and right alternatives for each scenario were generated by randomly sampling cost values between 1 and 20 euros and travel times between 10 and 60 minutes.

The synthetic respondents make choices based on the utility function given in Equation 2.14 which incorporates both the cost and travel time of each alternative and the error components drawn from an Extreme Value Type 1 (Gumbel) distribution with variances according to Table 2.4. For each choice, the utility of both the left and right alternatives is calculated using the costs and times from the synthetic choice experiment, along with the error specific to the respondent's subset. The choice is then determined by comparing the utilities: the respondent selects the alternative with the higher utility value. These choices, and the subset to which a synthetic respondent belongs, are recorded for model estimation.

As the variance is inversely proportional to the scale parameter the correct specification would be that of Equation 2.16. However, we deliberately misspecify the model and assume respondents are making choices according to the utility specification in Equation 2.15. To reiterate; we ”program” respondents to make choices with varying error variances but we wrongly hypothesise that these respondents have different VTTs.

$$U_{ij} = \mu_s \cdot (\text{cost}_{ij} + 10 \cdot \text{time}_{ij}) + \varepsilon_s \qquad \text{Decision rule} \qquad (2.14)$$

$$U_{ij} = \mu \cdot \left(\text{cost}_{ij} + \left(VTT \cdot (1 + \sum_{s \in S} \delta_s \cdot D_s)\right) \cdot \text{time}_{ij}\right) + \varepsilon_{ij} \qquad \text{Misspecified} \qquad (2.15)$$

$$U_{ij} = \mu \cdot \left(1 + \sum_{s \in S} \mu_s \cdot D_s\right) \cdot (\text{cost}_{ij} + VTT \cdot \text{time}_{ij}) + \varepsilon_{ij} \qquad \text{Correctly specified} \qquad (2.16)$$

Where $D_s$ is a dummy variable to indicate whether a respondent belongs to subsample $s \in S$.

Now that we obtained the synthetic choices we can estimate a multinomial logit model (McFadden, 1973). We thus assume for a moment that the $VTT$ across sub-samples is different. Therefore we include an interaction with the VTT, as in Equation 2.15. For clarification, we deliberately misspecify our model to illustrate the relevance of error heterogeneity across sub-samples. By including the interaction terms $\sum_{s \in S} \delta_s \cdot S_s$, we allow the $VTT$ to vary across the different sub-samples. But, we only estimate a single scale parameter for all sub-samples. The results of this empirical example are given in Table 2.5. Do note that we also included the results with the theoretically correct utility specification of Equation 2.16.

**Table 2.5:** MNL estimation results based on synthetic choice data.

| | | Misspecified Model | | Correct model | |
|---|---|---|---|---|---|
| Number of individuals | | 10000 | | 10000 | |
| Number of modelled outcomes | | 80000 | | 80000 | |
| Estimated parameters | | 6 | | 6 | |
| LL(final) | | -18190.9 | | -16946.38 | |
| Adj.Rho-square (0) | | 0.6718 | | 0.6943 | |
| **Parameters** | **Unit** | **Estimate** | **t-ratio(0)** | **Estimate** | **t-ratio(0)** |
| $\mu$ | - | -0.6076 | -124.87 | -0.592 | -56.87 |
| $\mu_1$ | - | 0 | NA | 2.1228 | 20.49 |
| $\mu_2$ | - | 0 | NA | 0.3843 | 10.5 |
| $\mu_3$ | - | 0 | NA | 0 | NA |
| $\mu_4$ | - | 0 | NA | -0.1628 | -8.01 |
| $\mu_5$ | - | 0 | NA | -0.2808 | -16.33 |
| $VTT$ | €$/h$ | 10.1745 | 66.49 | 9.9572 | 171.38 |
| $\delta_1$ | - | 0.0359 | 1.66 | 0 | NA |
| $\delta_2$ | - | -0.0041 | -0.19 | 0 | NA |
| $\delta_3$ | - | 0 | NA | 0 | NA |
| $\delta_4$ | - | -0.0725 | -3.53 | 0 | NA |
| $\delta_5$ | - | -0.0913 | -4.51 | 0 | NA |

One must interpret the $\delta_s$ and the $\mu_s$ parameters as a fraction. For example, the value of $\delta_5$ is -0.0913, indicating that respondents who belong to Sub-sample 5 have a $VTT$ which is 9% less as compared to the reference VTT, which in this case is the $VTT$ of Subsample 3. Do note that one cannot make statistical inferences between groups based on the magnitude of this $\delta_s$ parameter, as these results do not provide a statistical test for the difference between e.g. Subsample 4 and Subsample 5. What we *can* infer, is that the $VTT$ for Subsample 4 and 5 are statistically different from the reference $VTT$, which in this case specified to be that of Subsample 3.

This empirical example shows that it might indeed be possible that, an interaction on the $VTT$ which coincides with a sub-sample which has a different error term, can misattribute some of the scale heterogeneity to, in this case, the $VTT$.

## 2.5. Sign Effects

As stated before, a sign effect is a discontinuity in the respondent's preferences. In the context of $VTT$, this indicates a discontinuity at $\Delta t = 0$ and $\Delta C = 0$. This sign effect is well documented and appears in both scientific research (Hess & Train, 2017a) and applied studies (Kouwenhoven et al., 2022).

The $VTT$ thus becomes dependent on the sign of the travel time presented minus the reference travel time as in Equation 2.17.

$$VTT = \begin{cases} WTP & \text{if } \Delta t > 0 \\ WTA & \text{if } \Delta t < 0 \end{cases} \tag{2.17}$$

Where $\Delta t$ is the difference between the travel time presented and the reference travel time of the respondent, WTA is the willingness to accept a deterioration in travel time, and WTP is the willingness to pay for an improvement in travel time. In most research, one finds that the WTA is higher than the WTP (Kaa, 2010; Kouwenhoven et al., 2022). For respondents participating in the survey via an internet panel, the reference trip refers to a self-reported journey they have recently undertaken. If a respondent is intercepted during a journey, such as on a train, the reference trip is the one they are currently making. In the survey, respondents are asked to describe this reference trip, including the travel time and costs. When we refer to a respondent's reference time and costs, we are specifically referring to the time and cost associated with this reference trip.

#### Behavioural Notion

The fact that a WTA is higher than a WTP is often explained (Kaa, 2010) through the behavioural phenomena known as loss aversion (Tversky & Kahneman, 1974, 1991) and through the endowment effect (Kahneman et al., 1990). Loss aversion criticized the then-common microeconomic theory that people just aim to maximize utility. It is now demonstrated across a wide variety of experiments. One good example is that of Tversky and Kahneman (1991) themselves. One of their experiments involved giving participants a mug and then offering them the opportunity to trade it for an equally valued item, such as a pen. Despite the items being of equal value, participants were much less willing to trade away the mug once they owned it, indicating that the loss of the mug loomed larger than the potential gain of the pen. This behaviour illustrates that people tend to overvalue what they already possess compared to items they do not own. In the context of $VTT$ studies, Kaa (2010) among others, concluded that the sign effect follows from the notion of loss aversion. In other words; the notion that respondents are typically willing to pay more to prevent a deterioration in travel time than they are to improve it, can be explained through loss aversion.

#### Relevance for VTT

If the chosen model does take sign effects into account, disentangling for which groups this sign effect is more dominant can change the estimated $VTT$. Whenever the model applied to estimate a $VTT$ does not take sign effects into account; investigating this sign effect difference predominantly sheds light on why the $VTT_{nTCR}$ and the $VTT_{fTCR}$ could be different. If for example, a $VTT$ difference between TCR and nTCR respondents originates solely from the WTA and the WTP will be equal for TCR and nTCR respondents, then arguably the WTP will more closely resemble what respondents are willing to pay out of their own pocket. Lastly, it has been shown by Hess et al. (2017) that taking sign effects into account often better explains the observed choices as compared to models that do not take sign effects into account.

# 3
# Hypotheses

The goal of this chapter is to formalize the hypothesis stated in the Introduction and to provide the argumentation on why they are postulated in the first place. Do note that this chapter relies heavily on the theory which was discussed chapter 2. Thus whenever one misses background information one can go back to the Theory. Lastly, the method to test the hypotheses is not discussed in this chapter but left to chapter 4. Thus coding of variables, sample information, data suitability and model development are all discussed there.

## Hypothesis 1, The Value of Travel Time

Our hypothesis stated in chapter 1 stated that respondents with a TCR might have a higher $VTT$ compared to respondents not receiving a travel cost reimbursement. In the context of a stated choice experiment, we can formalize this hypothesis and its corresponding null hypothesis as Equation 3.1

$$H_0 : \quad \frac{\beta_{\text{time,reimb}}}{\beta_{\text{cost,reimb}}} = \frac{\beta_{\text{time,non-reimb}}}{\beta_{\text{cost,non-reimb}}} \tag{3.1}$$

$$H_1 : \quad \frac{\beta_{\text{time,reimb}}}{\beta_{\text{cost,reimb}}} > \frac{\beta_{\text{time,non-reimb}}}{\beta_{\text{cost,non-reimb}}} \tag{3.2}$$

The same hypothesis but now presented in willingness to pay space is given in equations Equation 3.3 and Equation 3.4.

$$H_0 : \quad \text{VTT}_{\text{reimb}} = \text{VTT}_{\text{non-reimb}} \tag{3.3}$$

$$H_1 : \quad \text{VTT}_{\text{reimb}} > \text{VTT}_{\text{non-reimb}} \tag{3.4}$$

To understand this hypothesis we provide a line of reasoning in the next section on *why* the $VTT$ could be different for respondents who receive a travel cost reimbursement.

### Behavioural Foundation

Several behavioural arguments and one microeconomic argument can be made to support Equation 3.1. Firstly, from a behavioural viewpoint, lack of payment consequentiality (Zawojska et al., 2019) for respondents with a TCR could increase the $VTT$. In the context of discrete choice experiments, *payment consequentiality* ensures that respondents treat the hypothetical choices as if they were real decisions involving actual financial costs or benefits. Respondents accustomed to reimbursements may struggle to internalize the hypothetical scenario where they bear the travel costs themselves. This discrepancy is primarily because their real-world decisions involve no direct financial consequences, leading to less careful consideration of cost implications in a hypothetical context such as a discrete choice experiment. This behavioural mechanism will be dependent on whether a respondent first pays for the travel costs themselves and is later reimbursed or if they are never confronted with the cost of travelling through their reimbursement as was elaborated on in subsection 2.3.1.

Secondly, also from a behavioural viewpoint, budget constraints (Ding et al., 2005) could also cause an increase in the $VTT$ for respondents receiving a TCR. The argument is quite simple, whenever a respondent does not receive a travel cost reimbursement, a change in the cost of their day-to-day commute could significantly impact their monthly finances. Therefore, their actual real-life budget for travel expenses could constrain how much they are willing to spend on travel costs. Respondents who do receive a fTCR or a pTCR do not have to consider these budget constraints.

Thirdly, Thaler's concept of mental accounting (Thaler, 1999) explains how individuals separate their finances into different accounts for various purposes. The "mental account" out of which travel expenses are paid is quite small for respondents who receive a fTCR or a pTCR. Whereas for respondents not receiving a TCR the "mental account" out of which travel expenses are paid naturally has to be larger. Therefore, the psychological cost associated with more expensive choices could be lower for respondents receiving a TCR and consequentially increase their $VTT$ compared to their non-reimbursed counterparts.

Lastly, from a microeconomic perspective, the actual disutility of costs in real-life decisions is lower whenever travel costs are reimbursed. Therefore time is relatively more important than costs which consequentially increases the $VTT$ (McFadden, 1973).

These behavioural phenomena are not new to practitioners, therefore what is often done in practice is to instruct respondents to, either take or not take their TCR into account. The purpose of such an instruction is to reduce the likelihood that respondents make assumptions about the choice situation. In our case, respondents in the NL2022 survey were instructed to; make decisions as if they were travelling in their own time and that they have to pay for the costs themselves. Respondents in the UK2014 survey were instructed to; take into account who normally pays for their journey when making choices. However, as is also widely established, respondents do not read instructions meticulously (Brosnan et al., 2019). Thus the behavioural phenomena as described earlier can still play a role in the decision-making process. The fact that respondents in the UK2014 and the NL2022 surveys got different instructions might impact the effect of TCR on the VTT. But as there are more differences between the two datasets we cannot with certainty attribute differences to the instruction type.

In summary, Hypothesis 1 posits that TCR increases the $VTT$ due to a combination of behavioural mechanisms and economic considerations. Despite instructions in the experiments aiming to neutralize these effects, the inherent workings of human behaviour may still lead to a higher VTT among reimbursed respondents.

## Hypothesis 2, Scale Heterogeneity

Our second hypothesis suggests that respondents with a TCR make less consistent choices compared to those without. This hypothesis explores the notion of scale heterogeneity, which refers to variations in the consistency of choices across different respondent groups. We can formalize this hypothesis and its corresponding null hypothesis as follows:

$$H_0: \quad \mu_{\text{reimb}} = \mu_{\text{non-reimb}} \tag{3.5}$$

$$H_1: \quad \mu_{\text{reimb}} > \mu_{\text{non-reimb}} \tag{3.6}$$

Scale heterogeneity is crucial as was stipulated in chapter 2 because heterogeneity in the VTT can be confounded with scale heterogeneity in some cases. The scale parameter ($\mu$) reflects the consistency of choices made by respondents or similarly the sensitivity of respondents to changing attribute levels, with a higher $\mu$ indicating more consistent choices or a higher sensitivity to attribute level changes.

### Behavioural Foundation

Several behavioral arguments suggest that TCR could lead to less consistent choices:

Firstly, respondents without reimbursement might be prone to start calculating what an alternative means for their monthly expenses. Respondents with a travel cost reimbursement might not feel the need to make such calculations, again related to lack of payment consequentiality (Zawojska et al., 2019). Therefore the respondents without a travel cost reimbursement could potentially put in more cognitive effort when making choices because they are making actual calculations or inferences about what an alternative would mean for their actual expenses. Alter et al. (2007) has shown, in a series of now-famous experiments, that cognitive effort could activate analytic reasoning. The idea by Alter et al. (2007) was such, that when faced with cognitive strain, individuals are more likely to reject intuitive answers and engage more thoroughly with the task to solve problems accurately. Kahneman (2011) described this as; cognitive strain activating analytic reasoning and cognitive effort. An increase in cognitive effort in its turn increases choice consistency.

Secondly, as indicated by Swait and Adamowicz (2001), the significance of a product class — measured by factors like the proportion of an individual's income— is related to the amount of cognitive effort people put into a decision. The significance of costs in proportion to an individual's income is naturally lower whenever an individual receives a TCR. Consequentially, respondents with a TCR could be less motivated to exert a substantial amount of cognitive effort when making choices, increasing randomness and therefore decreasing the magnitude of the scale parameter.

These behavioural theories collectively suggest that TCR could reduce the consistency of choices, supporting our hypothesis.

## Hypothesis 3, Sign Effects

Our third hypothesis stated that a difference in the $VTT$ might be explained through an asymmetry in the preferences of respondents with a TCR. Specifically in terms of their willingness to accept (WTA) a deterioration in travel time and their willingness to pay (WTP) for improved travel time. We can mathematically formalize the hypothesis per Equation 3.7.

$$\frac{(WTA_{TCR} - WTA_{nTCR})}{WTA_{nTCR}} > \frac{(WTP_{TCR} - WTP_{nTCR})}{WTP_{nTCR}} \tag{3.7}$$

Where

$WTA_{TCR}$ is the willingness to accept a deterioration in travel time for respondents receiving a travel cost reimbursement,

$WTA_{nTCR}$ is the willingness to accept a deterioration in travel time for respondents not receiving a travel cost reimbursement,

$WTP_{TCR}$ is the willingness to pay for an improvement in travel time for respondents receiving a travel cost reimbursement,

$WTP_{nTCR}$ is the willingness to pay for an improvement in travel time for respondents not receiving a travel cost reimbursement.

### Behavioural Foundation

The reasoning is simple; a respondent who receives a travel cost reimbursement might very well assume that their costs are reimbursed up to their normal travel costs. Therefore, costs which are below the respondents' reference might be of less importance to them, causing an increase in the $VTT$. However, this increase in the $VTT$ will thus not be constant across all cost levels since cost levels above the respondent reference might normally not be reimbursed for that respondent. As their willingness to accept, $WTA_{TCR}$, is determined based on choices with travel costs below their *reference* travel costs, the $WTA_{TCR}$ could be higher as compared to respondents who do not receive a TCR. Along the same line of reasoning, we can argue that, under the assumption that respondents receive a TCR up to their reference level, the $WTP_{TCR}$ should not be much different as compared to the willingness to pay for respondents not receive a travel cost reimbursement $WTP_{nTCR}$.

The second reason which motivates the hypothesis in Equation 3.7, is that there are types of travel cost reimbursement which are limited to a fixed amount. Applying microeconomic reasoning, the disutility of the travel costs is, therefore, lower below this fixed amount and higher above this fixed amount. Thus creating a discontinuity in the $VTT$ at that fixed amount. Regardless of the limit of the travel cost reimbursement, this will always result in a larger difference between the $WTA_{TCR}$ and the $WTP_{TCR}$ as compared to respondents who do not receive a TCR.

# 4 Method

This chapter is devoted to the method applied to be able to answer the research question and the sub-research questions. We first present the details of the datasets and methodology used in the research. It begins with a description of the dataset, including the choice tasks, reimbursement information retrieval, and filtering criteria for respondent instructions. We present a summary table, a table of reimbursement categories, and a reimbursement percentage distribution graph. Next, we assess data suitability and discuss limitations. The operationalization section covers coding of travel cost reimbursement variables, defining reimbursement bins, and presenting distribution tables and figures. The models applied include Multinomial Logit (MNL), MNL with heteroskedasticity, and Mixed Logit (ML). We describe the model development process, from initial analysis to evaluating reimbursement effects and adding interactions. We end this chapter by describing how we applied the developed model to assess policy implications.

## 4.1. Coding and Definition of TCR

We now define the $D_{\text{TCR}}$ variables as denoted in Equation 4.1 and Equation 4.2. To do so we discuss how TCR information is measured, how respondents were instructed to deal with an eventual TCR and the corresponding model parameters.

### Measuring TCR

Respondents self-reported whether they received a TCR. The options which were presented to the respondents are presented in Table 4.1.

In the NL2022 dataset, respondents were first instructed to think of a journey they recently made and to state e.g. with what type of mode and for what purpose they made this journey. Thereafter, respondents were told that the entire questionnaire would be about this journey they just described. Respondents were eventually asked to state if they received some form of reimbursement for this reference journey. Respondents could choose amongst the options presented in Table 4.1.

The UK2014 approach is, apart from specific wording, the same. Travel cost reimbursement information was retrieved similarly, by asking respondents to self-report the travel cost reimbursement of their reference trip, subsequently, it was stated that the questions in the survey would be about that reference trip. The descriptions are matched to those in the NL2022 dataset in Table 4.1.

**Table 4.1:** Options for travel expense reimbursement

| Option | NL2022 | UK2014 |
|---|---|---|
| 1 | Yes, in full | Employer/company paid |
| 2 | Yes, partially | - |
| 3 | Yes, travel costs were shared with fellow passengers | Shared with other travellers |
| 4 | No, no reimbursement | I paid all costs |
| 5 | There were no travel expenses for me | All paid by other travellers, Colleague /partner/family member paid |
| 6 | Other, namely: ___ | Other (please specify) |

In the NL2022 dataset, when respondents reported receiving a partial travel cost reimbursement a subsequent question was asked; "What percentage of the travel costs was reimbursed?" Which again is thus a self-reported metric. However one can argue that a respondent with a reimbursement of e.g. 10% has in fact paid most of the trip out of their own pocket. We therefore opted to redistribute the categorical reimbursement statuses. Respondents with a partial reimbursement were redistributed according to Table 4.3. Thus respondents who reported receiving a partial travel cost reimbursement of e.g. $9$% were added to the *No reimbursement category*.

### Instructions

The meaning of the TCR variables differs slightly between the NL2022 and the UK2014 datasets as the respondents got different instructions on how to deal with their TCR.

For the UK2014 dataset, the respondents were instructed to:

> "Please imagine that each situation is exactly the same as for the actual #MODE# journey at the time you made the journey, except for the travel costs and travel time. Choose the one you prefer, keeping in mind all of the circumstances of the #MODE# journey at the time you made the journey and assuming that the only things that vary are the costs and times shown. For each of the following pairs of options, carefully compare the two options, bearing in mind who paid for the actual journey and who else is travelling with you."(ARUP, 2015)

For the NL2022 dataset respondents were instructed to:

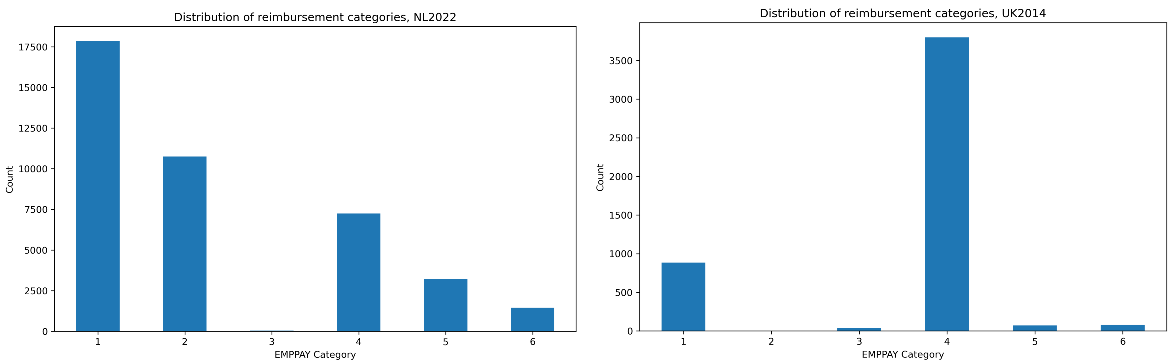
> "Imagine that you can choose between two different routes. The travel time can vary, for example, because one trip is shorter or less congested, and not because you accelerate more quickly. The costs can also vary, for example, because one trip requires less fuel.

> Each time, we ask which trip you prefer. For each choice, assume that both trips are feasible (even if they seem unrealistic), you travel in your own time, all costs are for your account, and all other characteristics are equal for both trips (equally safe, equally pretty, parking costs are equal, etc.). Additionally, all other circumstances, such as the weather and your appointments and activities that day, are the same as during the journey you have described."(Kouwenhoven et al., 2022)

Note that the respondents of the UK2014 dataset were instructed **not** to ignore their travel cost reimbursement. We argue that this is an advantage because this difference might explain differences in the TCR interactions.

### Filtering

In the NL2022 dataset, only respondents with reimbursement options 1,2 and 4, in Table 4.3, were selected for the analysis. This was done since for these options the meaning of the type of reimbursement is unambiguous and the number of observations was satisfactory, as can be observed in Figure 4.1a In the UK2014 dataset (Figure 4.1b) only respondents with reimbursement options 1 and 4 were selected for the analysis as the other options were rarely selected by respondents.




**(a)** Distribution of reimbursement categories for the NL2022 dataset. **(b)** Distribution of reimbursement categories for the UK2014 dataset.

**Figure 4.1:** Comparison of distribution of reimbursement categories for NL2022 and UK2014 datasets. The categories correspond to those in Table 4.1

Furthermore, respondents were also filtered based on several rules. For example, whenever the self-reported travel cost was unrealistic, respondents were removed from the dataset. It is outside the scope of this study to cover all these filtering procedures. Instead, we provide Table 4.2 with an overview of the final selection of observations. Thus denoting the number of completed choice tasks, not the number of respondents.

**Table 4.2:** NL2022 and UK2014, Observations distributed per mode

| **Mode** | **Commute** (1) | **Business** (2) | **Total** |
|---|---|---|---|
| Car (1) | 11680 | 5008 | 16688 |
| Train (2) | 7832 | 2456 | 10288 |
| BTM (3) | 6848 | 2024 | 8872 |
| **Total (NL2022)** | 26360 | 9488 | 35848 |
| Car (1) | 4835 | 8650 | 13485 |
| Train (2) | 4585 | 8790 | 13375 |
| LRT (3) | 2900 | 2300 | 5200 |
| Bus (UK2014, 4) | 1675 | 200 | 1875 |
| **Total (UK2014)** | 13995 | 19940 | 33935 |

### Coding

To include the TCR in the model specification, dummy coding was applied. The reason for applying dummy coding instead of effects coding has to do with the way the parameters can be interpreted. In effects coding the parameters should be interpreted relative to the sample average $VTT$, when dummy coding we interpret the parameters as absolute values for every subgroup. The absolute values for these subgroups are important in the calculation of national averages. Therefore dummy coding is preferred over effects coding. The category which was fixed was determined for every estimation separately, as fixing the category with the highest number of observations yields the highest t-ratios.

**Table 4.3:** Coding for reimbursement levels, $D_{\text{nTCR}}$, $D_{\text{pTCR}}$ and $D_{\text{fTCR}}$

| **Variable** | | **Coding** | | |
|---|---|---|---|---|
| **Reimbursement** | | $D_{\text{nTCR}}$ | $D_{\text{pTCR}}$ | $D_{\text{fTCR}}$ |
| 0 to 15% | No | 1 | 0 | 0 |
| 15 to 85% | Partial | 0 | 1 | 0 |
| 85 to 100% | Full | 0 | 0 | 1 |
| **Quadrants** | | $D_{\text{WTP}}$ | $D_{\text{WTA}}$ | |
| WTP quadrant | | 1 | 0 | |
| WTA quadrant | | 0 | 1 | |

with this definition of the $D_{TCR}$ variables, the corresponding parameters have to be interpreted accordingly. To illustrate; a $\delta_{fTCR}$ value of $0.09$ and a t-ratio of $2.05$, would indicate a 9% increase in the $VTT$ compared to the category which is fixed, ($0.09 \cdot 100\% = 9\%$). The t-ratio of $2.05$ would indicate that we can reject the null hypothesis that the $VTT$ is equal between groups with 95% certainty.

### The Sample

Both national stated preference surveys subdivided their experiments into sub-experiments. Not all of these sub-experiments are relevant to this study. For example, respondents who travelled by boat have by definition the travel purpose other, as this is typically a leisure activity. Only sub-experiments which were used to test the hypothesis are presented in Table 4.4.

The relevant sub-experiments were all presented as unlabeled route-choice experiments, where respondents had to choose between two hypothetical alternatives, each described by the attributes as in Table 4.4. Every respondent had to make 9 of these route choice decisions. One of these choice situations contained an alternative which dominated the other alternative. This was to ensure respondents were not randomly making decisions or e.g. always choosing the left option. During model estimation, this check question was not considered as it does not provide any information on the trade-offs respondents had to make.

**Table 4.4:** Overview of datasets, attributes, modes and hypotheses

| Dataset | Name | Attributes | Modes | Hypotheses |
|---|---|---|---|---|
| NL2022 | SP1A | time, cost | Car, Train, BTM | $H_1$, $H_2$, $H_3$ |
| | SP2A | time, cost, reliability | Car, Train, BTM | $H_1$ |
| | SP3A | time, cost, access/egress time, transfer time, number of transfers | Train, BTM | $H_1$ |
| | SP4A | time, cost, crowding, ability to sit, frequency | Train, BTM | $H_1$ |
| UK2014 | SP1 | time, cost | Car, Train, LRT, Bus | $H_1$, $H_2$ |
| | SP2 | time, cost, reliability | Car, Train, LRT, Bus | $H_1$ |

#### Data Suitability

We now discuss to what extent the datasets are suitable to answer the research question and suitable to test the hypotheses. We do so by discussing a-priori limitations.

Firstly, the datasets from the 2022 Dutch VTT study and the 2014 UK VTT study are from different periods. Economic conditions, travel behaviours, and policies may have changed between these periods, potentially affecting the comparability of the datasets. However, as we are investigating a relative difference as compared to the VTT, absolute changes in the VTT can safely be ignored.

Secondly, the 2022 Dutch VTT study and the 2014 UK VTT study did not record the same socioeconomic data. Therefore, not all socioeconomic interactions were applied consistently. However, the most important socioeconomic interactions, interactions which could have explained away a potential reimbursement effect, are included across both datasets

Thirdly, as was mentioned earlier, the UK2014 respondents were not able to specify whether they received a *partial* travel cost reimbursement. Therefore it might have been possible that respondents with a partial travel cost reimbursement chose either the option; ”Employer/company paid” or ”I paid all costs.” However, we expect that whenever a respondent receives some form of travel cost reimbursement they are more likely to opt for the ”Employer/company paid” option. Thus we further refer to the UK2014 respondents as fTCR respondents.

## 4.2. Model Development

**H1** As we hypothesised that respondents with TCR have a higher VTT we added an interaction with the $VTT$ and TCR, which is denoted by $\delta_{TCR}$. The definition and coding of the TCR variable were discussed in section 4.1. Furthermore, we added the $\delta_{TCR}$ parameters as relative interactions. The magnitude of the $\delta_{TCR}$ parameter must thus be interpreted as a relative change to the reference $VTT$. This specification allows for easy interpretation and quick comparison with different magnitudes for the $VTT$ itself. The interaction as described above is shown below in Equation 4.1

$$V_{ij} = \mu \cdot \{C_{ij} + [VTT \cdot T_{ij} \cdot (1 + \delta_{\text{TCR}} \cdot D_{\text{TCR}})]\} \quad \forall i \in I, j \in J \tag{4.1}$$

Where parameters to be estimated are denoted in blue, and $C_{ij}$ and $T_{ij}$ are respectively the cost and time for alternative $j$ for respondent $i$. Furthermore, $D_{\text{TCR}}$ denotes the variable which corresponds with the TCR. Further definition of this variable and the subsequent interpretation of the $\delta_{\text{TCR}}$ parameter is discussed in section 4.1. We opted to use a significance level of 95% throughout this study, consequentially whenever the t-ratio $\geq 1.96$ of the $\delta_{\text{TCR}}$ parameter [1] we can reject the Null hypothesis that the $VTT$ is not dependent on TCR, and therefore supporting the alternative hypothesis that the $VTT$ is dependent on TCR. This hypothesis can naturally only be tested for experiments where a $VTT$ can be estimated, thus experiments which include a cost and a time attribute.

**H2** We expected respondents with a travel cost reimbursement to make more random choices which can be subsequently captured in the scale of the utility function as was discussed in chapter 2. Hence, we incorporate TCR interactions with the scale of the utility similarly to (Chen et al., 2016; Dellaert et al., 1999), denoted in Equation 4.2

$$V_{ij} = \mu \cdot (1 + \mu_{TCR} \cdot D_{\text{TCR}}) \{C_{ij} + [VTT \cdot T_{ij} \cdot (1 + \delta_{\text{TCR}} \cdot D_{\text{TCR}})]\} \quad \forall i \in I, j \in J \tag{4.2}$$

Where again, the parameters to be estimated are denoted in blue, As preliminary estimates indicated significant interactions between the $VTT$ and TCR we opted to not solely add the $\mu_{\text{TCR}}$ but instead also add the $\delta_{\text{TCR}}$ interactions as previously denoted in Equation 4.1. Following the warnings given by Hess and Rose (2012) and Hess and Train (2017b), and per our synthetic example in Table 2.5, simultaneously allowing for $VTT$ interaction circumvents the scale parameter from capturing differences in the $VTT$. With the specification as in Equation 3.5, this scale parameter interaction with TCR was evaluated on experiments with two attributes, where only the $VTT$ is allowed to be a random variable. This again circumvents capturing other sources of correlation in the scale parameter (Hess & Rose, 2012; Hess & Train, 2017b) This approach is very similar to the one applied in Chen et al. (2016), where they aimed to explain scale heterogeneity through respondent-experienced time pressure. To explicate; the scale parameter $\mu$ is thus *not* specified as a random variable, again to circumvent capturing other sources of correlation.

This approach enabled us to evaluate the first *and* the second hypothesis with this specification since we allow for differences in the $VTT$ (Hypothesis 1) and differences in scale (Hypothesis 2). By comparing the estimates resulting from both specifications we can evaluate to what extent allowing for scale heterogeneity changes the $VTT$ estimates.

**H3** Lastly, we hypothesised that a travel cost reimbursement effect on the $VTT$ would be more profound in the willingness-to-accept (WTA) compared to the willingness-to-pay (WTP). To test this hypothesis we could only make use of quadrant type questions. These are questions where one of the alternatives is exactly the same as the reference journey described by the respondent, and the other alternative is either an increase in time but a decrease in costs or vice-versa. Consequentially, the only sub-experiment which had these types of questions was the SP1A sub-experiment of the NL2022 dataset. These types of questions enabled estimating two separate values for the $VTT$, denoted by the WTA and the WTP.(Kouwenhoven et al., 2022) To evaluate our third hypothesis we added a similar interaction as denoted in Equation 4.1 but now separated on the WTA and the WTP parameters. This approach yielded the specification in Equation 4.3

[1] Naturally under the condition that the degrees of freedom are sufficient.

$$V_{ij} = \mu \cdot \Big\{ C_{ij} + \Big[ WTP \cdot T_{ij} \cdot D_{\text{WTP}} \cdot (1 + \delta_{\text{TCR,WTP}} \cdot D_{\text{TCR}}) \Big] + \Big[ WTA \cdot T_{ij} \cdot D_{\text{WTA}} \cdot (1 + \delta_{\text{TCR,WTA}} \cdot D_{\text{TCR}}) \Big] \Big\} \quad \forall i \in I, j \in J \tag{4.3}$$

Where; $D_{\text{TCR,WTA}}$, $D_{\text{TCR,WTP}}$ are dummy variables to denote whether a choice set belongs to either the WTA or the WTP quadrant and $\delta_{\text{TCR,WTA}}$, $\delta_{\text{TCR,WTP}}$ are the corresponding parameters. These parameters should again be interpreted as a relative change to the reference WTA and WTP. Testing the third hypothesis was done by comparing the magnitudes of the $\delta_{\text{TCR,WTA}}$, $\delta_{\text{TCR,WTP}}$ parameters. A higher $\delta_{\text{TCR,WTA}}$ than $\delta_{\text{TCR,WTP}}$ would support our hypothesis. We tested whether this difference was statistically different by (1), comparing it to a utility specification which allowed for WTA and WTP differences but restricted the TCR interaction to be equal for both quadrants and evaluating whether the likelihood ratio statistic yielded a significant improvement in model fit (2), evaluating whether the $95\%$ confidence intervals of the $\delta_{\text{TCR,WTA}}$, $\delta_{\text{TCR,WTP}}$ parameters overlapped. The utility specification which assumes a homogenous interaction on both the WTA and the WTP is specified in Equation 4.4.

$$V_{ij} = \mu \cdot \Big\{ C_{ij} + \Big( \Big[ WTP \cdot T_{ij} \cdot D_{\text{WTP}} \Big] + \Big[ WTA \cdot T_{ij} \cdot D_{\text{WTA}} \Big] \Big) \cdot (1 + \delta_{\text{TCR}} \cdot D_{\text{TCR}}) \Big\} \quad \forall i \in I, j \in J \tag{4.4}$$

Where the $\delta_{\text{TCR}}$ parameter interacts equally on both the WTP and the WTA in a relative sense, as it is defined as a multiplicative interaction. This specification is thus purely meant to provide a baseline where we do allow differences in the WTA and the WTP, and subsequently evaluate what the average interaction is over the WTA and the WTP quadrants.

The limitations as discussed above defined the types of choice experiments which were eligible for testing. A summary is provided in Table 4.4.

### 4.2.1. Isolating TCR Effects

We ensured that the parameters which interact with TCR only capture $VTT$ differences due to a change in travel cost reimbursement. This was done because respondents with e.g. a high income are known to generally have a higher $VTT$. We observed in our sample that respondents with a high income also had a higher likelihood of receiving a full Travel Cost Reimbursement. Therefore we examined whether the distribution of TCR was related to other socioeconomic or design-specific variables of which we *a-priori* knew from practical experience and or the body of scientific literature that they interacted with either the $VTT$ or the scale parameter ($\mu$).

Examining whether there was a significant association between receiving a TCR and other variables, was done by performing either a Chi-square test of association for categorical variables or a KrusKal-Wallis test for non-categorical variables (Kruskal & Wallis, 1952; Pearson, 1900). Whenever the Chi-Square test or the Kruskal-Wallis test indicated a statistically significant relation the variable was added as an interaction. Instead of discussing all the test results we opted to provide them in Table 4.5. Do note that contrary to our a-priori expectations all of the variables tested yielded significant differences between receiving or not receiving a TCR. One of these unexpected results was that gender is associated with travel cost reimbursement as can be seen in Figure 4.2.

Furthermore, whenever an association between the variable and the TCR was significant we calculated the Cramers-V for the Chi-Squared test or the $\eta^2$ for the Kruskal-Wallis test to indicate the size of the effect. [2]

Subsequently, to ensure that there were no other respondent characteristics or survey-related aspects associated with a respondent receiving a TCR or not, a correlation analysis was performed. Based on the outcome of this correlation analysis, employment status and ZZP were added as interaction variables with the $VTT$. Employment status refers to whether a respondent is currently in between jobs or not. ZZP refers to a variable which indicates whether a respondent is self-employed or not.

[2] Effect size was not considered in the decision whether to include it as an interaction in the model, this decision was purely based on the statistical tests itself.

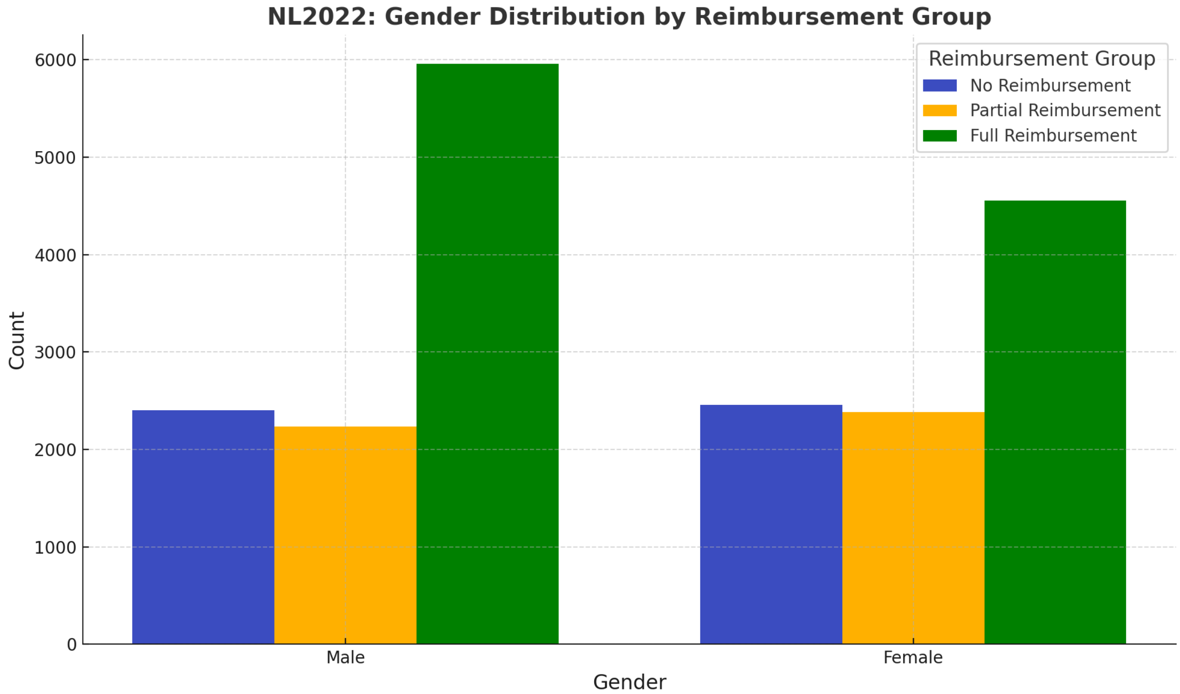



**Figure 4.2:** Counts of different types of travel cost reimbursement distributed across genders.

**Table 4.5:** Results of Chi-square and Kruskal-Wallis tests, to test whether a different number of reimbursed respondents across groups is statistically significant.

| Variable | Test | Statistic | p-value | D.O.F. | Cramér's V / $\eta^2$ | Effect Size |
|---|---|---|---|---|---|---|
| Gender | Chi-Square | 245.252*** | $< 0.001$ | 2 | 0.078 | Negligible |
| Income class | Chi-Square | 404.114*** | $< 0.001$ | 10 | 0.071 | Negligible |
| Age class | Chi-Square | 462.796*** | $< 0.001$ | 6 | 0.076 | Negligible |
| Education class | Chi-Square | 73.312*** | $< 0.001$ | 8 | 0.030 | Negligible |
| Reference time | Kruskal-Wallis | 1965.548*** | $< 0.001$ | - | 0.049 | Small |
| Reference costs | Kruskal-Wallis | 2186.181*** | $< 0.001$ | - | 0.055 | Small |
| Travel frequency | Kruskal-Wallis | 1863.306*** | $< 0.001$ | - | 0.047 | Small |
| Work situation | Chi-Square | 529.627*** | $< 0.001$ | 8 | 0.081 | Negligible |
| ZZP class | Chi-Square | 1000.340*** | $< 0.001$ | 2 | 0.158 | Small |
| Recruitment-type | Chi-Square | 155.648*** | $< 0.001$ | 2 | 0.062 | Negligible |
| Employment class | Chi-Square | 1959.111*** | $< 0.001$ | 4 | 0.157 | Small |

*Note.* *** p < 0.001.

## 4.2.2. Final Model Specification

The model development procedure as discussed above yielded the utility specification in Equation 4.5. This utility specification was used for the $SP1A$ experiment in the NL2022 dataset presented in Table 4.4. This specification was adopted where necessary to accommodate experiments with more attributes. We did this by simply adding the extra attribute multiplied by a parameter to the utility function. To accommodate the choice data from the UK2014 dataset only minor adjustments were made in the number of categories for socioeconomic interactions with the $VTT$.

$$V_{ij} = \mu \tag{4.5}$$
$$\cdot (1 + \mu_{\text{nTCR}} \cdot D_{\text{nTCR}} + \mu_{\text{pTCR}} \cdot D_{\text{pTCR}} + \mu_{\text{fTCR}} \cdot D_{\text{fTCR}}) \tag{4.6}$$
$$\cdot \left( \left( \frac{T_i^{\text{ref}}}{60} \right)^{\lambda_{T\mu}} \cdot \left( \frac{C_i^{\text{ref}}}{5} \right)^{\lambda_{C\mu}} \right) \tag{4.7}$$
$$\cdot \{ C_{ij} + [VTT \cdot T_{ij} \tag{4.8}$$
$$\cdot (1 + \delta_{\text{nTCR}} \cdot D_{\text{nTCR}} + \delta_{\text{pTCR}} \cdot D_{\text{pTCR}} + \delta_{\text{fTCR}} \cdot D_{\text{fTCR}}) \tag{4.9}$$
$$\cdot \left( 1 + \sum_{k=1}^{2} \delta_{\text{employment},k} \cdot D_{\text{employment},k} \right) \tag{4.10}$$
$$\cdot \left( 1 + \sum_{k=1}^{2} \delta_{\text{paidEmployment},k} \cdot D_{\text{paidEmployment},k} \right) \tag{4.11}$$
$$\cdot (1 + \delta_{\text{zzp}} \cdot D_{\text{zzp}}) \tag{4.12}$$
$$\cdot \left( 1 + \sum_{k=1}^{4} \delta_{\text{ageClass},k} \cdot D_{\text{ageClass},k} \right) \tag{4.13}$$
$$\cdot (1 + \delta_{\text{panel}} \cdot D_{\text{panel}} + \delta_{\text{intercept}} \cdot D_{\text{intercept}}) \tag{4.14}$$
$$\cdot \left( 1 + \sum_{k=1}^{5} \delta_{\text{edu},k} \cdot D_{\text{edu},k} \right) \tag{4.15}$$
$$\cdot (1 + \delta_{\text{male}} \cdot D_{\text{male}} + \delta_{\text{female}} \cdot D_{\text{female}}) \tag{4.16}$$
$$\cdot \left( 1 + \sum_{k=1}^{6} \delta_{\text{freq},k} \cdot D_{\text{freq},k} \right) \tag{4.17}$$
$$\cdot \left( \frac{T_i^{\text{ref}}}{T_0^{\text{ref}}} \right)^{\lambda_{\text{T}}} \cdot \left( \frac{C_i^{\text{ref}}}{C_0^{\text{ref}}} \right)^{\lambda_{\text{C}}} \tag{4.18}$$
$$\cdot \left( 1 + \sum_{k=1}^{6} \delta_{\text{inc},k} \cdot D_{\text{inc},k} \right) ] \} \tag{4.19}$$

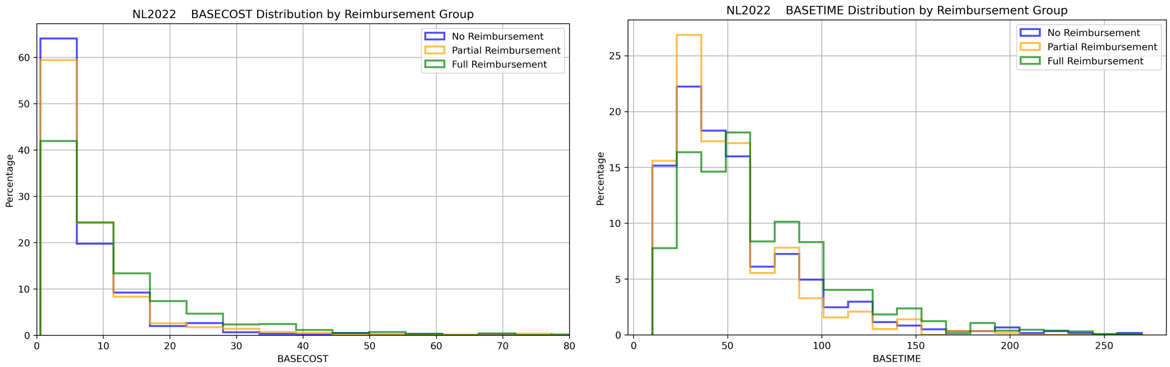
As mentioned earlier, the interactions added in Equation 4.5 were based on whether; (1) the variable of which was a-priori known to impact the $VTT$ or the scale had a statistically significant association with TCR or (2) no a-priori interaction with the $VTT$ or the scale parameter was expected but the variable correlated with TCR. These variables were added as multiplicative interactions. Categorical variables were dummy-coded. The interactions with the respondent's reference time and reference costs were added as power law interactions since the UK and the NL survey used what is called a pivot-type design.[3]

In a pivot-type design, the attribute levels are pivoted around the reference journey of the respondent. Thus respondents see time and cost levels which are familiar to them. For example; Respondents with a travel time of e.g. 10 minutes will thus be presented with attribute levels of 8 and 12 minutes. Respondents with a longer reference journey will be presented with alternatives with longer journey times. The attribute level range thus differs among respondents and is dependent on the time and cost levels of their reference journey. One can imagine that a respondent who is presented with an offer to reduce their travel time from 10 minutes to 8 minutes is more likely to be interested in that offer as it constitutes 20% of their travel time. However, when presenting a respondent to reduce their travel time from 60 minutes to 58 minutes, they might not be interested as this same 2-minute reduction is negligible on their total journey. Therefore respondents with lower travel times are more sensitive to absolute attribute level changes than respondents with high travel times. This is the reasoning behind adding the reference cost and reference time as interactions with the scale parameter as the

[3] The reference time and the reference costs, denoted by $T_i^{\text{ref}}$ and $C_i^{\text{ref}}$ respectively, reflect the time and the cost of the self-reported reference journey. Respondents were instructed to assume that all conditions were the same as their reference journey, except for the attributes varied in the experiment.

scale parameter captures the sensitivity of all the attributes relative to noise. The specific form of this interaction; is a power law is used as Kouwenhoven et al. (2022) found that these explained the data the best.

But why would it be important to add this power law interaction with the respondent's reference time and reference costs if you are interested in the effect of travel cost reimbursement? This was done because as visualised in Figure 4.3 respondents with different levels of TCR also had statistically significant different levels of reference time and reference costs, therefore there is a risk of confounding scale heterogeneity due to TCR with scale heterogeneity due to the pivot type design.




**(a)** Reference cost histogram

**(b)** Reference time histogram

**Figure 4.3:** Comparison of reference cost and reference time distributions across different levels of TCR.

## 4.3. Model Estimation

The developed model was estimated using Apollo Choice modelling software (Hess & Palma, 2019). For the Mixed Logit specifications, halton draws were used to simulate the randomly distributed $VTT$ (Halton, 1960). Sufficiently stable parameters were retrieved for 500 draws. The $VTT$ was given a lognormal distribution, as Kouwenhoven et al. (2022) found that after extensive testing this fitted the data the best. Furthermore, using the lognormal distribution ensures that the $VTT$ can only attain positive values in line with microeconomic theory.

The developed model was estimated on the other choice tasks with minimal adjustments to prevent confirmation bias. This approach was taken to ensure that the model's performance was assessed objectively across different choice tasks, without being influenced by prior expectations. By limiting adjustments, we aimed to validate the model's generalizability and robustness based on the data itself, rather than fitting it to preconceived outcomes. Consequently, this philosophy was also used when estimating the TCR interaction effects on the UK2014 data. Only necessary adjustments in, for example, the number of income classes and the number of age classes were accommodated. As a result, the model fit was not as good as it might have been able to, but we argue this approach increases the validity of the results.

## 4.4. Model application

After the model was developed we applied it to the societal cost-benefit analysis (SCBA) for the *Ring Utrecht* project (IenW, 2016), aiming to give an example of how incorporating or neglecting TCR can impact a SCBA. This provides a tangible example of policy implications. In order words; by applying our findings we aimed to assess whether it actually matters, in the context of an SCBA, if travel cost reimbursement is taken into account or neglected.

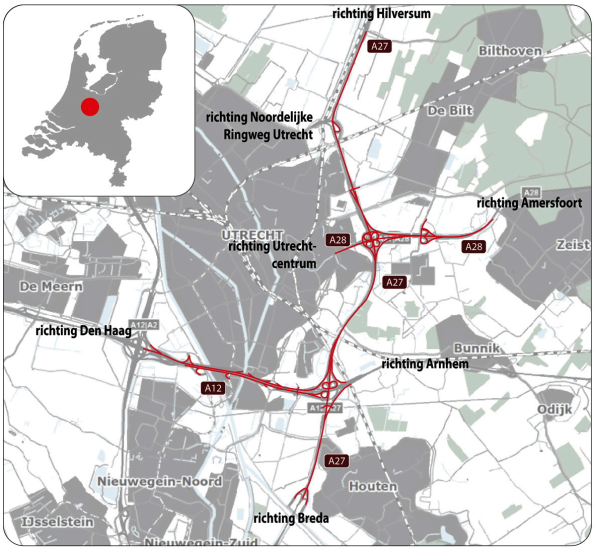



**Figure 4.4:** Overview Road Layout, Ring Utrecht (Ministerie van Infrastructuur en Waterstaat, 2022)

We opted for the SCBA of the *Ring Utrecht* project for several reasons. Firstly, for this specific SCBA the travel time reductions are approximately 75% of all benefits. Thus changes in the $VTT$ quickly echo through to the outcome of the SCBA, which makes it a good candidate to illustrate our findings. Secondly, this SCBA is currently the subject of a political discussion on whether or not the project or parts thereof should be executed. Therefore, we argue that possible implications could be even more socially relevant when illustrated on a project like the *Ring Utrecht*. Lastly, the SCBA itself is performed by a renowned research institute, *Decisio Amsterdam*, and it is presented as an example by Rijkswaterstaat (Dutch Road Authority), which makes it a credible SCBA to use.

The Ring Utrecht project itself is designed with dual objectives of improving traffic flow and safety while enhancing the quality of life for local residents. To achieve these goals, Rijkswaterstaat employed a multifaceted approach.

Firstly, the number of lanes on the Ring Utrecht will be increased to ensure a smoother and more efficient traffic flow. Secondly, traffic flows will be strategically separated to minimize congestion and reduce the risk of accidents. Additionally, a green crossing will be constructed at the nature reserve Amelisweerd, to minimize ecological disruption. The project also includes the renovation of existing ecopassages, ensuring safe crossings for wildlife. Furthermore, the implementation of quieter asphalt is planned to mitigate noise pollution. The new layout is depicted by Figure 4.4 as was published in Ministerie van Infrastructuur en Waterstaat (2022).

To assess whether accounting for TCR or neglecting it impacts the Ring Utrecht project, we first approximated the TCR effect on the $VTT$ suitable for use in SCBA's. Since the value used in an SCBA is *not* the same as the model average $VTT$. The reasoning is that the respondents in the sample are not necessarily representative of the Dutch population. As SCBA's rely on national average $VTT$ values it is necessary to perform a so-called *sample enumeration*. This is a procedure in which the $VTT$ values retrieved from the model are weighted according to the distributions in the population. This procedure ensures that the national average $VTT$ is representative of the entire population. These national average $VTT$ values are the values used in Dutch SCBA's. In our case study, we thus needed to calculate the national average $VTT$ for people who use a private car as their mode of transportation because the time benefits for the Ring Utrecht project are for car users only.

We illustrate how this sample enumeration works through $VTT$ differences in income. Let's for a moment assume that there are two income groups; low and high income. The high-income group has an arbitrary $VTT$ of € 20,-/h and the low-income group has an arbitrary $VTT$ of € 10,-/h. Whenever 70 % of the population in the Netherlands has a high income and 30 % has a low income, the weighted average $VTT$ is thereafter simply calculated as in Equation 4.20.

$$\frac{\text{VTT}_{\text{high}} \cdot \text{weight}_{\text{high}} + \text{VTT}_{\text{low}} \cdot \text{weight}_{\text{low}}}{\text{weight}_{\text{high}} + \text{weight}_{\text{low}}} = \frac{(20€/h \cdot 0.7 + €\,10,-/h \cdot 0.3)}{(0.7 + 0.3)} = €\,17,-/h \tag{4.20}$$

The weights and groups for the entire population in the Netherlands are naturally distributed among more dimensions, such as age, gender, education, travel purpose, etc, but the concept remains the same.

The weights we applied to ensure representativeness in the Dutch population were retrieved from *ODIN 2023* (CBS, 2024). This is a survey which contains information on the mobility patterns of Dutch citizens. In the *ODIN 2023* survey, participants were asked to document their destinations, travel purposes and modes of transportation. Furthermore, participants provided data on personal and household characteristics, including driver's license and vehicle ownership. Since the 2022 survey travel cost reimbursement has been added as an additional variable. The Dutch Bureau for Statistics,*CBS* ensures that responses in this survey are weighted to ensure any calculations done on the survey are representative of the Dutch population. The *ODIN2023* survey is also used by the *CBS* itself to publish official statistics on the travel behaviour of Dutch citizens. We were therefore confident that the *ODIN2023* data is suitable and valid.

We filtered the *ODIN2023* sample to include respondents who are employed and have either received or not received travel cost reimbursement. Furthermore, we only selected commute trips for the car and for which the respondent was the driver of the car. This created a subsample of $5664$ respondents who recorded $12.341$ trips. After the filtering, we enumerated over the remaining sample and assigned a $VTT$ value to every respondent according to the estimated $VTT$ values of our model.

However, *ODIN2023* does not provide information on whether a respondent is fully reimbursed for their travel expenses or only partly. Interestingly *ODIN2023* does provide information on the type of reimbursement, thus whether it is for example distance-based or a fixed amount. Because of this dichotomy between what is recorded in stated preference questionnaires and what is recorded in the *ODIN2023* dataset, we opted to explore the two extremes; (1) Respondents who reported receiving *any* travel cost reimbursement were assigned the $\text{VTT}_{\text{TCR}}$. (2) Respondents who reported receiving *any* travel cost reimbursement were assigned the $\text{VTT}_{\text{pTCR}}$ value. Naturally, respondents who did not receive a TCR were assigned the $\text{VTT}_{\text{nTCR}}$ value.

Thereafter we applied Equation 4.21 to calculate the weighted average for the two options.

$$\overline{VTT} = \frac{\sum_{i\in I}\sum_{t\in T_i} v_{ij}\cdot \text{VTT}i}{\sum_{i\in I}\sum_{j\in T_i} v_{ij}} \tag{4.21}$$

where $v_i j$ is the weight for trip $j$ for respondent $i$, $\text{VTT}_i$ is the $VTT$ value for respondent $i$, the set $T_i$ is the set of recorded trips by respondent $i$ and $\overline{VTT}$ is the weighted $VTT$. This equation was applied separately for *four* scenarios; with TCR effects and with TCR + scale effects & using the $\delta_{\text{pTCR}}$ and the $\delta_{\text{fTCR}}$ parameters. These were compared to the *baseline* scenario where we neglect TCR altogether[4]. This enabled us to approximate to what extent different model specifications impact a SCBA. The baseline scenario is thus in essence the scenario where TCR does not affect the $\overline{VTT}$ at all. Thus in this baseline scenario; $\overline{VTT} = \text{VTT}_i, \quad \forall i \in I$.

Our last step to approximate the impact of our findings on the *Ring Utrecht* project was to scale the $VTT^{\text{SCBA}}$ used in the SCBA according to the weighted averages of our four scenarios. We applied the relative difference in $\overline{VTT}$ between the scenario and the baseline scenario to the SCBA. Put differently; the relative difference between accounting and neglecting TCR is the multiplication factor we apply to the SCBA, mathematically denoted in Equation 4.22.

$$\text{VTT}_s^{\text{SCBA}} = \left(1 + \left(\frac{\overline{VTT_s} - \overline{VTT_b}}{\overline{VTT_b}}\right)\right)\cdot \text{VTT}^{\text{SCBA}}_{\text{current}} \tag{4.22}$$

Where $\text{VTT}^{\text{SCBA}}_{\text{current}}$ is the $VTT$ which was used in the SCBA of the Ring Utrecht project, $\overline{VTT_b}$ is the $\overline{VTT}$ resulting from the sample enumeration procedure for the baseline scenario, $\overline{VTT_s}$ is the $\overline{VTT}$ resulting from the sample enumeration procedure for the scenarios and where $\text{VTT}_s^{\text{SCBA}}$ is the $VTT$ value for the SCBA for scenario $s$.

We can illustrate this procedure by evaluating the scenario equivalent to the baseline scenario. Thus evaluate what happens in the scenario where we *do not* take travel cost reimbursement into account. The scenario we are evaluating is thus equivalent to the baseline; $\overline{VTT}_s = \overline{VTT}_b$. Inserting this equality in Equation 4.22 thus produces Equation 4.23.

[4]The scenario without TCR effects is denoted as the *baseline* scenario

$$\text{VTT}_s^{\text{SCBA}} = \left(1 + \left(\frac{0}{\overline{VTT_b}}\right)\right) \cdot \text{VTT}_{\text{current}}^{\text{SCBA}} \tag{4.23}$$

Here we can immediately observe that $\text{VTT}_s^{\text{SCBA}} = \text{VTT}_{\text{current}}^{\text{SCBA}}$, thus not taking travel cost reimbursement into account does not alter the outcome of the SCBA regardless of the absolute value of $\overline{VTT_b}$. This yields the desired outcome as the $\text{VTT}_{\text{current}}^{\text{SCBA}}$ originally did not take travel cost reimbursement into account. Our baseline scenario also did not take travel cost reimbursement into account, thus the effect on the SCBA should therefore be nonexistent, which is the case with this method.

We opted for this multiplication method (denoted in Equation 4.22) since retrieving the original $\overline{VTT_b}$ which the researchers used in 2013 would require using the same (older) ODIN dataset and would require to estimate TCR effects on the national VTT study of Significance et al. (2013). Both are impossible since neither the ODIN dataset nor the national VTT study recorded information on travel cost reimbursement. However, we are confident that our approach is representative as we ensured consistency with the baseline scenario by directly comparing scenarios with and without TCR. The multiplication method we used was robust as it produced no change in the SCBA when the tested scenario was identical to the baseline.

# 5
# Results

The goal of this chapter is to present the results and test the hypotheses. First, the effects on the VTT are analysed for models both including and excluding scale heterogeneity, thereafter we discuss the results regarding the scale parameter itself. Next, we discuss the results regarding the model which incorporates sign-effects and we end the results chapter with the findings from the *Ring Utrecht* SCBA.

## 5.1. Hypothesis 1, $VTT$

We first hypothesised that respondents with any travel cost reimbursement have a higher $VTT$ as compared to respondents who do not receive any travel cost reimbursement. As described in the method section we added dummy coded multiplicative interactions with the $VTT$ to test this hypothesis. We can thus evaluate the first hypothesis by observing the statistical significance of these interaction parameters, a t-ratio greater than 1.96 means that we can reject the Null hypothesis that the $VTT$ is the same as the reference group, in our case the nTCR respondents. As was already mentioned in the method chapter, we also included $VTT$ interactions for the heteroskedastic model. This enables us to (1) evaluate the first hypothesis for the heteroscedastic model, and (2) observe to what extent incorporating scale heterogeneity affects the evaluation of the first Hypothesis. In this first section, we focus on the first Hypothesis and evaluate it on the models including and excluding scale heterogeneity. The results regarding the degree of scale heterogeneity, thus the results regarding the second hypothesis are discussed in the next section.

As mentioned above, we can reject the first Null hypothesis whenever the TCR interaction parameters can be estimated with a t-ratio exceeding 1.96. This reflects a significance level of $\alpha = 5$%. This is equivalent to rejecting the Null hypothesis whenever the 95% confidence interval does not overlap with the VTT of respondents who do not receive any travel cost reimbursement. These equivalent methods are shown in Figure 5.1, Figure 5.2 Figure 5.3, Figure 5.4 and Figure 5.5. The coloured t-ratios correspond with the colours of the projected confidence intervals. The values correspond with the magnitude of the relative difference in $VTT$ and are both plotted graphically and presented in the table below the graph. All the values are presented as relative increases/decreases to the reference $VTT$, in our case the $VTT$ of respondents who did not receive any travel cost reimbursement. The values in the table below the graphs are the actual estimated values, thus they are reported as fractions. An increase in the $VTT$ of e.g. 31% is therefore reported in the table as 0.31. The figures for the NL2022 data report both the estimates for respondents receiving a full Travel Cost Reimbursement (fTCR) and a partial Travel Cost Reimbursement (pTCR). As the UK2014 data did not provide any information on what portion of the reference journey was reimbursed, the UK2014 figures only provide one reimbursement interaction per estimation. We added the number of observations to allow for a more nuanced interpretation of the results because an insignificant difference can sometimes be attributed to a lack of observations.

As shown in Figure 5.1 and in Figure 5.2 the results are mixed. We found that for a simple time-money tradeoff, denoted by SP1A in Figure 5.1 and in Figure 5.2 commuter respondents with a full Travel Cost Reimbursement (fTCR) had a significantly higher VTT as compared to respondents with no Travel Cost reimbursement (nTCR) thus supporting our first Hypothesis that respondents with a travel cost reimbursement have a higher $VTT$ than respondents who do not receive a travel cost reimbursement. This is the case for models with and without scale heterogeneity. The only exception is the Light Rail Transit (LRT) commute estimation in Figure 5.2. However, do observe that the number of observed choices for respondents with a fTCR was 180. Therefore a likely explanation is that these respondents coincidentally had a VTT which closely matched that of the nTCR respondents.

The business estimates of the time-money tradeoff (Figure 5.1 & Figure 5.2) were less convincing than the commute estimates. Three out of the six business estimates across the NL2022 and the UK2014 datasets showed statistically significant fTCR interactions at a 5% confidence level. Interestingly, the estimates for Business Light Rail Transit (LRT) in and for Business Bus Tram Metro (BTM) do indicate statistically significant differences in VTT compared to the reference situation of not receiving a Travel Cost Reimbursement.

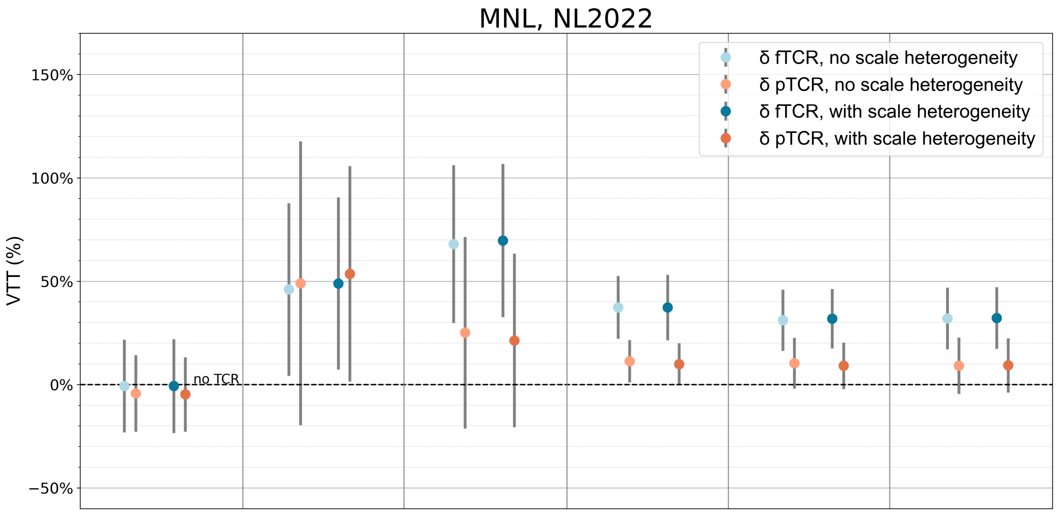
The estimates for the more complex experiments, SP3A & SP4A, were not disaggregated across modes and purposes to obtain realistic model parameters. Instead, we opted to create a model which included a mode- and purpose-specific constant to account for potential VTT differences. The estimates for these models are presented in Figure 5.5, almost none of the results yielded significant effects. The only exceptions are Car Business and Car Commute for SP2A, where respondents receiving a full Travel Cost Reimbursement (fTCR) had a significantly higher VTT as compared to respondents who received no Travel Cost Reimbursement (nTCR).

The mixed Logit specification generally widened the confidence intervals of the TCR interaction. This deterioration in statistical significance did not cause any fTCR interactions to become insignificant at a 95% confidence level. Conversely, when examining the Commute estimates in Figure 5.1 and Figure 5.3, we can observe that the mixed logit specification did cause the interactions with the $VTT$ and the partial travel cost reimbursement, denoted by pTCR, to decrease in magnitude and as a consequence, the difference in $VTT$ for pTCR respondents is no longer statistically significant w.r.t. respondents not receiving a travel cost reimbursement.

Interestingly, when comparing the Mixed Logit and the MNL estimates for Train Commute (Figure 5.1, Figure 5.3) we can observe that the magnitude of the $\delta_{\text{fTCR}}$ parameter increased when comparing the MNL and the Mixed Logit specifications. For the model excluding scale heterogeneity, this was an increase from 32% (t-ratio, 4.36) to 45% (t-ratio, 4.04) and for the heteroscedastic model, this was an increase from 32% (t-ratio, 4.11) to 43% (t-ratio, 3.79).

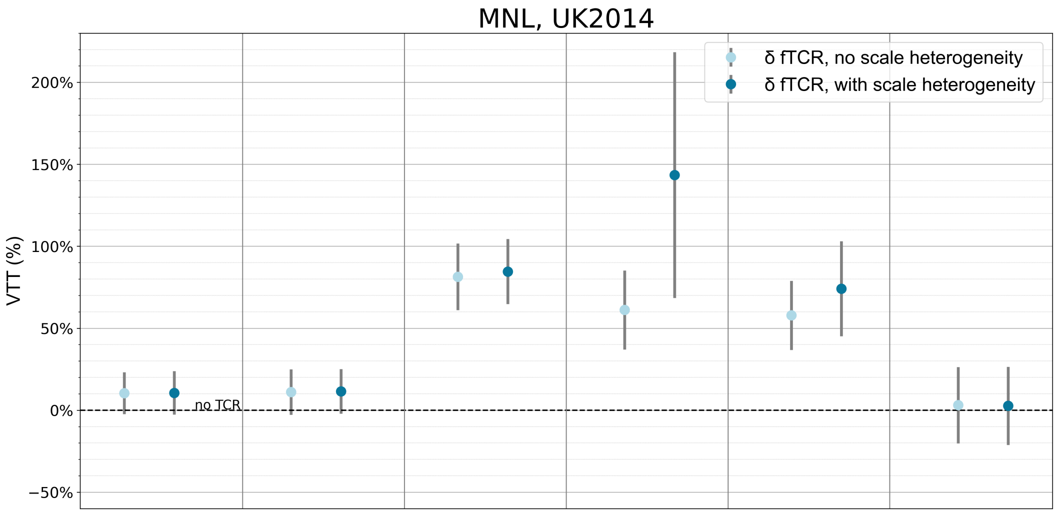
The findings, for commuter respondents with a full Travel Cost Reimbursement across the UK2014 and the NL2022 datasets, between incorporating and neglecting scale heterogeneity and for Mixed- and Multinomial Logit, supported our hypothesis that respondents with a travel cost reimbursement have a higher $VTT$. The results indicated that the difference in $VTT$ is somewhere between 29% and 143%. The lowest reported t-ratio was 2.48 whenever we excluded the LRT commute estimation since it lacked observations.

The other findings strongly suggested that TCR affects the $VTT$ but were either inconsistent or insignificant.



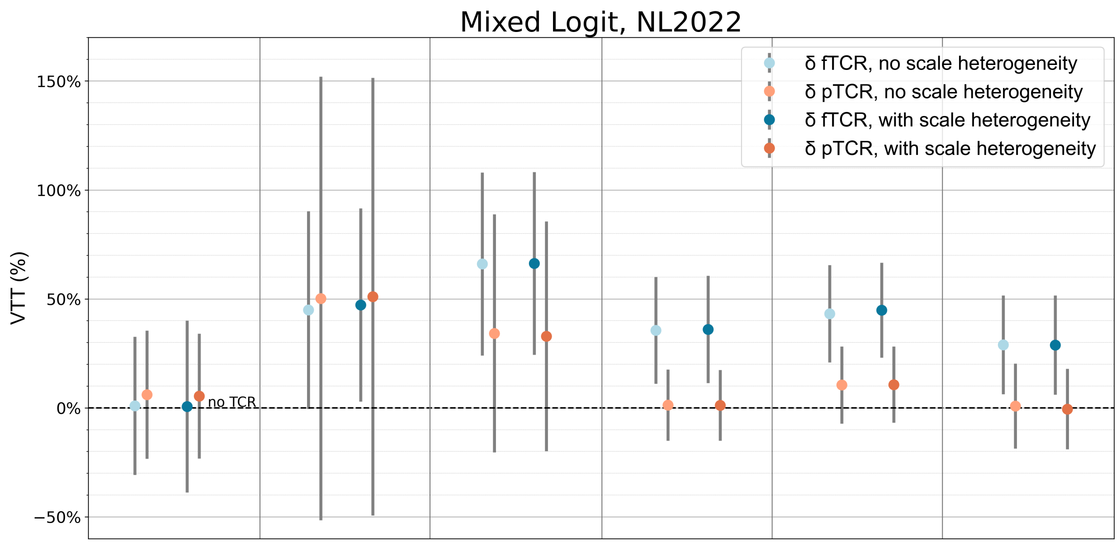

| | | | | | | |
|---|---|---|---|---|---|---|
| δ_fTCR. | -0.01, (-0.06) | 0.46, (2.16) | 0.68, (3.49) | 0.37, (4.83) | 0.31, (4.11) | 0.32, (4.21) |
| δ_pTCR. | -0.04, (-0.45) | 0.49, (1.4) | 0.25, (1.06) | 0.11, (2.17) | 0.1, (1.64) | 0.09, (1.32) |
| δ_fTCR | -0.01, (-0.06) | 0.49, (2.3) | 0.7, (3.69) | 0.37, (4.62) | 0.32, (4.36) | 0.32, (4.23) |
| δ_pTCR | -0.05, (-0.52) | 0.54, (2.02) | 0.21, (1.0) | 0.1, (1.93) | 0.09, (1.57) | 0.09, (1.39) |
| Δ(μ) fTCR | 0.08, (0.71) | -0.07, (-0.4) | -0.24, (-1.29) | -0.09, (-1.17) | -0.16, (-1.84) | -0.09, (-1.14) |
| Δ(μ) pTCR | 0.25, (1.91) | 0.96, (0.98) | -0.04, (-0.12) | 0.29, (3.05) | 0.09, (0.73) | 0.12, (1.05) |
| Modes | Car | Train | BTM | Car | Train | BTM |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 1376 | 1192 | 752 | 2120 | 3136 | 1936 |
| #obs pTCR | 624 | 48 | 152 | 2168 | 752 | 872 |
| #obs nTCR | 504 | 248 | 328 | 1552 | 872 | 1352 |

**Figure 5.1:** NL2022, time-money tradeoff results, including 95% confidence intervalls. $\Delta(\mu)$ estimates are reported as fractional differences with respect to the nTCR observations including their t-statistic in between parentheses.



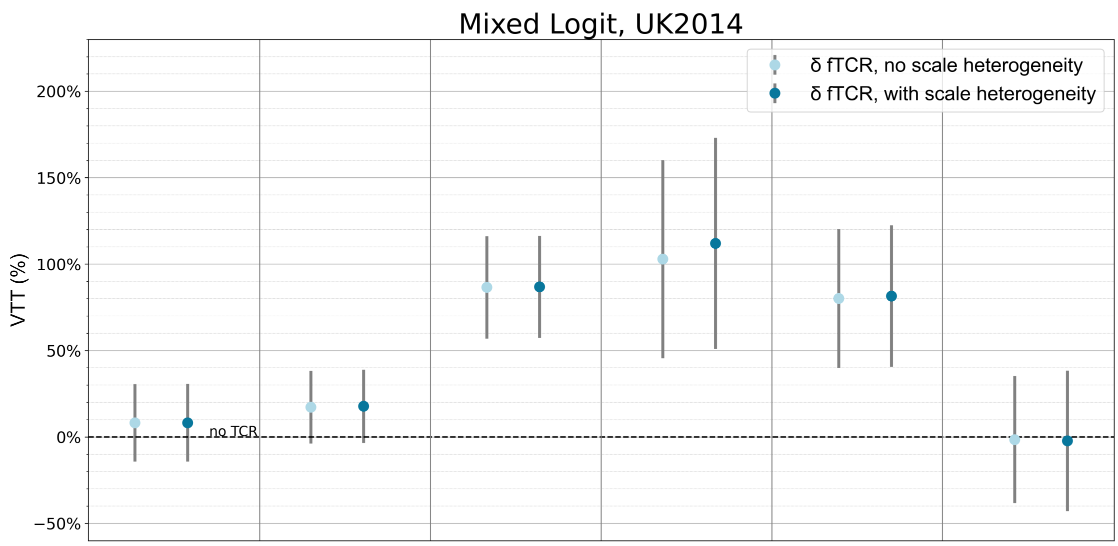

| | | | | | | |
|---|---|---|---|---|---|---|
| δ_fTCR. | 0.1, (1.59) | 0.11, (1.57) | 0.81, (7.87) | 0.61, (4.97) | 0.58, (5.39) | 0.03, (0.26) |
| δ_fTCR | 0.11, (1.57) | 0.11, (1.65) | 0.85, (8.34) | 1.43, (3.75) | 0.74, (5.01) | 0.03, (0.22) |
| Δ(μ) fTCR | -0.01, (-0.1) | -0.06, (-0.61) | -0.28, (-1.73) | -0.62, (-7.34) | -0.26, (-2.9) | 0.02, (0.14) |
| Modes | Car | Train | LRT | Car | Train | LRT |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 3360 | 3455 | 495 | 895 | 810 | 180 |
| #obs pTCR | 0 | 0 | 0 | 0 | 0 | 0 |
| #obs nTCR | 965 | 940 | 655 | 3940 | 3775 | 2720 |

**Figure 5.2:** UK2014, time-money tradeoff results, including 95% confidence intervalls. $\Delta(\mu)$ estimates are reported as fractional differences with respect to the nTCR observations including their t-statistic in between parentheses.



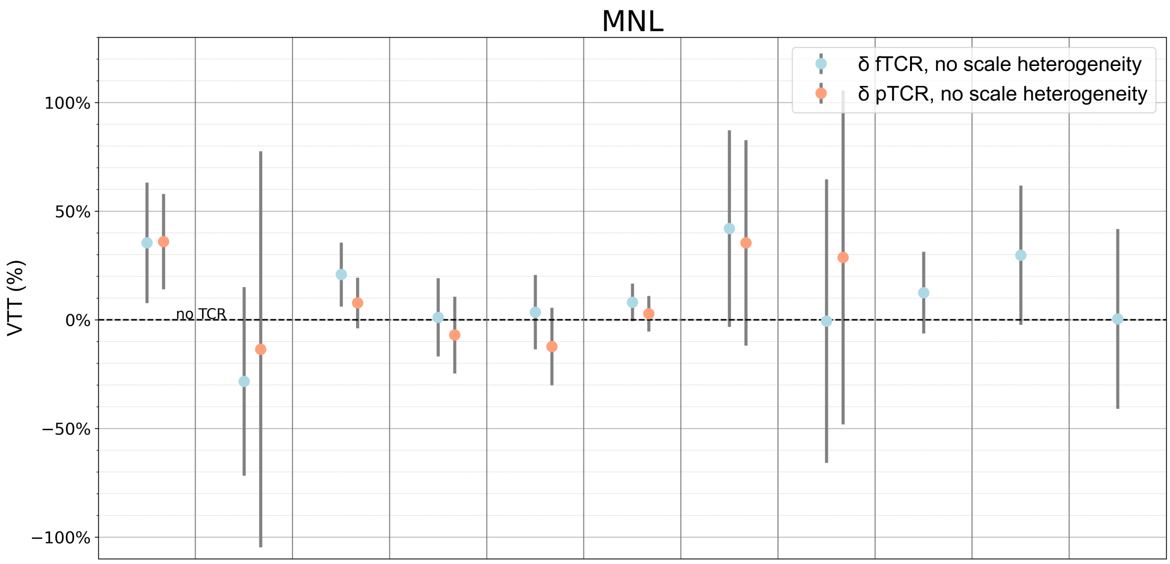

| | | | | | | |
|---|---|---|---|---|---|---|
| δ_fTCR. | 0.01, (0.06) | 0.45, (1.94) | 0.66, (3.08) | 0.36, (2.85) | 0.43, (3.79) | 0.29, (2.5) |
| δ_pTCR. | 0.06, (0.4) | 0.5, (0.97) | 0.34, (1.23) | 0.01, (0.15) | 0.11, (1.16) | 0.01, (0.09) |
| δ_fTCR | 0.01, (0.03) | 0.47, (2.09) | 0.66, (3.1) | 0.36, (2.87) | 0.45, (4.04) | 0.29, (2.48) |
| δ_pTCR | 0.05, (0.37) | 0.51, (1.0) | 0.33, (1.22) | 0.01, (0.14) | 0.11, (1.2) | -0.01, (-0.06) |
| Δ(μ) fTCR | 0.24, (1.73) | -0.24, (-1.02) | -0.07, (-0.34) | -0.11, (-1.2) | -0.24, (-2.24) | -0.11, (-1.18) |
| Δ(μ) pTCR | 0.59, (2.92) | -0.25, (-0.48) | 0.13, (0.37) | 0.1, (0.97) | -0.18, (-1.5) | 0.43, (2.46) |
| Modes | Car | Train | BTM | Car | Train | BTM |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 1376 | 1192 | 752 | 2120 | 3136 | 1936 |
| #obs pTCR | 624 | 48 | 152 | 2168 | 752 | 872 |
| #obs nTCR | 504 | 248 | 328 | 1552 | 872 | 1352 |

**Figure 5.3:** NL2022, time-money tradeoff results, including 95% confidence intervalls. $\Delta(\mu)$ estimates are reported as fractional differences with respect to the nTCR observations including their t-statistic in between parentheses.




| | | | | | | |
|---|---|---|---|---|---|---|
| δ_fTCR. | 0.08, (0.72) | 0.17, (1.61) | 0.87, (5.74) | 1.03, (3.52) | 0.8, (3.91) | -0.01, (-0.08) |
| δ_fTCR | 0.08, (0.72) | 0.18, (1.64) | 0.87, (5.77) | 1.12, (3.59) | 0.82, (3.9) | -0.02, (-0.11) |
| Δ(μ) fTCR | 0.11, (1.23) | 0.16, (1.56) | -0.04, (-0.22) | -0.35, (-3.98) | -0.07, (-0.55) | -0.2, (-1.34) |
| Modes | Car | Train | LRT | Car | Train | LRT |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 3360 | 3455 | 495 | 895 | 810 | 180 |
| #obs pTCR | 0 | 0 | 0 | 0 | 0 | 0 |
| #obs nTCR | 965 | 940 | 655 | 3940 | 3775 | 2720 |

**Figure 5.4:** UK2014, time-money tradeoff results, including 95% confidence intervalls. $\Delta(\mu)$ estimates are reported as fractional differences with respect to the nTCR observations including their t-statistic in between parentheses.



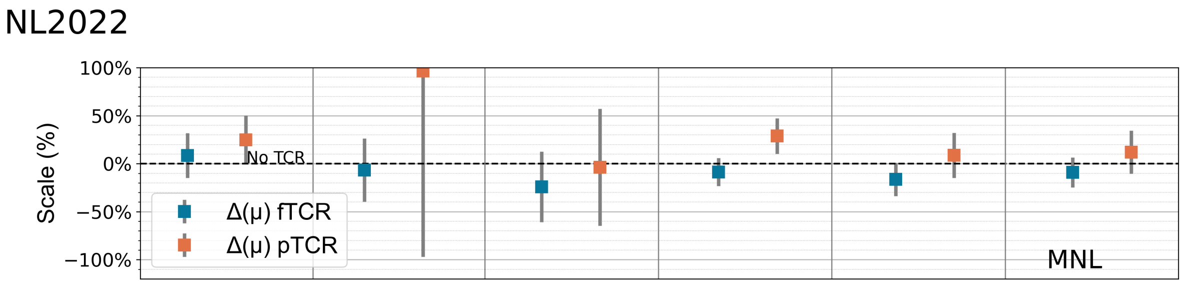

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| δ_fTCR. | 0.35, (2.5) | -0.28, (-1.28) | 0.21, (2.78) | 0.01, (0.12) | 0.03, (0.4) | 0.08, (1.82) | 0.42, (1.82) | -0.01, (-0.02) | 0.12, (1.3) | 0.3, (1.82) | 0.0, (0.02) |
| δ_pTCR. | 0.36, (3.22) | -0.14, (-0.29) | 0.08, (1.31) | -0.07, (-0.78) | -0.12, (-1.36) | 0.03, (0.67) | 0.35, (1.47) | 0.29, (0.73) | nan, (nan) | nan, (nan) | nan, (nan) |
| Dataset | NL2022 | NL2022 | NL2022 | NL2022 | NL2022 | NL2022 | NL2022 | NL2022 | UK2014 | UK2014 | UK2014 |
| Sub-experiment | SP2A | SP2A | SP2A | SP2A | SP2A | SP2A | SP3A | SP4A | SP2A | SP2A | SP2A |
| Modes | Car | Train | Car | Train | BTM | all | BTM Train | BTM Train | Car | Train | LRT |
| Purposes | Business | Business | Commute | Commute | Commute | all | all | all | Business | Business | Business |
| #obs fTCR | 1376 | 792 | 2120 | 2048 | 1312 | 8136 | 1280 | 1096 | 3360 | 3455 | 495 |
| #obs pTCR | 624 | 32 | 2168 | 464 | 544 | 3920 | 344 | 352 | 0 | 0 | 0 |
| #obs nTCR | 504 | 144 | 1552 | 560 | 832 | 3808 | 528 | 520 | 965 | 940 | 655 |

**Figure 5.5:** Estimates for more complex choice tasks across both the NL2022 and the UK2014 dataset. For SP3A and SP4A only estimates which combined modes and purposes are included to achieve to necessary number of observations.

## 5.2. Hypothesis 2, Scale Heterogeneity

As a second hypothesis, we expected that respondents with a travel cost reimbursement make less consistent choices as compared to respondents who did not receive any travel cost reimbursement. We tested this hypothesis by incorporating interactions with the scale of the utility function. Our findings did generally *not* support our hypothesis at a significance level of 95%. In other words; we did not find convincing evidence which supported our hypothesis that respondents with a travel cost reimbursement make less consistent choices.

The way we reported our results is similar to the style of reporting for the first hypothesis, only now we report the difference in scale as opposed to the difference in $VTT$. The results are presented in Figure 5.6 and Figure 5.7. The colours in the table again match those in the graph. In contrast to the figures for the first hypothesis we now included both the Mixed Logit and the MNL estimates in one figure. Whenever a $\Delta(\mu)$ estimate is lower than 0, it indicates that respondents with a travel cost reimbursement had a lower scale than respondents without a travel cost reimbursement. A $\Delta(\mu)$ estimate higher than 0 naturally implies the opposite. Whenever the t-ratio > 1.96 or similarly whenever the 95% confidence interval does not overlap with the horizontal 0% line, we reject the Null hypothesis for the respective estimate at a significance level of $\alpha = 0.05$, therefore supporting our hypothesis that TCR reduces choice consistency.




| | | | | | | |
|---|---|---|---|---|---|---|
| Δ(μ) fTCR | 0.08, (0.71) | -0.07, (-0.4) | -0.24, (-1.29) | -0.09, (-1.17) | -0.16, (-1.84) | -0.09, (-1.14) |
| Δ(μ) pTCR | 0.25, (1.91) | 0.96, (0.98) | -0.04, (-0.12) | 0.29, (3.05) | 0.09, (0.73) | 0.12, (1.05) |
| Modes | Car | Train | BTM | Car | Train | BTM |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 1376 | 1192 | 752 | 2120 | 3136 | 1936 |
| #obs pTCR | 624 | 48 | 152 | 2168 | 752 | 872 |
| #obs nTCR | 504 | 248 | 328 | 1552 | 872 | 1352 |
| Δ(μ) fTCR | 0.24, (1.73) | -0.24, (-1.02) | -0.07, (-0.34) | -0.11, (-1.2) | -0.24, (-2.24) | -0.11, (-1.18) |
| Δ(μ) pTCR | 0.59, (2.92) | -0.25, (-0.48) | 0.13, (0.37) | 0.1, (0.97) | -0.18, (-1.5) | 0.43, (2.46) |

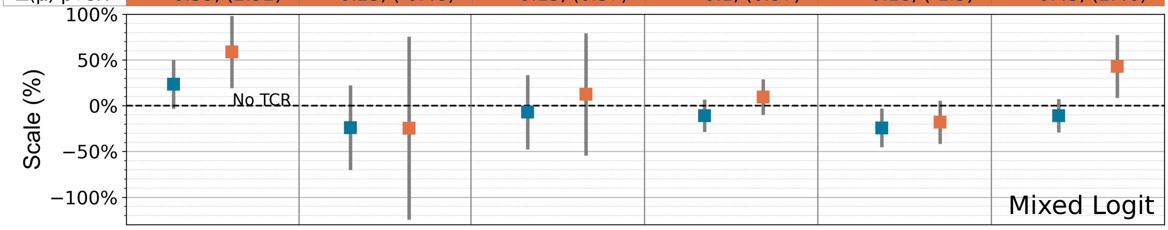


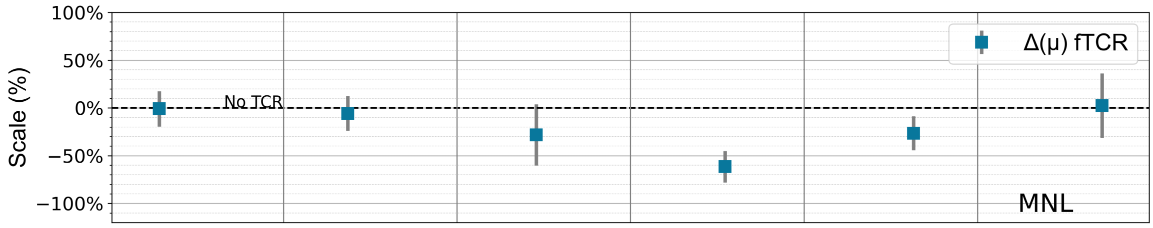

**Figure 5.6:** NL2022 time-money tradeoff results for the models which included scale heterogeneity. The difference in scale is presented in the graph and reported in the table. The graph includes the 95% confidence interval whilst the table includes the t-statistic in between parentheses.

Examining Figure 5.6, we found that, across respondents with a full Travel Cost Reimbursement for both the Mixed- and Multinomial Logit models the $\Delta(\mu)$ estimates are of the expected sign, except for the car business estimate, hinting that respondents with a fTCR might have a lower scale parameter as compared to respondents without a TCR. However, none of these interactions were statistically significant at $\alpha = 0.05$, except for the Mixed Logit estimation for Train Commute.

Interestingly for the MNL Car Commute estimate and the Mixed Logit BTM commute estimate the interactions regarding a partial Travel Cost Reimbursement indicated a higher scale compared to nTCR respondents. This is not in line with our second hypothesis which stated that an increase in TCR would lead to a decrease in scale. These estimations thus indicate that respondents with a partial travel cost reimbursement make more consistent choices compared to nTCR respondents.

The scale interaction estimates, denoted by $\Delta(\mu)$fTCR, for UK2014 in Figure 5.2 are quite similar. All of the interactions are of the expected sign, hinting that respondents with a full Travel Cost Reim-

UK2014




| Δ(μ) fTCR | -0.01, (-0.1) | -0.06, (-0.61) | -0.28, (-1.73) | -0.62, (-7.34) | -0.26, (-2.9) | 0.02, (0.14) |
|---|---|---|---|---|---|---|
| Modes | Car | Train | LRT | Car | Train | LRT |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 3360 | 3455 | 495 | 895 | 810 | 180 |
| #obs pTCR | 0 | 0 | 0 | 0 | 0 | 0 |
| #obs nTCR | 965 | 940 | 655 | 3940 | 3775 | 2720 |
| Δ(μ) fTCR | 0.11, (1.23) | 0.16, (1.56) | -0.04, (-0.22) | -0.35, (-3.98) | -0.07, (-0.55) | -0.2, (-1.34) |

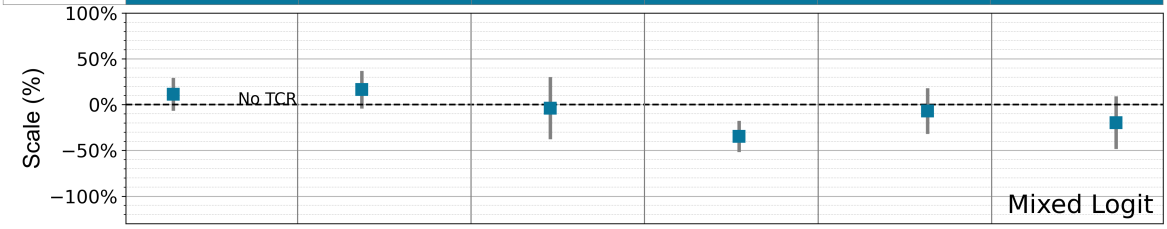



**Figure 5.7:** UK2014 time-money tradeoff results for the models which included scale heterogeneity. The difference in scale is presented in the graph and reported in the table. The graph includes the 95% confidence interval whilst the table includes the t-statistic in between parentheses.

bursement (fTCR) had on average a higher scale compared to respondents who had to pay for the travel costs themselves but again these interactions were not significant. The only estimates with significant scale interactions were Car Commute and Train Commute; -0.62 t-ratio (7.34) and -0.26 t-ratio (2.9) respectively. The $\Delta(\mu)$fTCR UK2014 Car Commute estimate was the largest in magnitude indicating a scale reduction of 62% compared to respondents who received no Travel Cost Reimbursement. Additionally, this relatively large difference in scale also reduced the VTT interaction for this particular estimate as can be observed in Figure 5.2. Whenever scale heterogeneity was not taken into account the VTT difference was 61% (t-ratio, 4.97). Conversely, taking scale heterogeneity into account yielded a VTT difference of, on average, 143% (t-ratio, 3.75).

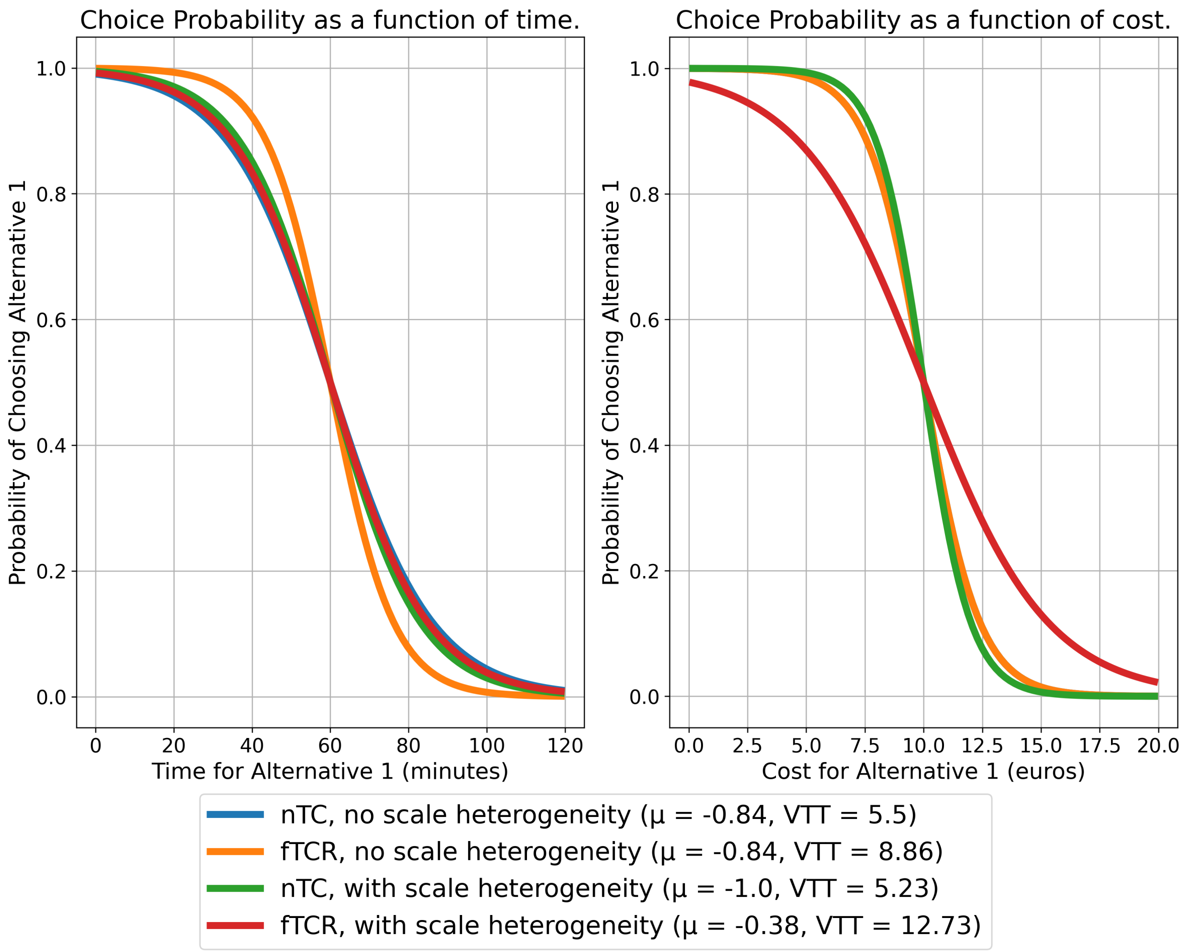



**Figure 5.8:** Choice probabilities UK2014 Car Commute

To further investigate this effect we plotted the corresponding choice probabilities in Figure 5.8 for UK2014 Car Commute. We can observe that the higher VTT for fTCR respondents yielded a slightly higher sensitivity to attribute level changes for time. However, when taking scale heterogeneity into account the sensitive w.r.t. time decreased again because of the dramatic scale decrease. The increase in $VTT$ but the decrease in scale seemingly cancelled each other out. Naturally, the sensitivity to changes in cost remained the same whenever the scale parameter was fixed. But when allowing scale heterogeneity the sensitivity to changes in the time attribute is reduced drastically.

This gives some conceptual insight into why the VTT is not higher for the estimates which ignore scale heterogeneity. If the VTT for fTCR respondents were even higher it would approximate the "true" VTT better, but increasing the VTT would also imply a higher sensitivity to changes in time, whereas the VTT difference was in this estimate dominated by a decreased sensitivity to changes in cost. Put differently, we can observe that for the UK2014 Car Commute respondents, taking scale heterogeneity into account, the VTT difference between nTCR and fTCR respondents can be traced back to a decreased sensitivity to changes in cost in an MNL specification.

The difference in scales for the Car Commute estimate as was observed in the MNL specification, was not as apparent in the Mixed Logit specification, as can be observed in Figure 5.7. In the MNL specification, the scale of the utility for fTCR respondents was 62% lower relative to nTCR respondents, whereas for the Mixed Logit specification the difference in scale was reduced to just 35%, (t-ratio, 3.98). Simultaneously, the indirect effect of the scale difference on the $VTT$ is also greatly reduced as can be observed by revisiting Figure 5.4 Generally speaking, the differences in scales when comparing MNL and Mixed Logit related to scale heterogeneity do not seem to follow a particular pattern. The Mixed Logit specification, does not cause scale heterogeneity between nTCR and fTCR respondents to increase or decrease.

## 5.3. Hypothesis 3, Sign Effects

To achieve statistically significant parameters and test whether or not there is a disparity between the willingness to pay and the willingness to accept, the model incorporated all modes and purposes in one estimation. This was a necessary aggregation to achieve significant results. However, the model did include parameters to account for apriori known differences between the VTT across modes and purposes. The results of this estimation are presented in Table 5.1. Interpretation of these results is less trivial than before. Firstly, both models, including a TCR sign effect and excluding a TCR sign effect *both*, allowed us to estimate a VTT in the WTP quadrant and the WTA quadrant. We can observe that for both estimations, the VTT in the WTP quadrant is higher as compared to the VTT in the WTA quadrant, denoted by the *VTT(WTP)* and the *VTT(WTA)* parameters. Conceptually this means respondents were willing to pay more to avert a deterioration in travel time than they were willing to pay to improve their travel time. This aligns with empirical evidence (Hess et al., 2007; Kouwenhoven et al., 2022) and with the behavioural notion of loss aversion (Tversky & Kahneman, 1974). The fact that these were different from one another thus implies that the iso-utility curve is no longer a straight line, but has a discontinuity at the reference time and cost of the respondent. We originally hypothesised that this difference in willingness to accept (WTA) and willingness to pay (WTP) might be the sole reason that there is a difference in the VTT to begin with. To test this we estimated a travel cost reimbursement (TCR) interaction on both sides of the discontinuity, on the VTT(WTP) and the VTT(WTA), denoted by the estimate on the left in Table 5.1. We expected that the TCR interaction with the VTT would be explained solely by the interaction with the VTT(WTA). Our results supported this hypothesis. The nTCR interaction with the VTT(WTA), denoted by nTCR(WTA), indicated a significant decrease of 25% in the VTT (t-ratio, 3.23), whilst the nTCR(WTP) parameter indicated an insignificant decrease of 10% (t-ratio, 1.01). This indicated that, in this sample, the VTT difference can be mostly attributed to a difference in the WTA quadrant. However, when comparing both models in terms of their LogLikelyhood the Ben-Akiva and Swait test indicated that the model improvement is not statistically significant at a 5% level. Thus, the model including a TCR sign effect did not outperform the model excluding a TCR sign effect in terms of model fit. Therefore, it could be due to random chance that we observed a slight improvement in LogLikelihood.

**Table 5.1:** Estimation results Including and Excluding TCR Sign Effect. Both models account for a discontinuity in the VTT at the reference time and costs of every individual respondent.

| | | (1) Including TCR sign effect | | (2) Excluding TCR sign-effect | |
|---|---|---|---|---|---|
| Modelled outcomes | | 3690 | | 3690 | |
| LL(final) | | -1653.62 | | -1654.28 | |
| Adj.Rho-square (C) | | 0.3404 | | 0.3405 | |
| Parameters | Unit | Estimate | t-ratio(0) | Estimate | t-ratio(0) |
| $\mu$ | - | -0.89 | -21.88 | -0.89 | -21.95 |
| VTT(wtp) | €$/h$ | 8.77 | 8.61 | 8.97 | 8.77 |
| VTT(WTA) | €$/h$ | 11.79 | 9.04 | 11.62 | 9.07 |
| fTCR(WTP) | % | 0.00 | NA | - | - |
| nTCR(WTP) | % | -10% | -1.01 | - | - |
| fTCR(WTA) | % | 0.00 | NA | - | - |
| nTCR(WTA) | % | -25% | -3.23 | - | - |
| fTCR VTT | % | - | - | 0.00 | NA |
| nTCR VTT | % | - | - | -19% | -3.05 |

## 5.4. Application: Ring Utrecht SCBA

This section describes the results regarding the calculation of the weighted average as denoted in section 4.4 and the application to the SCBA of the *Ring Utrecht* project.

**Amenability** As we were unable to prove that travel cost reimbursement has a significant effect on the $VTT$ for business respondents using the car, we evaluated the *Ring Utrecht* SCBA assuming that the $VTT$ for business travellers is unaffected by travel cost reimbursement. For Car Commute, significant effects were found for both the MNL- and the Mixed Logit models. In the NL2022 dataset estimated $VTT$ differences were significant for both model specifications; without scale effects $\delta_{\text{TCR}}$ = 36%, t-ratio (2.85) and with scale effects $\delta_{\text{TCR}}$ = 36%, t-ratio (2.87) for the Mixed-Logit estimates. Therefore we opted to only vary the $VTT$ in the *Ring Utrecht* SCBA for commuters. We opted for the Mixed-Logit estimations as these fitted the data significantly better than the MNL estimations.

**Interpretation** We present our findings in Table 5.2. It should be interpreted as follows; the five scenarios as described in the procedure of section 4.4 are listed as the columns of the table. The base scenario, where travel cost reimbursement interactions are not taken into account, is included in the first column. The other columns present the scenarios where TCR is taken into account, with and without scale heterogeneity and with either using $\delta_{\text{fTCR}}$ or $\delta_{\text{pTCR}}$ as inputs for the sample enumeration.

Going from top to bottom in the table; model estimates are presented first, they are the same as the model estimates presented earlier in Figure 5.1. These model estimates were used as input for the sample enumeration procedure. The complete model estimations used in the *Ring Utrecht* case study can be found in Appendix C. Further down, the weighted values of travel time as calculated through the sample enumeration procedure are presented. These weighted values are denoted by $\overline{VTT}$.

Lastly, the different cases of the *Ring Utrecht* SCBA are presented. The cases are defined by; (1) different road layouts denoted by *Selecteren* and *Compact* and (2) under different economic scenarios, denoted by *GE** for Global Economies and *RC** for Regional Communities. The economic scenarios are similar to the Dutch WLO scenarios, which are defined by the CPB, the Netherlands Bureau for Economic Policy Analysis, and the PBL, the Netherlands Environmental Assessment Agency. A recent version of these scenarios can be found at PBL and CPB (2020). The cases and their background are discussed in detail in the SCBA itself (IenW, 2016). The *Value of Travel Time Savings* in the first scenario (presented in the first column) are exactly the same as in the *Ring Utrecht* SCBA (IenW, 2016). For every scenario the Travel Time Savings in terms of the number of hours is constant. However, the monetary valuation of these hours differs according to the increase of the $\overline{VTT}$ compared to the $\overline{VTT}$ of the Baseline scenario. Table 5.2 ends with the monetary impact on respectively the yearly and the perpetually discounted commuter and total values of travel time savings.

**Synthesis** The $VTT$ effects correspond with the model estimations which were presented earlier in Figure 5.3. The presented t-ratios correspond with the confidence intervals presented earlier in Figure 5.4. The VTT values are very similar between neglecting and including scale heterogeneity. Both indicated a significant difference of approximately 36% in $VTT$ between receiving- and not receiving a travel cost reimbursement. Incorporating TCR caused the $VTT_{\text{nTCR}}$ and the $VTT_{\text{pTCR}}$ to reduce and the $VTT_{\text{fTCR}}$ to increase relative to the baseline scenario. This is to be expected as when we do not consider travel cost reimbursement the $VTT$ is estimated on respondents who receive a fTCR, on respondents who receive a pTCR and on respondents who do not receive a travel cost reimbursement.

Next, the average weights are presented. They are simply calculated as;

$$\overline{v_{\text{TCR}}} = \left( \sum_{i \in I_{\text{TCR}}} \sum_{t \in T_i} v_{it} \right) / \left( \sum_{i \in I} |T_i| \right) \tag{5.1}$$

Where, $I_{\text{TCR}}$ denotes the set of respondents who receive any Travel Cost Reimbursement, $|T_i|$ denotes the number of trips in the set of trips for respondent $i$ and $v_{it}$ denotes the weight for trip $j$ made by respondent $i$. Thus in the *ODIN2023* survey, 78% of Car Commute trips were made by respondents who reported receiving some form of travel cost reimbursement, whereas 22% of trips were made without any form of travel cost reimbursement. This distribution is very similar to the 80% reported by

**Table 5.2:** Comparison of VTT with and without accounting for TCR. The Baseline is the model which does not take TCR into account. $\overline{VTT}$ is the weighted VTT as calculated through the sample enumeration procedure. VTTS denotes the Value of Travel Time Savings. The grey VTT values are not used in the sample enumeration.

| **Scenario** | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** |
|---|---|---|---|---|---|
| | | **Excl. heterskedasticty** | | **Incl. heterskedasticty** | |
| | **Baseline** | **fTCR** | **pTCR** | **fTCR** | **pTCR** |
| **Mixed Logit Model Estimations** | | | | | |
| $VTT_{\text{fTCR}}$ (€/hour) | 14.82 | 16.99 | 16.99 | 17.06 | 17.06 |
| $VTT_{\text{pTCR}}$ (€/hour) | 14.82 | 12.74 | 12.74 | 12.75 | 12.75 |
| $VTT_{\text{nTCR}}$ (€/hour) | 14.82 | 12.53 | 12.53 | 12.55 | 12.55 |
| **Sample Enumeration on ODIN2023** | | | | | |
| ODIN weight $\overline{v_{\text{TCR}}}$ | | | 0.78 | | |
| ODIN weight $\overline{v_{\text{nTCR}}}$ | | | 0.22 | | |
| $\overline{VTT}$ (€/hour) | 9.24 | 10.66 | 8.29 | 10.69 | 8.28 |
| $\overline{VTT}$ increase | 0.00% | 15.38% | -10.28% | 15.65% | -10.40% |
| ***Ring Utrecht* SCBA** | | | | | |
| **RC* selecteren** | | | | | |
| **Yearly** | | | | | |
| TTS Car commute | | | 454k [hours] | | |
| VTTS car commute | €5.20M | €6.00M | €4.67M | €6.01M | €4.66M |
| | | 15.38% | -10.28% | 15.65% | -10.40% |
| VTTS total | €16.90M | €17.70M | €16.37M | €17.71M | €16.36M |
| | | 4.73% | -3.16% | 4.82% | -3.20% |
| Absolute Increase | €- | €0.80M | -€0.53M | €0.81M | -€0.54M |
| **Discounted perpetuity** | | | | | |
| VTTS total | €187.00M | €195.85M | €181.08M | €196.00M | €181.01M |
| | | 4.73% | -3.16% | 4.82% | -3.20% |
| All benefits − All costs | -€485.00M | -€476.15M | -€490.92M | -€475.99M | -€490.99M |
| | | 1.82% | -1.22% | 1.86% | -1.23% |
| Absolute Increase | €- | €8.85M | -€5.92M | €9.00M | -€5.99M |
| **GE* Selecteren Compact** | | | | | |
| **Yearly** | | | | | |
| TTS Car commute | | | 712k [hours] | | |
| VTTS car commute | €9.00M | €10.38M | €8.07M | €10.41M | €8.06M |
| | | 15.38% | -10.28% | 15.65% | -10.40% |
| VTTS total | €37.90M | €39.28M | €36.97M | €39.31M | €36.96M |
| | | 3.65% | -2.44% | 3.72% | -2.47% |
| Absolute Increase | €- | €1.38M | -€0.93M | €1.41M | -€0.94M |
| **Discounted perpetuity** | | | | | |
| VTTS total | €885.00M | €917.32M | €863.39M | €917.89M | €863.13M |
| | | 3.65% | -2.44% | 3.72% | -2.47% |
| All benefits − All costs | €589.00M | €621.32M | €567.39M | €621.89M | €567.13M |
| | | 5.49% | -3.67% | 5.58% | -3.71% |
| Absolute Increase | €- | €32.32M | -€21.61M | €32.89M | -€21.87M |

Bachaus (2023). A small difference was to be expected as Bachaus (2023) did not only cover the car as a mode of transportation.

Thereafter, the $\overline{VTT}$'s resulting from the sample enumeration procedure are presented. They are approximately €1.42/hour (+15%) higher and €0.95/hour (-10%) lower than the baseline $\overline{VTT}$ for models using the fTCR and the pTCR estimates respectively. There were negligible differences between the models including and excluding scale heterogeneity. Thus accounting for travel cost reimbursement does change the $\overline{VTT}$, also after applying representative weights for the Dutch population. However, we can observe that the impact on the $\overline{VTT}$ is heavily dependent on whether we adopt the fTCR or the pTCR estimates. The actual effect will be somewhere in between as a portion of the population receives a fTCR and another portion a pTCR.

Evaluating the commute $VTTS$ with the increases to the baseline sketches the impact on the *Ring Utrecht* SCBA. For instance, in the "RC Selecteren" case, the baseline value of travel time savings was €5.20M, which increased to €6.00M with the $VTT_{\text{fTCR}}$ estimate adopted and reduced to €4.67M with the $VTT_{\text{pTCR}}$ estimate adopted. Similarly, the "GE Compact" case showed an increase and decrease from €9.0M baseline to €10.38M and €8.07M. Which in turn reflects a value of $+$€32.32M and $-$€21.61M on the total travel time benefits of the *Ring Utrecht* project.

# 6

# Conclusion & Discussion

## Method Summary

The method applied in this study investigated the impact of full- and partial travel cost reimbursement (TCR) on the value of travel time ($VTT$) using both the NL2022 (Netherlands) and UK2014 (United Kingdom) datasets (ARUP, 2015; Kouwenhoven et al., 2022). We tested three hypotheses: whether TCR influenced the $VTT$, if respondents with TCR made less consistent choices, and if these respondents exhibited a higher willingness to accept (WTA) compared to those without TCR. Utilising Multinomial Logit ($MNL$) and Mixed Logit ($ML$) models, the analysis incorporated evaluating the interaction terms between TCR and $VTT$, while allowing for discrete scale heterogeneity. Furthermore, by considering the respondents' reference situation, we separated a TCR interaction with $VTT$ into willingness to accept (WTA) and willingness to pay (WTP) components. Lastly, we evaluated to what extent accounting for TCR impacted the *Ring Utrecht* case study.

## Evaluation Hypotheses

**When does a travel cost reimbursement impact the VTT?** Our 2-attribute results indicated that for a 2-attribute choice experiment commuter respondents with a full Travel Cost Reimbursement (fTCR) had a significantly higher $VTT$ compared to those without a TCR, supported by; (1) data from both the UK2014 and the NL2022 national value of travel time study, (2) MNL and Mixed Logit models, (3) including and excluding scale heterogeneity. This effect was significant and consistent for five out of six 2-attribute experiments. The likely cause for the insignificant effect of this one estimation was likely due to the lack of the number of respondents receiving a TCR. The Mixed Logit specifications reduced the statistical significance of the difference in $VTT$ between the fTCR and the nTCR groups, but this did not result in any of the abovementioned observations becoming statistically insignificant at a 5% significance level. This might be attributed to the Mixed Logit model's ability to account for unobserved heterogeneity in respondents' preferences. Unlike the Multinomial Logit (MNL) model, which only allows for observed heterogeneity in the VTT, the Mixed Logit model allows for inter-individual variations in the $VTT$. Thus some of the reimbursement effects captured in the interaction parameters for the MNL specificion might now have been captured in the distribution of the VTT.

Interestingly, as the number of attributes in the experiments increased the difference in $VTT$ between the nTCR group and the fTCR group was no longer significant. In practice, this is similar to what happens to interactions with socioeconomic variables, whenever the number of attributes in the experiment increases, these interactions typically also become less significant.

We thus found a difference in the $VTT$ between commuter respondents who receive a full travel cost reimbursement (fTCR) and those who do not receive a TCR in a 2-attribute discrete choice experiment. The difference ranged from 30% to 50% for the NL2022 data and from 80% to 140% for the UK2014 data. This supports our hypothesis that respondents who receive a TCR have a higher value of travel time.

The estimates for business respondents were generally not significant enough to support our hypothesis that respondents receiving a fTCR have a higher $VTT$. Only the 2-attribute estimates for the closely related Bus, Tram & Metro (BTM) and Light Rail Transit (LRT) modes showed significant and consistent differences in $VTT$ between respondents who receive a full travel cost reimbursement fTCR and those who do not receive a TCR. These effects were consistent for the MNL and Mixed Logit models both excluding and including scale heterogeneity. The differences in $VTT$ for BTM and LRT were between 65% & 70% for the NL2022 dataset and between 80% & 85% for the UK2014 dataset.

The interaction between the $VTT$ and partial Travel Cost Reimbursement (pTCR) was generally lower in magnitude and results were more often than not insignificant. Thus the observed difference in $VTT$ between respondents not receiving a TCR and respondents receiving a pTCR was lower as compared to the effects observed for respondents receiving a fTCR. The signs of the effects were generally in line with our hypothesis that travel cost reimbursement increases the $VTT$. However, we cannot formally prove a relationship between pTCR and the $VTT$ at a satisfactory significance level.

Thus to answer the first research question; a full travel cost reimbursement significantly increased the $VTT$ for commute respondents in a 2-attribute discrete choice experiment supported by data from both the Dutch national $VTT$ study and the UK national $VTT$ study. It did not significantly affect the $VTT$ for 3 or more attribute experiments or business respondents.

**Do respondents with a travel cost reimbursement make less consistent choices?** The direction of the scale interaction effects is as expected for respondents receiving a full Travel Cost Reimbursement (fTCR), their scale parameter estimate is generally lower compared to respondents not receiving a travel cost reimbursement. The signs of the scale parameter interactions for pTCR respondents were not as expected, 8/12 estimations yielded *higher* scale parameters for the pTCR group compared to the nTCR group. However, these abovementioned scale interactions were only significant in 5/24 estimations. In other words; we were unable to find convincing evidence which supports the notion that the scale parameter might be statistically different for respondents receiving any form of TCR. Thus we were unable to find convincing evidence to support our hypothesis that respondents who receive a travel cost reimbursement make less consistent choices.

**Does a travel cost reimbursement impact the WTP and the WTA differently?** We were able to find evidence which supports our third hypothesis. Namely, the interaction with the $VTT$ was more predominant in the WTA quadrant compared to the WTP quadrant. However, the model which took this into account did not perform statistically better in terms of model fit compared to a model which only allowed for equal TCR interaction on both the WTA and the WTP quadrant. This suggests that, although there may be a nuanced difference in how TCR influences WTA and WTP, this distinction might not be substantial enough to justify a more complex model that separately accounts for these interactions. Therefore, for practical purposes and model simplicity, treating the TCR effect as uniform across both WTA and WTP might be sufficient, without sacrificing the accuracy or predictive power of the model.

**To what extent do sign effects and choice consistency explain an effect of TCR on the VTT?** Our results generally indicated that taking scale differences into account between respondents receiving a TCR and not receiving a TCR did not significantly alter the $VTT$ estimates. This is probably because the scale differences were not too large between groups.

Only one estimation yielded greatly different $VTT$ estimates due to a scale decrease of approximately 62% for TCR observations as compared to nTCR observations. However, the confidence intervals of the $VTT$ taking scale heterogeneity into account and the $VTT$ when neglecting scale heterogeneity still overlapped. Thus stating that the scale for this estimation was indeed statistically different cannot be concluded. When we assumed the $VTT$ to be randomly distributed following a lognormal distribution, the scale difference was reduced, and simultaneously the $VTT$ difference was no longer apparent. We are unsure about the exact underlying mechanism, but it might be possible that the randomly distributed $VTT$ in the Mixed Logit specification captures effects which were previously captured in the scale parameter in the MNL specification.

Nonetheless, we did observe hints of what we refer to as a *dampening effect*. Thus whenever the scale for the TCR group was lower than for the nTCR group, incorporating scale heterogeneity slightly increased the $VTT$ difference between the two groups. This was in line with our theory, as a decrease in the scale of the overall utility might be misattributed to a reduction in the $VTT$.

**What is the impact of the findings on cost-benefit analysis?** We evaluated this research question by applying our findings to the *Ring Utrecht* societal cost-benefit analysis (SCBA) and assessed the extent to which accounting for TCR impacted the travel time reduction benefits of this project. Based on the conclusions of the other research questions, we considered TCR effects only for commuters. Given that the time benefits for the *Ring Utrecht* project are primarily associated with car transportation, we approximated the time benefits if TCR for car-using commuters had been incorporated in the calculation for national averages. When calculating these national averages we separately adopted the pTCR and the fTCR estimates as the ODIN2023 database used for these calculations did not include any information on the magnitude of the travel cost reimbursement. We also opted to evaluate the *Ring Utrecht* SCBA for the model which accounts for scale heterogeneity even though we concluded earlier that there are no significant scale effects. By applying this more sophisticated model, we aimed to confirm that the conclusions drawn from the SCBA are robust across different modelling approaches. We however found no substantial differences in the total travel time benefits when using the model which allows for scale heterogeneity.

We did find that adopting the models which account for travel cost reimbursement significantly affect the national $VTT$, with this value for car commuters increasing by approximately 15.38% to 15.65% when using the $VTT_{\text{fTCR}}$ estimate, and decreasing by 10.28% to 10.40% when using the $VTT_{\text{pTCR}}$ estimate. After mapping these changes onto the travel time savings for commuting traffic in the *Ring Utrecht* SCBA, the total travel time savings of the entire project either increased by up to 5.58% or decreased by up to 3.71%, corresponding to a monetary impact ranging from an increase of €32.32M to a decrease of €21.61M in the total travel time benefits of the project. These findings underscore the importance of considering travel cost reimbursement when calculating the value of travel time, as its impact on an SCBA can be substantial. Other SCBAs might be affected to an even greater extent whenever the share of commuters is larger.

## Discussion

We made a distinct contribution by focusing on the interaction between the $VTT$ and TCR within the context of stated preference experiments. While prior research, including studies by Baldassare et al. (1998), Bueno et al. (2017), and Evangelinos et al. (2018) investigated the effects of financial incentives and penalties on travel decisions, they did not specifically address how TCR impacts the $VTT$. Furthermore, our method of incorporating discrete scale heterogeneity to assess choice consistency provided a nuanced understanding of how TCR influences decision-making processes. In contrast to a large body of scientific literature, who conclude that marginal willingness to pay can be significantly distorted when not taking scale heterogeneity into account, (Alter et al., 2007; Deshazo, 2002; Fiebig et al., 2010; Oehlmann et al., 2017; Swait & Adamowicz, 2001; Zhang & Adamowicz, 2011) our results generally indicated that discrete scale differences between mutually exclusive and assumed homoskedastic subsets did not significantly alter the $VTT$. Significant effects on the $VTT$ were observed only in one case where the scale of the observed utility for respondents who received a travel cost reimbursement was reduced by as much as 62% compared to the nTCR group. The reason that other studies have found more sizable differences whenever allowing for scale heterogeneity might be because it is very easy to confound scale heterogeneity with other sources of correlation. Scale heterogeneity is in essence a special form of correlation where the parameters for all the attributes increase or decrease simultaneously in magnitude across subsets. This is exactly the reason why we only allowed for scale heterogeneity in 2-attribute experiments, as the only correlation is the correlation between the parameters of the 2-attributes. Otherwise, an increase in the correlation between for example the parameter for the number of transfers and the $VTT$ can be misattributed to an increase in the scale parameter.

Furthermore, a potential limitation in the interpretation of travel cost reimbursement among business respondents, particularly those using a car, should be considered. It is possible that some business respondents who indicated that they do not receive a travel cost reimbursement might actually be driving a company car. These respondents may have interpreted the term ”travel cost reimbursement” differently, potentially excluding the value of having a company car from their response. This could have led to an underestimation of the true impact of travel cost reimbursement on their VTT, as the benefit of a company car effectively reduces their out-of-pocket travel costs, even if they do not receive a direct monetary reimbursement. Such a mechanism could explain the reduced effects we observed for business respondents using the car. Future research could address this by clarifying the definition of travel cost reimbursement in surveys or by specifically asking about the use of company cars to ensure that all forms of cost compensation are adequately captured.

For future research, we suggest a methodological improvement regarding the evaluation of socioeconomic interactions. In this study, we added and evaluated each socioeconomic interaction step by step, assessing them individually before incorporating them into the model. In hindsight, it may have been more efficient to include all potential socioeconomic interactions simultaneously and then evaluate whether the TCR interactions remain significant.

On a last note; we had two options for the sample enumeration procedure; either enumerate over trips which were made by car for commuting purposes or over respondents who frequently commute by car. Travel cost reimbursement information is stored as a personal characteristic, thus not as a trip characteristic. Therefore enumerating based on persons would perhaps make sense. However, we do not know whether the TCR effect on the $VTT$ also applies to trips which the employer does not reimburse. We therefore opted to enumerate over the trips, thus assuming the reimbursement effects *only* apply to commuting trips.

## Methodological Implications

The car commute estimate for the UK2014 dataset was the only estimate which showed a sizeable change in $VTT$ whenever we allowed for scale heterogeneity between reimbursed and non-reimbursed respondents. This showed that it is worthwhile to check the homoscedasticity assumption for MNL models to ensure unbiased parameter estimates. However one must do so very carefully because it is very easy to misattribute taste heterogeneity to scale heterogeneity.

On another note, as we found significant differences in the $VTT$ for reimbursed and non-reimbursed respondents in the UK2014 dataset, future $VTT$ studies might want to incorporate this effect when calculating national averages. However, the UK National Travel Survey (NTS), the UK counterpart of the *ODIN2023* dataset does not include information on travel cost reimbursement. Therefore, if the British authorities wish to incorporate travel cost reimbursement in the calculation of national averages, they must (1) ensure that their NTS includes travel cost reimbursement data, and (2) integrate travel cost reimbursement into their models.

A similar but slightly different argument can be made for the *ODIN2023* database. This national travel survey does collect information on travel cost reimbursement and even asks respondents what type of reimbursement they receive, for example; fixed amount, distance-based or any other type. However the magnitude of the reimbursement is not documented, thus to what extent this reimbursement covers their trip expenses is not known. Conversely, in the choice experiments information on the magnitude of the reimbursement was collected, but information on the type of reimbursement was lacking. Thus for future $VTT$ studies, we recommend aligning the travel cost reimbursement questions with the questions asked for the *ODIN2023* survey. As for the *ODIN2023* survey itself, we recommend also asking respondents about what portion of their trip is reimbursed. We would advise making this a trip-specific characteristic as a reimbursement of a fixed amount means that the proportion of the trip which is reimbursement changes with the trip costs.

One last methodological implication is related to the type of instructions respondents are given, recall; in the NL2022 dataset commuting respondents were generally instructed to make decisions as if they have to pay for the travel costs themselves and in the UK2014 dataset commuting respondents were instructed to make choices taking into account who paid for their actual reference journey. As there are more differences between the studies other than the type of instructions we cannot claim that any difference in the extent to which TCR affects the $VTT$ can be attributed to the instruction. Nonetheless, our results do indicate that the $VTT$ was affected more radically in the UK2014 dataset than in the NL2022 dataset. Future $VTT$ studies could validate this by assigning for example half of the sample with a UK2014-type instruction and the other half of the sample with a NL2022-type instruction. Through such an approach one would be able to make more thorough claims on the impact of the instruction as the instruction would be the only variable which is varied. Any differences in the results aside; As we want to know what choices respondents would make in real life it is arguably better to follow the UK2014 approach, as in real life people presumably do take TCR into account when making choices. However, an approach like that could potentially lead to more lexicographic choice behaviour as more people could be prone to always choose the fastest alternative. As lexicographic choice behaviour does not provide any information on the time-cost tradeoff a respondent makes, common practice is to discard these observations from the sample. Therefore instructing respondents with a UK2014-type instruction could effectively reduce the number of usable observations and therefore reduce the statistical significance of the model estimates. But whether a TCR causes lexicographic choice behaviour was not the topic of our study, thus this is purely speculative.

## Policy Implications

We found that respondents with a TCR have a higher $VTT$ in 2-attribute choice experiments, particularly among commuters. If the goal is to ascertain what individuals are willing to pay out-of-pocket for travel time, it is more appropriate to base the $VTT$ on respondents who paid for their reference journey themselves.

However, when using the $VTT$ in national cost-benefit analysis or when determining network effects as a result of an infrastructural change we argue that travel cost reimbursement should be taken into account as in practice the national average $VTT$ is often based on 2-attribute experiments. Thus one should weigh the $VTT$ according to the number of people who receive a TCR in the general population. The reason is, that the utilisation of e.g. a new piece of road is also dependent on people who receive a travel cost reimbursement. In real life, these respondents are not instructed to ignore the fact that they receive a reimbursement, and their subsequent choices will thus be made under the influence of them receiving a travel cost reimbursement.

We evaluated such a cost-benefit analysis by approximating to what extent the SCBA of the *Ring Utrecht* project would change as a consequence of taking travel cost reimbursement into account. We found substantial differences in this specific case where the total travel time benefits could differ as much as 3.72%, and we found a heavy dependence on whether the estimates for a partial- or a full Travel Cost Reimbursement were adopted. This further strengthens the argument that future cost-benefit analysis should be based on models which incorporate travel cost reimbursement effects on the $VTT$ as the results could be even more dramatic for other projects due to; (1) the ratio of commuters to other road users changing, (2) the mode of transportation not being limited to cars.

When further zooming out, we observed in our samples that the number of commuter respondents who received a travel cost reimbursement is substantially higher in the NL2022 dataset compared to the UK2014 dataset (73% versus 19% respectively for Car Commute). We argue that this difference in the number of reimbursement respondents across both datasets is probably because there is no tax-free travel cost reimbursement in the UK as we were unable to find any other reasonable explanations. Uncovering to what extent a tax-free travel cost reimbursement impacts the number of employers who receive a TCR was naturally not the goal of either the Dutch (Kouwenhoven et al., 2022) or the UK (ARUP, 2015) study but it does provide an interesting perspective. Since the $VTT$ is significantly higher for reimbursed respondents, increasing the number of people in the population who receive a travel cost reimbursement also increases the average $VTT$ in the population. Which in turn tips the scale of cost-benefit analysis more often to the benefit side as travel time savings are valued more. This subsequently increases the overall efficiency of a transport network as for more projects the benefits outweigh the costs. This is an interesting policy effect as a consequence of the number of people who receive a travel cost reimbursement in the population, which is likely determined by there being any tax benefits.

One another note; as TCR seems to affect the $VTT$ predominantly for commuters, evaluating a SCBA in which the share of commuters is larger than that of the *Ring Utrecht* SCBA will make the impact of TCR even greater. A typical policy where the ratio of commuters in proportion to the total traffic flow is very high is during rush hour. Thus an SCBA to evaluate a policy aimed at for example reducing traffic flow during rush hour can be affected to a great extent. A controversial policy idea focussed on exactly that is that of congestion charging or ”rekeningrijden” in Dutch. It aims to reduce the number of vehicles in congested areas or congested periods by imposing time or location-based fees on drivers. As we found that TCR significantly affects the VTT for commuters the perceived benefits of congestion charging (e.g., reduced travel times) might be heavily undervalued if TCR is not considered. Policymakers should therefore adjust weighted $VTT$'s in such an SCBA to reflect the proportion of reimbursed commuters, ensuring that the benefits of congestion charging are accurately assessed. When we assess the congestion charging policy from a more broader perspective we can uncover more mechanisms where our results are relevant. If a substantial portion of drivers are reimbursed for their travel costs, their higher VTT might make them less sensitive to congestion charges, as they perceive their time as more valuable and may be willing to pay the charge. Thus we might see less reduction in traffic volumes from congestion charging than anticipated as the effectiveness of the policy could be moderated by the extent to which commuters receive TCR. Besides, employers might opt to also reimburse the congestion charge further reducing the effectiveness of the policy. Policymakers could consider varying congestion charges based on whether or not drivers are reimbursed for their travel costs or implement complementary measures to ensure that commuters with a travel cost reimbursement also ”feel” the financial stimulus to change their travel behaviour. Our findings also suggest a potential approach to ensuring fairness in congestion charging. By offering a form of travel cost reimbursement from the government to low-income road users, we could increase their $VTT$ to align more closely with that of high-income road users. This adjustment would effectively level the playing field, ensuring that all individuals experience the congestion charging stimulus on equal terms, thereby enhancing the policy’s equity.

## Limitations

One limitation of the study was the definition of reimbursement categories, where the percentages were chosen through reasoning semantically about what percentage of reimbursement can be defined as a full Travel Cost Reimbursement. To recall; respondents with a reimbursement of [0%, 15%] were set as respondents receiving no travel cost reimbursement, (15%, 85%) were set as respondents receiving a partial travel cost reimbursement and [85%, 100%] was defined as a full travel cost reimbursement. It would be interesting to explore how the results might have changed as a function of these definitions, potentially providing more nuanced insights. One such alteration was briefly evaluated where a comparison was made between the bin as defined in this study and the bins defined in a more trivial manner, where any fraction of reimbursement other than 0% or 100% is denoted as a partial travel cost reimbursement. We observed that the estimated interactions did not differ dramatically. Nonetheless, a robustness analysis which evaluates the reimbursement effect on the $VTT$ whilst varying the reimbursement categories would be a valuable contribution.

Furthermore, as our research question was about the difference in $VTT$ between nTCR and TCR respondents, we focused on the relative difference in $VTT$ between the two groups. However, we later observed that when allowing for scale heterogeneity the reference $VTT$, thus the $VTT$ for the nTCR group sometimes also changed slightly. This effect was implicitly taken into account for the sample enumeration procedure as the estimated $VTT$ values themselves are used and not solely the difference thus it will not have affected the results for the *Ring Utrecht* SCBA. Nonetheless, it would be an interesting avenue for further analysis from a methodological viewpoint.

Moreover, as mentioned earlier ODIN2023 does not provide information on the magnitude of the travel cost reimbursement. Thus for the evaluation of the textitRing Utrecht project, we used the parameter estimates for a full Travel Cost Reimbursement and applied the corresponding $VTT$ values to the respondents who reported receiving *any* travel cost reimbursement in ODIN2023. Whenever the yearly ODIN survey also reports on the magnitude of the reimbursement it will be possible to distinguish between a partial and a full travel cost reimbursement which might alter the impact on the *Ring Utrecht* SCBA.

## Further Research

This study opens several avenues for further research, all aimed at deepening our understanding of how travel cost reimbursement (TCR) influences the Value of Travel Time (VTT) across different contexts and populations.

An intriguing yet unanswered question is whether a travel cost reimbursement increases the $VTT$ to the same extent for low-income individuals as it does for high-income individuals. This could be studied through a similar analysis as was performed in this study, only now allowing for a heterogeneous interaction across income groups. This also leads to the methodological implication listed earlier; namely that for future national $VTT$ surveys it could be beneficial to retrieve more specific information on what type of travel cost reimbursement respondents receive. One can imagine that reimbursement effects on the $VTT$ might be very similar for low- and high-income respondents whenever their travel costs are *directly* and in full paid by the employer. However, when people first have to pay for the travel cost themselves and are only thereafter compensated a fixed amount, a financial incentive to make cheaper choices remains, and a different interaction for low- and high-income groups might show.

Another avenue for further research would be to ensure whether the observed difference in the $VTT$ is a reflection of true preferences and not limited to a stated preference study, we suggest performing a revealed preference study with the same goal as this study. This might even be possible on existing data, e.g. by exploring smart-card data. However, we expect that finding a significant interaction between the $VTT$ and travel cost reimbursement will be very difficult if not impossible due to multicollinearity. In other words; the number of occurrences in real life where one can choose between a slower but cheaper route is very limited.

Furthermore, another avenue could be to confirm whether partial travel cost reimbursement indeed affects the $VTT$ to a lesser extent than a full travel cost reimbursement. To do so in a future study it would be very valuable to also record the *type* of reimbursement respondents receive to ensure the difference between a partial- and a full travel cost reimbursement is not due to another confounding variable; the type of travel cost reimbursement.

Additionally, further research could be to ask respondents whether they usually receive a travel cost reimbursement for their day-to-day commute but ask them to keep a trip in mind which is *not* reimbursed. Thus asking respondents with a reference journey which is not reimbursed if they receive a TCR for commuting. This approach would enable us to evaluate whether *usually* receiving a TCR affects the $VTT$ even if the respondent thinks of another reference journey. In other words; this would enable us to evaluate whether a travel cost reimbursement changes a person's $VTT$ for all trips or if a person's $VTT$ is only affected for the commute trip to which the reimbursement applies. Because even if respondents are thinking of a non-reimbursed trip, the habit of receiving reimbursement could influence their perception of costs and benefits, potentially altering their VTT.

An additional evaluation could consider whether a reference-based design impacts the extent to which travel cost reimbursement effects occur. [1] Specifically, it is plausible that a choice set which is based on the reference (reimbursed) trip influences the extent to which individuals with travel cost reimbursement take their reimbursement into account as such a choice set more closely resembles their reimbursed reference trip.

Another interesting avenue for further research would be to investigate whether respondents with a lower-scale parameter are also respondents who make faster choices. Put simply; does scale correlate with response time? Such a relation was already found by Chen et al. (2016) for externally imposed response time limitations but to our knowledge not yet for response time without externally imposed limits. If a strong and consistent relation is found this could be a nice check for researchers when assuming homoscedasticity of their data. A strong correlation between the two would also be in line with our expectation that task motivation might be one of the causes of scale differences as we would expect lower motivated respondents to make choices more quickly. Furthermore, following the same arguments, respondents with a TCR could also show more non-trading behaviour. Non-trading behaviour in this case would mean that respondents always choose the alternative with the lowest travel time.

One last potential avenue for further research is the exploration of the interaction parameters between reimbursement bins and the $VTT$ by assigning them a random distribution and estimating a mixed Logit Model. This approach may allow for the analysis of heterogeneity within these interaction parameters since it might be possible that some respondents assume that their reimbursement may not apply to the tradeoffs presented and therefore do not show an increased $VTT$. However, applying this procedure is guaranteed to result in a higher model fit, as a more flexible utility specification is set. Thus evaluating whether this method yields greater predictive power will be difficult.

[1] A reference-based design in short, is where the reference cost- and time values of the respondent are included in either one or both of the alternatives. Thus there is always some value in the choice set which is the same as the reference journey which was described by the respondent.

# References


Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*, 569–576. https://doi.org/10.1037/0096-3445.136.4.569

ARUP. (2015, August). *Provision of market research for value of travel time savings and reliability: Phase 2 report* (tech. rep. No. Issue 14 August 2015) (In partnership with Ove Arup & Partners Ltd, Institute for Transport Studies (ITS), University of Leeds, and Accent. Contributors: Stefan Sanders, Efie Drivyla, Jeremy Bourdon, Richard Batley, Manuel Ojeda Cabral, Phani Kumar Chintakayala, Charisma Choudhury, Andrew Daly, Thijs Dekker, Anthony Fowkes, Stephane Hess, Daniel Johnson, James Laird, Peter Mackie, Mark Wardman, John Bates, Michiel Bliemer, Maria Börjesson, John Parkin, Tom Worsley, Chris Heywood, Miranda Mayes, Rachel Risely, Nancy Curzon, Rob Sheldon.). Department for Transport. London, United Kingdom.

Bachaus, A. (2023). Evaluatie onbelaste reiskostenvergoeding. www.ce.nl

Baldassare, M., Ryan, S., & Katz, C. (1998). Suburban attitudes toward policies aimed at reducing solo driving.

Belastingdienst. (2024). U reist naar uw werk [Accessed: 2024-06-07]. https://www.belastingdienst.nl/wps/wcm/connect/bldcontentnl/belastingdienst/prive/auto_en_vervoer/u_reist_naar_uw_werk/

Bradly, M., & Daly, A. (1996). Estimation of logit choice models using mixed stated-preference and revealed-preference information see profile. https://www.researchgate.net/publication/230663774

Brosnan, K., Babakhani, N., & Dolnicar, S. (2019). “i know what you’re going to ask me” why respondents don’t read survey questions. *International Journal of Market Research*, *61*(4), 366–379. https://doi.org/10.1177/1470785318821025

Bueno, P. C., Gomez, J., Peters, J. R., & Vassallo, J. M. (2017). Understanding the effects of transit benefits on employees’ travel behavior: Evidence from the new york-new jersey region. *Transportation Research Part A: Policy and Practice*, *99*, 1–13. https://doi.org/10.1016/j.tra.2017.02.009

CBS. (2024). *ODiN 2023*. https://doi.org/10.17026/SS/FNXJEU

Chen, C., Chorus, C., Molin, E., & van Wee, B. (2016). Effects of task complexity and time pressure on activity-travel choices: Heteroscedastic logit model and activity-travel simulator experiment. *Transportation*, *43*(3), 455–472. https://doi.org/10.1007/s11116-015-9584-3

Dellaert, B. G. C., Brazell, J. D., & Louviere, J. J. (1999). The effect of attribute variation on consumer choice consistency.

Deshazo, J. R. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, *44*. https://doi.org/10.1006rjeem.2001.1199

Ding, M., Grewal, R., & Liechty, J. (2005). Incentive-aligned conjoint analysis. *Journal of Marketing Research*, *42*, 67–82. https://doi.org/10.1509/jmkr.42.1.67.56890

Dupuit, J. (1844). De la mesure de l’utilité des travaux publics [Translated by R.H. Barback as ”On the Measurement of the Utility of Public Works”, in International Economic Papers, 1952, no. 2, pp. 83–110]. *Annales des ponts et chaussées*, *8*(2), 332–375.

Evangelinos, C., Tscharaktschiew, S., Marcucci, E., & Gatta, V. (2018). Pricing workplace parking via cash-out: Effects on modal choice and implications for transport policy. *Transportation Research Part A: Policy and Practice*, *113*, 369–380. https://doi.org/10.1016/j.tra.2018.04.025

Farahmand, Z. H., Gkiotsalitis, K., & Geurs, K. T. (2021). Mobility-as-a-service as a transport demand management tool: A case study among employees in the netherlands. *Case Studies on Transport Policy*, *9*, 1615–1629. https://doi.org/10.1016/j.cstp.2021.09.001

Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2010). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *29*, 393–421. https://doi.org/10.1287/mksc.l090.0508

GOV.UK. (2024). Tax on company benefits: Tax on company cars [Accessed: 2024-08-07].

Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, *2*(1), 84–90. https://doi.org/10.1007/BF01386213

Hensher, D. A. (2006). Towards a practical method to establish comparable values of travel time savings from stated choice experiments with differing design dimensions. *Transportation Research Part A: Policy and Practice*, *40*, 829–840. https://doi.org/10.1016/j.tra.2006.02.001

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

Hess, S., & Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, *32*. https://doi.org/10.1016/j.jocm.2019.100170

Hess, S., & Rose, J. M. (2012). Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation*, *39*(6), 1225–1239. https://doi.org/10.1007/s11116-012-9394-9

Hess, S., Rose, J. M., & Hensher, D. A. (2007). Asymmetric preference formation in willingness to pay estimates in discrete choice models. *Transportation Research Part E: Logistics and Transportation Review*, *44*, 847–863. https://doi.org/10.1016/j.tre.2007.06.002

Hess, S., & Train, K. (2017a). Correlation and scale in mixed logit models. *Journal of Choice Modelling*, *23*, 1–8. https://doi.org/10.1016/j.jocm.2017.03.001

Hess, S., & Train, K. (2017b). Correlation and scale in mixed logit models. *Journal of Choice Modelling*, *23*, 1–8. https://doi.org/10.1016/j.jocm.2017.03.001

HM Revenue Customs. (2024). Ordinary commuting and private travel (490, chapter 3) [Retrieved July 3, 2024]. https://www.gov.uk/guidance/ordinary-commuting-and-private-travel-490-chapter-3

IenW. (2016, August). *Mkba ring utrecht* (tech. rep.) (Accessed: 2024-07-18). Ministerie van Infrastructuur en Milieu. https://www.rwseconomie.nl/onderwerpen/veel-gestelde-vragen/documenten/rapporten/2016/augustus/augustus/mkba-ring-utrecht

ITF. (2019). What is the value of saving travel time? *ITF Roundtable Reports*, *176*.

Jang, S., Rasouli, S., & Timmermans, H. (2022). The effect of task complexity on stated choice processes: The moderating role of cognitive ability. *Transportation Research Part E: Logistics and Transportation Review*, *163*. https://doi.org/10.1016/j.tre.2022.102744

Kaa, J. V. D. (2010). Sign-dependent value of time in stated preference: Judgment bias or exposure of genuine preference? *EJTIR Issue*, *10*, 347–367. www.ejtir.tbm.tudelft.nl

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus; Giroux.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy*, *98*(6), 1325–1348. https://doi.org/10.1086/261737

Kouwenhoven, M., Muller, J., Thoen, S., Willigers, J., & de Jong, G. (2022). Values of time, reliability and comfort in the netherlands 2022 new values for passenger travel and freight transport.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. https://doi.org/10.2307/2280779

Liebe, U., Meyerhoff, J., Kroesen, M., Chorus, C., & Glenk, K. (2018). From welcome culture to welcome limits? uncovering preference changes over time for sheltering refugees in germany. *PloS one*, *13*(8), e0199923. https://doi.org/10.1371/journal.pone.0199923

Louviere, J. J., & Hensher, D. (1983). Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event. *Journal of Consumer Research*, *10*(3), 348–361. https://doi.org/10.1086/208974

Louviere, J. J., & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. *Journal of Marketing Research*, *20*(4), 350. https://doi.org/10.2307/3151440

McFadden, D., & Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, *15*(5), 447–470.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.

Ministerie van Infrastructuur en Waterstaat. (2022). Staatscourant 2022, 22046: Tracébesluit a27/a12 ring utrecht 2022 [Official decision by the Minister of Infrastructure and Water Management, Netherlands. Accessed: 2024-07-22]. https://zoek.officielebekendmakingen.nl/stcrt-2022-22046.html

Munizaga, M. A., Heydecker, B. G., Dios, J. D., & Uzar, O. (2000). Representation of heteroskedasticity in discrete choice models. www.elsevier.com/locate/trb

Oehlmann, M., Meyerhoff, J., Mariel, P., & Weller, P. (2017). Uncovering context-induced status quo effects in choice experiments. *Journal of Environmental Economics and Management*, *81*, 59–73. https://doi.org/10.1016/j.jeem.2016.09.002

PBL & CPB. (2020). *Nederland in 2030 en 2050: Twee referentiescenario's-wlo* (Report) (https://www.pbl.nl/sites/default/files/downloads/PBL-CPB-2020-Nederland-in-2030-en-2050-twee-referentiescenario's-WLO.pdf). PBL Netherlands Environmental Assessment Agency and CPB Netherlands Bureau for Economic Policy Analysis. https://www.pbl.nl/sites/default/files/downloads/PBL-CPB-2020-Nederland-in-2030-en-2050-twee-referentiescenario's-WLO.pdf

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157–175. https://doi.org/10.1080/14786440009463897

Raux, C., Croissant, Y., & Pons, D. (2015). Would personal carbon trading reduce travel emissions more effectively than a carbon tax? *Transportation Research Part D: Transport and Environment*, *35*, 72–83. https://doi.org/10.1016/j.trd.2014.11.008

Romijn, G., & Renes, G. ( (2013). *Algemene leidraad voor maatschappelijke kosten-batenanalyse*.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behavior. *Economica*, *5*(17), 61–71. https://doi.org/10.2307/2548836

Significance, Amsterdam, V. U., Services, J. B., TNO, NEA, NIPO, T., & PanelClix. (2013). *Values of time and reliability in passenger and freight transport in the netherlands* (tech. rep.) ((Revised final version)). Ministry of Infrastructure and the Environment. Den Haag.

Swait, J., & Adamowicz, W. (2001). Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes*, *86*, 141–167. https://doi.org/10.1006/obhd.2000.2941

Thaler, R. H. (1999). Mental accounting matters.

Train, K., & Weeks, M. (2015). In preference space and willingness-to-pay space. In *The economics of non-market goods and resources* (Vol. 6). Springer. https://doi.org/10.1007/1-4020-3684-1_1

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. https://about.jstor.org/terms

Zawojska, E., Bartczak, A., & Czajkowski, M. (2019). Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. *Journal of Environmental Economics and Management*, *93*, 63–84. https://doi.org/10.1016/j.jeem.2018.11.007

Zhang, J., & Adamowicz, W. L. (2011). Unraveling the choice format effect: A context-dependent random utility model.

# A

# Utility Function

This is the utility function which was used to estimate a combined model for SP1A and SP2A. However, for the SP1A model specification interactions which are only relevant for the SP2A choice tasks were constrained to zero.

```
V[["LEFT"]] = (
    (
        mu
        * (1 + sc_cost_same_mu * DUMMY_COST_SAME)  # Interaction of mu with the same cost dummy variable
        * (1 + sc_sp2_mu * DUMMY_SP2A)  # Interaction of mu with the sp2 dummy variable
        * (1 + b_zelfBetalen_mu * EMPPAYBINS_Low
           + b_geheelVergoed_mu * EMPPAYBINS_High
           + b_gedeeltelijkVergoed_mu * EMPPAYBINS_Medium)  # EMPPAY interaction with mu based on employment bins
        * ((BASETIME / BaseTime0) ** lmda_t_mu
           * (BASECOST / BaseCost0) ** lmda_c_mu)  # Basetime and basecost interaction with mu
    )
    * (
        COST_L + (
            (vtt * TIME_MIN_L / 60
             + vttr * SDEV_ACTUAL_L / 60 * DUMMY_SP2A)  # Value of time and standard deviation with sp2 dummy variable
            * (1 + vtt_employment_1 * PAID_EMPLOYMENT_1
               + vtt_employment_2 * PAID_EMPLOYMENT_2)  # VTT interaction with employment types
            * (1 + vtt_zzp * ZZP)  # VTT interaction with self-employment
            * (1 + vtt_age_1 * AGE_CLASS_1
               + vtt_age_2 * AGE_CLASS_2
               + vtt_age_3 * AGE_CLASS_3
               + vtt_age_4 * AGE_CLASS_4)  # VTT interaction with age classes
            * (1 + vtt_panel * RECRUITMENTTYPE_1
               + vtt_intercept * RECRUITMENTTYPE_2)  # VTT interaction with recruitment types
            * (1 + vtt_edu_1 * EDU_1
               + vtt_edu_2 * EDU_2
               + vtt_edu_3 * EDU_3
               + vtt_edu_4 * EDU_4
               + vtt_edu_5 * EDU_5)  # VTT interaction with education levels
            * (1 + vtt_male * GENDER_1
               + vtt_female * GENDER_2)  # VTT interaction with gender
            * (1 + vtt_freq_1 * FREQ_1
               + vtt_freq_2 * FREQ_2
               + vtt_freq_3 * FREQ_3
               + vtt_freq_4 * FREQ_4
               + vtt_freq_5 * FREQ_5
               + vtt_freq_6 * FREQ_6)  # VTT interaction with frequency
            * (1 + sp2a_vtt * DUMMY_SP2A)  # VTT interaction with sp2 dummy variable
            * (1 + b_zelfBetalen_vtt * EMPPAYBINS_Low
               + b_geheelVergoed_vtt * EMPPAYBINS_High
```

```
                + b_gedeeltelijkVergoed_vtt * EMPPAYBINS_Medium)  # EMPPAY interaction with
                    VTT based on employment bins
             * (BASETIME / BaseTime0) ** lmda_t_vtt
             * (BASECOST / BaseCost0) ** lmda_c_vtt  # Basetime and basecost interaction with
                 VTT
             * (1 + vtt_inc1 * INC_1
                + vtt_inc2 * INC_2
                + vtt_inc3 * INC_3
                + vtt_inc4 * INC_4
                + vtt_inc5 * INC_5
                + vtt_inc6 * INC_6)  # Income interaction with VTT
        )
    )
)
```

# B Correlation table

| Variable | % TCR | TCR group |
|---|---|---|
| RESP_ID | -0.03 | 0.04 |
| question | -0.00 | -0.00 |
| SP | -0.01 | -0.00 |
| Block | -0.00 | 0.00 |
| ATTR_ORDER | -0.01 | 0.00 |
| ALTORDERCHANGE | 0.01 | -0.01 |
| QUESTIONX | 0.00 | -0.00 |
| CHOICE | -0.00 | 0.01 |
| COST_L | 0.16 | -0.15 |
| COST_R | 0.16 | -0.15 |
| TIME_MIN_L | 0.16 | -0.12 |
| TIME_MIN_R | 0.16 | -0.12 |
| USEDESTR_L | -0.01 | 0.02 |
| USEDESTR_R | 0.00 | 0.00 |
| REL1_L | 0.16 | -0.13 |
| REL2_L | 0.17 | -0.13 |
| REL3_L | 0.17 | -0.13 |
| REL4_L | 0.17 | -0.14 |
| REL5_L | 0.17 | -0.14 |
| REL1_R | 0.16 | -0.13 |
| REL2_R | 0.17 | -0.13 |
| REL3_R | 0.17 | -0.14 |
| REL4_R | 0.17 | -0.14 |
| REL5_R | 0.17 | -0.14 |
| SDEV_DESIGN_L | 0.11 | -0.09 |
| SDEV_DESIGN_R | 0.11 | -0.09 |
| SDEV_ACTUAL_L | 0.12 | -0.10 |
| SDEV_ACTUAL_R | 0.12 | -0.10 |
| IS_REFBASEID | 0.00 | -0.01 |
| QUADRANT | 0.00 | -0.00 |
| RECRUITMENTTYPE | -0.03 | 0.04 |
| EXP2 | 0.02 | 0.01 |
| CURRENTMODE | 0.05 | -0.00 |
| PURPOSE | 0.08 | -0.06 |
| USEHENSHER | -0.01 | 0.02 |
| PAID_EMPLOYMENT | -0.44 | 0.46 |

| Variable | % TCR | TCR group |
|---|---|---|
| DECISION_BY_EMPLOYER | -0.17 | 0.17 |
| BASECOST | 0.17 | -0.16 |
| BASETIME | 0.16 | -0.13 |
| GENDER | -0.06 | 0.03 |
| AGE_CLASS | -0.06 | 0.08 |
| EDU_CLASS | 0.01 | -0.01 |
| INCOME_CLASS | 0.01 | -0.01 |
| WORK_SITUATION_CLASS | -0.05 | 0.07 |
| ZZP | -0.19 | 0.20 |
| HHSIZE | -0.03 | -0.03 |
| GROUPSIZE_CAR | 0.06 | 0.11 |
| GROUPSIZE_OTHER | -0.02 | -0.02 |
| FREQ_0 | -0.01 | 0.04 |
| FREQ_A | 0.00 | -0.04 |
| ACTIVITY_ORIG | -0.01 | 0.02 |
| ACTIVITY_DEST | -0.06 | 0.08 |
| AIR_ACCESS_MODE | -1.00 | 1.00 |
| AIR_EGRESS_MODE | 1.00 | -1.00 |
| PTMODES_1 | 0.17 | -0.15 |
| PTMODES_2 | 0.09 | -0.07 |
| PTMODES_3 | 0.03 | -0.02 |
| PTMODES_4 | -0.10 | 0.09 |
| PTMODES_5 | -0.08 | 0.07 |
| PTMODES_6 | -0.03 | 0.04 |
| PTMODES_7 | -0.05 | 0.03 |
| PTMODES_8 | -0.05 | 0.05 |
| PTMODES_9 | 0.06 | -0.06 |
| PTMODES_10 | 0.03 | -0.04 |
| WHEN_AUTO_OV_ACTIVE | -0.01 | 0.05 |
| FF_TRAVTIME | 0.17 | -0.13 |
| TRAVTIME | 0.16 | -0.13 |
| DISTANCE | 0.17 | -0.16 |
| SPEED | 0.10 | -0.12 |
| NTRANSFERS | 0.07 | -0.05 |
| EMPPAY | -0.93 | 1.00 |
| EMPPAY_PERC_DEF | 1.00 | -0.93 |

# C
# Estimates SCBA Ring Utrecht

*Table on the next page*

**Table C.1:** Mixed-Logit estimates used in the *Ring Utrecht* SCBA. The increase in LogLikelihood, DoF, LRS, and Better model at $\alpha = 1\%$ are comparisons to the model to its neighbor on the **left**. The parameters denoted by either $\delta_{\text{sbscrpt}}$ or $\mu_{\text{sbscrpt}}$ are interaction parameters with the $VTT$ and the scale parameter $\mu$. They should be interpreted as relative changes irrespective of the other interaction parameters. Thus, $\delta_{\text{female}} = 0.18$ denotes that females have a $VTT$ which is 18% higher than those of males ceteris paribus. The Obs column lists the number of observations for which the corresponding dummy variable is equal to one. For example, the sample contained 3408 observations from males and 2432 from females.

| | | Not accounting for TCR | | TCR on VTT only | | TCR on VTT + scale | |
|---|---|---|---|---|---|---|---|
| Number of individuals | | 730 | | 730 | | 730 | |
| Observations | | 5840 | | 5840 | | 5840 | |
| Estimated parameters | | 27 | | 29 | | 31 | |
| Estimation time (hh:mm:ss) | | 00:32:20 | | 00:34:59 | | 00:28:51 | |
| Iterations | | 32 | | 31 | | 30 | |
| LL(final) | | -2304.17 | | -2299.72 | | -2296.76 | |
| Adj.Rho-square (0) | | 0.42 | | 0.42 | | 0.43 | |
| Adj.Rho-square (C) | | 0.42 | | 0.42 | | 0.43 | |
| BIC | | 4842.51 | | 4850.94 | | 4862.37 | |
| Increase in LL | | - | | 4.45 | | 2.96 | |
| Increase in DoF | | - | | 2 | | 2 | |
| LRS | | - | | 8.90 | | 5.92 | |
| Better model at $\alpha = 1\%$ | | - | | No | | No | |
| Better model at $\alpha = 5\%$ | | - | | Yes | | No | |
| Better model at $\alpha = 10\%$ | | - | | Yes | | Yes | |
| Parameter | Obs | estimate | t-ratio(0) | estimate | t-ratio(0) | estimate | t-ratio(0) |
| $\mu$ | | -1.37 | -17.13 | -1.37 | -17.13 | -1.25 | -14.00 |
| $\mu_{\text{fTCR}}$ | 1552 | 0.00 | NA | 0.00 | NA | 0.00 | NA |
| $\mu_{\text{pTCR}}$ | 2168 | 0.00 | NA | 0.00 | NA | 0.22 | 2.19 |
| $\mu_{\text{nTCR}}$ | 2120 | 0.00 | NA | 0.00 | NA | 0.12 | 1.20 |
| $\lambda_{T\mu}$ | | -1.03 | -11.37 | -1.03 | -11.40 | -1.01 | -11.07 |
| $\lambda_{C\mu}$ | | -0.11 | -1.79 | -0.11 | -1.80 | -0.12 | -1.92 |
| $VTT$ (mean lognormal dist.) | | 14.82 | | 16.99 | | 17.06 | |
| $VTT_{\mu}$ | | 2.16 | 12.90 | 2.30 | 13.27 | 2.30 | 13.28 |
| $VTT_{\sigma}$ | | 1.04 | 25.08 | 1.03 | 24.91 | 1.03 | 25.00 |
| $\delta_{\text{fTCR}}$ | 1552 | 0.00 | NA | 0.00 | NA | 0.00 | NA |
| $\delta_{\text{pTCR}}$ | 2168 | 0.00 | NA | -0.25 | -2.99 | -0.25 | -3.05 |
| $\delta_{\text{nTCR}}$ | 2120 | 0.00 | NA | -0.26 | -2.85 | -0.26 | -2.87 |
| $\lambda_T$ | | 0.28 | 2.18 | 0.25 | 1.93 | 0.25 | 1.94 |
| $\lambda_C$ | | 0.46 | 5.21 | 0.47 | 5.25 | 0.47 | 5.27 |
| $\delta_{\text{inc,1}}$ | 848 | -0.19 | -1.40 | -0.18 | -1.27 | -0.18 | -1.25 |
| $\delta_{\text{inc,2}}$ | 1496 | -0.18 | -1.58 | -0.15 | -1.28 | -0.15 | -1.24 |
| $\delta_{\text{inc,3}}$ | 744 | -0.36 | -3.24 | -0.35 | -3.15 | -0.35 | -3.15 |
| $\delta_{\text{inc,4}}$ | 1288 | 0.00 | NA | 0.00 | NA | 0.00 | NA |
| $\delta_{\text{inc,5}}$ | 472 | -0.01 | -0.06 | -0.02 | -0.12 | -0.03 | -0.13 |
| $\delta_{\text{inc,6}}$ | 992 | -0.23 | -1.90 | -0.23 | -1.85 | -0.23 | -1.86 |
| $\delta_{\text{male}}$ | 3408 | 0.00 | NA | 0.00 | NA | 0.00 | NA |
| $\delta_{\text{female}}$ | 2432 | 0.01 | 0.07 | 0.05 | 0.45 | 0.05 | 0.45 |
| $\delta_{\text{freq,1}}$ | 216 | 0.10 | 0.34 | 0.06 | 0.21 | 0.06 | 0.21 |
| $\delta_{\text{freq,2}}$ | 480 | -0.27 | -1.90 | -0.29 | -2.04 | -0.29 | -2.05 |
| $\delta_{\text{freq,3}}$ | 688 | -0.01 | -0.06 | -0.01 | -0.05 | -0.00 | -0.02 |
| $\delta_{\text{freq,4}}$ | 1368 | 0.00 | NA | 0.00 | NA | 0.00 | NA |
| $\delta_{\text{freq,5}}$ | 1816 | -0.13 | -1.08 | -0.12 | -1.01 | -0.12 | -1.00 |
| $\delta_{\text{freq,6}}$ | 1272 | 0.28 | 1.47 | 0.28 | 1.45 | 0.28 | 1.45 |
| $\delta_{\text{edu,1}}$ | 680 | -0.18 | -1.26 | -0.20 | -1.45 | -0.20 | -1.44 |
| $\delta_{\text{edu,2}}$ | 1792 | -0.18 | -1.91 | -0.21 | -2.28 | -0.21 | -2.31 |
| $\delta_{\text{edu,3}}$ | 712 | -0.09 | -0.66 | -0.10 | -0.75 | -0.10 | -0.74 |
| $\delta_{\text{edu,4}}$ | 2624 | 0.00 | NA | 0.00 | NA | 0.00 | NA |
| $\delta_{\text{edu,5}}$ | 32 | 0.45 | 0.52 | 0.59 | 0.62 | 0.59 | 0.62 |
| $\delta_{\text{panel}}$ | 5120 | 0.00 | NA | 0.00 | NA | 0.00 | NA |
| $\delta_{\text{intercept}}$ | 720 | 0.27 | 1.46 | 0.31 | 1.58 | 0.30 | 1.57 |
| $\delta_{\text{age,1}}$ | 1752 | 0.01 | 0.10 | 0.03 | 0.22 | 0.03 | 0.21 |
| $\delta_{\text{age,2}}$ | 2120 | 0.00 | NA | 0.00 | NA | 0.00 | NA |
| $\delta_{\text{age,3}}$ | 1848 | -0.14 | -1.33 | -0.12 | -1.16 | -0.12 | -1.18 |
| $\delta_{\text{age,4}}$ | 120 | -0.21 | -0.93 | -0.16 | -0.71 | -0.17 | -0.74 |
| $\delta_{\text{zzp}}$ | 288 | 0.41 | 1.38 | 0.43 | 1.40 | 0.42 | 1.39 |

# D

# Paper

*Starts on the next page*

# The Impact of Travel Cost Reimbursement on the Value of Travel Time

Levi Mulder

**Abstract**

To determine whether the investment in an infrastructure project, e.g. a new train track, is worthwhile often a cost-benefit analysis is used. This cost-benefit analysis (CBA) is used to determine whether the benefits of a particular investment outweigh the costs. In the context of infrastructure projects, the largest benefits are often the travel time savings as a result of a shorter travel time for the (new) users of that piece of infrastructure. To trade off the monetary costs required to build e.g. the railroad and the travel time savings one needs a "conversion factor". This factor is referred to in practice as the value of travel time ($VTT$). Since governments have limited resources for infrastructure projects, accurately determining the $VTT$ helps prioritize projects that deliver the greatest societal benefits.

In practice, this $VTT$ is often determined through discrete choice experiments (DCE). In a DCE a respondents are asked to make trade-offs between time, money and other attributes. Through the estimation of a mathematical model, one can extract the $VTT$. However one can imagine that making this tradeoff is somewhat different for respondents who are normally reimbursed for their travel cost. These respondents normally do not have to consider costs, or part of them, in their day-to-day travel decisions. Therefore their value of travel time might be different. If such a difference exists one would want to take this into account when calculating national $VTT$ averages.

Thus the research question posed is; *To what extent does a travel cost reimbursement in a stated choice experiment affect the value of travel time?*

We investigated the impact of travel cost reimbursement (TCR) on the value of travel time ($VTT$) using both the NL2022 (Netherlands) and UK2014 (United Kingdom) datasets. We hypothesized three mechanisms through which a $VTT$ difference could appear: whether TCR influenced the $VTT$ directly, if respondents with a TCR made less consistent choices, and if these respondents exhibited a higher willingness to accept (WTA) compared to those without TCR. Utilizing Multinomial Logit (MNL) and Mixed Logit (ML) models, we analysed the effect of receiving TCR on the $VTT$ while allowing for discrete scale heterogeneity. Furthermore, by considering the respondents' reference situation, we separated a TCR interaction on the $VTT$ into willingness to accept (WTA) and willingness to pay (WTP) components. The models developed for this research accounted for socio-economic interactions and ensured that the findings were robust across different experimental setups.

The results indicated that commuter respondents, in the NL2022 dataset, with a full Travel Cost Reimbursement had a $VTT$ which was approximately 29% - 45% higher compared to respondents who did not receive a reimbursement. The same type of respondents in the UK2014 dataset reported a $VTT$ which was approximately 58% - 143% higher. This effect was significant and consistent for experiments in which only time and cost were presented as attributes of the alternatives. However, as the number of attributes in the choice task increased the difference in $VTT$ was no longer significant. Furthermore, we were unable to discover any consistent patterns regarding scale heterogeneity between the $VTT$ for nTCR and TCR respondents. Therefore we cannot accept our hypothesis that respondents who receive a travel cost reimbursement make less consistent choices. Moreover, the interaction with the $VTT$ was more predominant in the $WTA$ domain compared to the $WTP$ domain. However, the model which took this into account did not perform statistically better in terms of model fit compared to a model which only allowed for equal TCR interaction on both the $WTA$ and the $WTP$ quadrant.

We concluded that the $VTT$ is significantly different for respondents receiving and not receiving a travel cost reimbursement, but that these changes cannot be explained through scale heterogeneity or a WTA, WTP disparity. We argue that national $VTT$ values should take travel cost reimbursement into account as national $VTT$ values are in practice based on the 2-attribute time-cost experiments.

**Index Terms**

Choice Modelling, Heteroskedasticty, MNL, Mixed Logit, Scale heterogenetiy, Travel Cost Reimbursement, VTT.

Levi Mulder is a graduate student from the Technical University of Delft

# I. Introduction

TO determine whether the investment in an infrastructure project, e.g. a new train track, is worthwhile, governments often use a cost-benefit analysis. This cost-benefit analysis (CBA), as conceived by [1] is used to determine whether the benefits of a particular investment outweigh the costs. In the context of infrastructure projects, the largest benefits are often the travel time savings as a result of a shorter travel time for the (new) users of that piece of infrastructure. [2] To trade off the monetary costs required to build e.g. a railroad and the travel time savings one needs a ”conversion factor”. This factor is referred to as the value of travel time (VTT). Since governments have limited resources for infrastructure projects, accurately determining the VTT helps prioritize projects that deliver the greatest societal benefits.

The trivial way to determine this VTT would be to ask people directly, *”What are you willing to pay to shorten your journey by 15 minutes?”*. However, people are not ”programmed” to explicate the way we make choices. This is also empirically proven in for example location research into refugee camps [3]. When people are asked whether to explicate their choices they tend to give socially desirable answers. In this example by [3] respondents *told* that they would be willing to welcome refugees, but whenever they were able to actually choose the extent to which their neighbourhood would welcome refugees, they were not so welcoming anymore. Thus when observing their actual choice behaviour their underlying preferences tell a different story. Furthermore, economic theory [4] is based on what people do, thus what choices they make, not on their judgements. Therefore, in practice and academics one often uses stated preference as introduced by [5], [6] to determine the VTT.

In stated preference, respondents (people who answer the questionnaire) are presented with hypothetical choice situations which can be fully controlled by the researcher. Respondents are asked to make a tradeoff between travel cost and travel time. One of the options is cheaper but slower, the other one is quicker but more expensive. Furthermore, contrary to RP, in SP the alternatives (choice options) and the number of attributes (characteristics) can be fully controlled by design. It is thus possible to present respondents with choices which they normally are not able to make in real life. These characteristics make SP a popular method to determine the VTT and are one of the reasons this method is used in the Netherlands [7].

To calculate the VTT, from an SP study, one needs to present respondents with variations in the time and cost attribute in a choice task. Thereafter, the researcher can calculate the value of travel time by analysing the choices using a mathematical model.

One can imagine that making this tradeoff could potentially be somewhat distanced for respondents who receive a travel cost reimbursement (TCR). These respondents suddenly have to take costs into account, which they normally might not or only partially. Travel costs might not show up in their bank account and it often does not impact their monthly expenses. Therefore, if such a respondent bases their choices on their usual day-to-day travel decisions, their VTT could be significantly higher as compared to respondents who have to normally pay for the travel costs themselves. It would be relevant to evaluate whether the VTT for respondents with and without a travel cost reimbursement differs indeed, as An impressive 80% of employees in the Netherlands receive some form of travel cost reimbursement. Thus the number of people in the population to whom a different VTT might apply is extensive [8]. This relationship is, to the author’s knowledge, not yet covered in the scientific literature, as will be elaborated on in subsection I-A.

As the value of travel time is important in allocating money to projects which yield the greatest societal benefits, it is important to investigate to what extent and for which choice tasks a travel cost reimbursement impacts the VTT.

## *A. Research Gap*

As was stipulated above the focus of this paper is to investigate whether respondent travel cost reimbursement impacts the value of travel time. In this section, this aim is analysed by synthesising the body of scientific knowledge.

Many different forms of reimbursement have been examined. Some are incentivizing travel through a financial bonus, [9], [10] others penalizing some form of travel through a financial penalty. [11], [10], [12] All aim to quantify the effects of some policy on respondent preferences. A good example is provided by [9] who investigated the relationship between commuter benefits and transportation mode choice. They found that benefit programs designed to encourage walking, cycling, and using public transport were effective in increasing the use of these modes of transportation. One interesting finding is that all studies focus on some form of reimbursement scheme and mode-choice behaviour. None of these articles review to what extent a travel cost reimbursement could impact the VTT.

If respondents with a TCR base their choices on their reference situation we would expect them to more frequently choose the more expensive option because cost is of less, or no, importance in their reference situation, therefore increasing their VTT. The size of the potential difference will thus be dependent on the extent to which these respondents base their choices on their “normal reimbursed” situation. Why respondents would base their decisions on their reimbursed situation is discussed in subsection I-C.

This culminates in the **first hypothesis**; We hypothesise that respondents with a travel cost reimbursement will have a significantly higher VTT as compared to respondents who do not receive a travel cost reimbursement.

Additionally, there might be other behavioural mechanisms which could cause respondents with a travel cost reimbursement to alter their choice behaviour and therefore indirectly affect the VTT. Respondents with a travel cost reimbursement might have a lower choice task motivation because their choices normally have no direct consequences in real life. This is called lack of payment consequentiality [13] or lack of budget constraints [14]. From the point of view of the respondent, this in essence means that they feel like; "I don't really care about these choices as my travel costs are reimbursement anyhow." Therefore they could exert less cognitive effort when making choices. It has been shown that less cognitive effort and choice task motivation could lead to a lower choice consistency [15], [16]. Choice consistency in its turn can bias model parameters [17]. The reasoning is that a higher degree of randomness in the choices can be captured as a lesser degree of importance of the model parameters, therefore lowering the VTT. As such, a lower degree of choice consistency for respondents with a travel cost reimbursement could indirectly lower the VTT, whilst in essence, the VTT might actually be higher. This effect could therefore also dampen the direct effect postulated in Hypothesis 1. The exact workings of how choice consistency can bias model parameters is left to subsection I-C. Furthermore, other potential behavioural mechanisms are also discussed in subsection I-C. This culminates in our **second hypothesis**; We hypothesise that respondents with a travel cost reimbursement make less consistent choices as compared to respondents without a travel cost reimbursement.

There is a third way, a third hypothesis, through which travel cost reimbursement could impact the VTT. It might be possible that respondents with a travel cost reimbursement reasonably assume that cost levels up to their reference costs are reimbursed and costs above their reference level are not. This assumption is not unreasonable when one considers that there are many forms of travel cost reimbursement which are fixed at a certain limit. If this assumption echoes through in their choices, they will opt more often for the more faster but expensive route if the costs are below their reference, and they will more often choose the cheaper but slower route whenever the costs are above their reference costs. This would thus imply two different values for the VTT. We provide a visual interpretation in Figure 1.

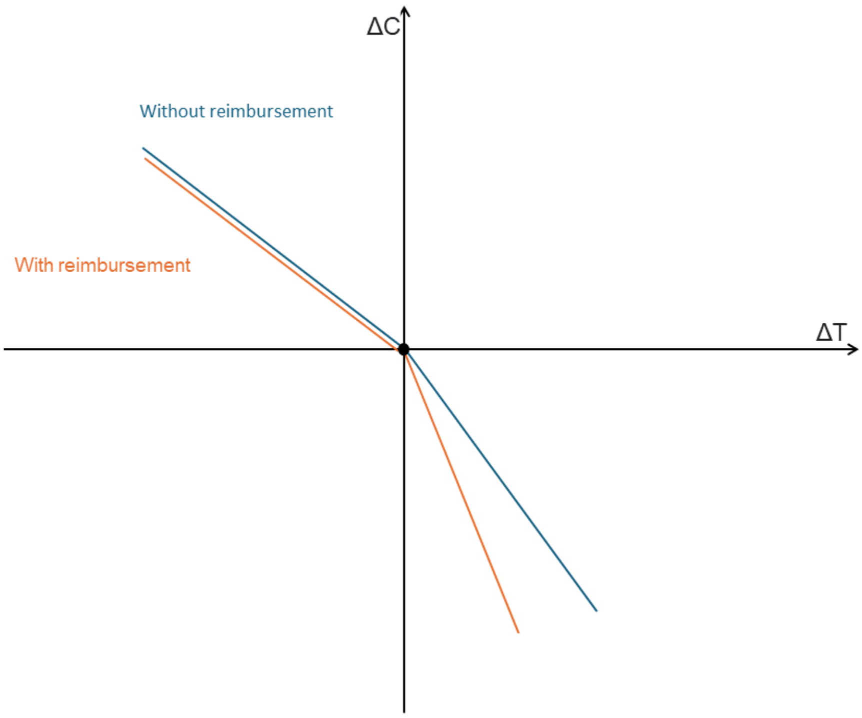



Fig. 1: Visual representation of Hypotheses 3. Indifference curves for a two-attribute time, money trade-off. The graph is centred on the reference travel time and travel costs of the respondents. $\Delta T$ and $\Delta C$ respectively denote the time and the cost difference relative to the respondent's reference time and cost levels.

Do observe that the discontinuity in Figure 1 as described above is also present for respondents without a reimbursement. This discontinuity at $\Delta T = 0, \Delta C = 0$ is referred to as a *sign effect* in the choice modelling literature[18]. It is often explained through the behavioural phenomena known as loss aversion[19]. This sign-effect is well documented and it appears in both scientific [20] and applied work [7]. However, as also visualised Figure 1 we argue this well-known discontinuity might very well be more prominent for respondents with a reimbursement as compared to respondents without. To summarise our **third hypothesis**; We hypothesise that respondents with a travel cost reimbursement have a significantly higher willingness to accept (WTA) as compared to respondents without a travel cost reimbursement.

### B. Goal & Research Question

Thus to further understand the relationship between travel cost reimbursement and the value of travel time the author aims to answer the following research question:

*"To what extent does a travel cost reimbursement in a stated choice experiment affect the value of travel time?"*

The sub-questions are presented below following the previously stated hypotheses. subsection I-C further elaborates on the necessary theory and relevant literature for each of these sub-questions.

1) When does a travel cost reimbursement impact the $VTT$?
2) Do respondents with a travel cost reimbursement make less consistent choices?
3) Does a travel cost reimbursement impact the $WTP$ and the $WTA$ differently?
4) To what extent do sign effects and choice consistency explain an effect of $TCR$ on the $VTT$?
5) What is the impact of the findings on national cost-benefit analysis?

By answering this research question and its subquestions the contribution to the literature is **threefold**: It will potentially provide analysts tasked with calculating value of travel time metrics grips on how to take travel cost reimbursement effects into account. Secondly, it will provide researchers insight into the choice behaviour of respondents with a travel cost reimbursement. Lastly, it could further inform policymakers about the effects of travel cost reimbursement and how it changes people's preferences.

We tested our hypotheses on the Dutch national [7] value of travel time study and the UK national [21] value of travel time study. These studies both documented whether or not a respondent received a travel cost reimbursement. As we a priori expected potential effects to be small, the large sample sizes of these studies were deemed to be of use. Secondly, these studies can both be subdivided into multiple sub-studies, which varied the number of attributes. This enabled us to test the hypotheses across these sub-studies and therefore allow for a more varied analysis and generalize the results to a greater extent.

Furthermore, these two datasets enabled us to evaluate the hypotheses for two different types of instructions. Respondents for the NL2022 dataset were instructed to ignore their reimbursement and make choices as if they had to pay out of their own pocket. The UK2014 survey instructed respondents to take a travel cost reimbursement into account if applicable.

To generalize the results as much as possible we aimed to develop both Multinomial- [4] and Mixed Logit models [22]. Furthermore, non-parametric and parametric analyses had to be performed to ensure that the travel cost reimbursement parameters did not capture any other socioeconomic effects. Testing the statistical significance of the final model parameters, enabled the evaluation of possible reimbursement effects. The specifics of the methods are elaborated in section III.

This paper structured as follows; In subsection I-C the necessary literature and theory are presented to fully understand the hypotheses which are formalized in section II. Subsequently, in section III the method to test these hypotheses is discussed. This Method chapter covers the model development process, the data samples used, model estimation and finally model application on a case study. Thereafter section IV discusses the results structured around the hypotheses and ends with the model application results on the case study. The findings are discussed and summarized in section VI and discusses subsequently the; methodological implications, policy implications, limitations and suggestions for further research.

### C. Choice Consistency and Error Heterogeneity

Linking choice consistency to error heterogeneity was done by several papers [23], [24], [16], [25], [15], [26], [27] all related a some form of choice consistency to the scale parameter. However, scale parameter variations are often directly linked with choice task complexity. [26] is one of the few who formally links error heterogeneity to the notion of choice consistency. This is arguably a better definition as the actual observation in the choice data is a variation in choice consistency which is captured by a heterogeneous error term. In many papers, the link with choice task complexity is the main research objective and the link with error heterogeneity is made without mentioning choice consistency. We therefore adopt the definition of [26], who states that:

> "Choice consistency is measured as the variance of the random error component in the consumer utility function: the smaller this variance, the higher choice consistency." [26]

We can thus relate choice consistency to the magnitude of the scale parameter $\mu$. However, we must stress that we can only talk about choice consistency relative to the observed utility. Thus conceptually, choice consistency is in essence the degree to which respondents "follow" the utility function defined by the researcher. In other words, how well the utility function can explain the observed choices.

### D. Modelling Scale Parameter Variations

There are several options available to model scale parameter variations. These scale parameter variations are sometimes referred to as, error heterogeneity, scale heterogeneity or heteroskedasticty [26]. These terms are used interchangeably in the scientific literature. [28] were one of the first to provide a model for (discrete) error heterogeneity. It was employed to be able to account for error heterogeneity across datasets to be able to estimate a combined RP, and SP model. It involves a nesting structure where respondents are assigned to nests based on whether they were part of the RP or the SP dataset. They estimated a parameter which denoted the difference in scale across the two datasets. [29] further developed this to what they call a Generalized Mixed Multinomial Logit Model, to be able to account for both taste and scale heterogeneity. This approach enabled [29] to model a continuous form of heteroskedasticity. He based his approach on the idea that the degree of randomness in choices has a certain distribution in the population. The work by [29] work was criticised by [20]. They argue that the G-MNL model is not a generalisation of the Mixed Logit Model but instead a more restrictive version. This argument is based on the fact that mixed-logit when specified properly can accommodate a full covariance matrix between the randomly distributed attributes. In the standard G-MNL specification, this covariance matrix in essence collapses to a single parameter. As we aim to model heteroskedasticity across, what we assume are homoskedastic subsets, the more discrete approach provided by [28] provides the best framework.

## II. Hypotheses

The goal of this chapter is to formalize the hypothesis stated in the Introduction. The method to test the hypotheses is not discussed in this chapter but left to section III. Thus coding of variables, sample information, data suitability and model development are all discussed there.

### *Hypothesis 1, The Value of Travel Time*

Our hypothesis stated in section I stated that respondents with a $TCR$ might have a higher $VTT$ compared to respondents not receiving a travel cost reimbursement. The hypothesis stated in the introduction is as follows; In the context of a stated choice experiment, we can formalize this hypothesis and its corresponding null hypothesis as Equation 1

$$H_0: \quad \frac{\beta_{\text{time,reimb}}}{\beta_{\text{cost,reimb}}} = \frac{\beta_{\text{time,non-reimb}}}{\beta_{\text{cost,non-reimb}}} \tag{1}$$

$$H_1: \quad \frac{\beta_{\text{time,reimb}}}{\beta_{\text{cost,reimb}}} > \frac{\beta_{\text{time,non-reimb}}}{\beta_{\text{cost,non-reimb}}} \tag{2}$$

The same hypothesis but now presented in willingness to pay space is given in equations Equation 3 and Equation 4.

$$H_0: \quad \text{VTT}_{\text{reimb}} = \text{VTT}_{\text{non-reimb}} \tag{3}$$

$$H_1: \quad \text{VTT}_{\text{reimb}} > \text{VTT}_{\text{non-reimb}} \tag{4}$$

Several behavioural arguments and one microeconomic argument can be made to support Equation 1. Firstly, from a behavioural viewpoint, lack of payment consequentiality [13] for respondents with a $TCR$ could increase the $VTT$. In the context of discrete choice experiments, *payment consequentiality* ensures that respondents treat the hypothetical choices as if they were real decisions involving actual financial costs or benefits. Respondents accustomed to reimbursements may struggle to internalize the hypothetical scenario where they bear the travel costs themselves. This discrepancy is primarily because their real-world decisions involve no direct financial consequences, leading to less careful consideration of cost implications in a hypothetical context such as a discrete choice experiment.

Secondly, also from a behavioural viewpoint, budget constraints [14] could also cause an increase in the $VTT$ for respondents receiving a $TCR$. The argument is quite simple, whenever a respondent does not receive a travel cost reimbursement, a change in the cost of their day-to-day commute could significantly impact their monthly finances. Therefore, their actual real-life budget for travel expenses could constrain how much they are willing to spend on travel costs. Respondents who do receive a $fTCR$ or a $pTCR$ do not have to consider these budget constraints.

Thirdly, Thaler's concept of mental accounting [30] explains how individuals separate their finances into different accounts for various purposes. The "mental account" out of which travel expenses are paid is quite small for respondents who receive a $fTCR$ or a $pTCR$. Whereas for respondents not receiving a $TCR$ the "mental account" out of which travel expenses are paid naturally has to be larger. Therefore, the psychological cost associated with more expensive choices could be lower for respondents receiving a $TCR$ and consequentially increase their $VTT$ compared to their non-reimbursed counterparts.

Lastly, from a microeconomic perspective, the actual disutility of costs in real-life decisions is lower whenever travel costs are reimbursed. Therefore time is relatively more important than costs which consequentially increases the $VTT$ [4].

These behavioural phenomena are not new to practitioners, therefore what is often done in practice is to instruct respondents to, either take or not take their $TCR$ into account. The purpose of such an instruction is to reduce the likelihood that respondents make assumptions about the choice situation. In our case, respondents in the NL2022 survey were instructed to; make decisions as if they were travelling in their own time and that they have to pay for the costs themselves. Respondents in the UK2014 survey were instructed to; take into account who normally pays for their journey when making choices. However, as is also widely established, respondents do not read instructions meticulously [31]. Thus the behavioural phenomena as described earlier can still play a role in the decision-making process. The fact that respondents in the UK2014 and the NL2022 surveys got different instructions might even increase the extent to which our results are generalisable.

### *Hypothesis 2, Scale Heterogeneity*

The second hypothesis stated in the introduction is as follows: *"We hypothesise that respondents with a travel cost reimbursement make less consistent choices as compared to respondents without a travel cost reimbursement.* In we can formalize this hypothesis and its corresponding null hypothesis as Equation 5.

$$H_0: \quad \mu_{\text{reimb}} = \mu_{\text{non-reimb}} \tag{5}$$

$$H_1: \quad \mu_{\text{reimb}} > \mu_{\text{non-reimb}} \tag{6}$$

To understand Equation 5 we again ought to define a few concepts, starting with the interpretation of the scale parameter. Thereafter, we explain, with the body of scientific literature and with a synthetic example, why the scale parameter is relevant for the VTT.

We now argue why travel cost reimbursement could impact choice consistency based on several behavioural notions. Firstly, respondents without reimbursement might be prone to start calculating what an alternative means for their monthly expenses. Respondents with a travel cost reimbursement might not feel the need to make such calculations, again related to lack of payment consequentiality [13]. Therefore the respondents without a travel cost reimbursement could potentially put in more cognitive effort when making choices because they are making actual calculations or inferences about what an alternative would mean for their actual expenses. [32] has shown, in a series of now-famous experiments, that cognitive effort could activate analytic reasoning. The idea by [32] was such, that when faced with cognitive strain, individuals are more likely to reject intuitive answers and engage more thoroughly with the task to solve problems accurately. [33] described this as; cognitive strain activating analytic reasoning and cognitive effort. An increase in cognitive effort in its turn increases choice consistency.

Secondly, as indicated by [15], the significance of a product class —measured by factors like the proportion of an individual's income— is related to the amount of cognitive effort people put into a decision. The significance of costs in proportion to an individual's income is naturally lower whenever an individual receives a $TCR$. Consequentially, respondents with a $TCR$ could be less motivated to exert a substantial amount of cognitive effort when making choices, increasing randomness and therefore decreasing the magnitude of the scale parameter.

Note that these three behavioural theories are analogous and are all able to explain a potential effect on choice consistency.

*Hypothesis 3, Sign Effects*

Our third hypothesis stated that a difference in the $VTT$ might be explained through an asymmetry in the preferences of respondents with a TCR. We can mathematically formalize the hypothesis per Equation 7.

$$\frac{(WTA_{TCR} - WTA_{nTCR})}{WTA_{nTCR}} > \frac{(WTP_{TCR} - WTP_{nTCR})}{WTP_{nTCR}} \tag{7}$$

Where

$WTA_{TCR}$ is the willingness to accept a deterioration in travel time for respondents receiving a travel cost reimbursement,
$WTA_{nTCR}$ is the willingness to accept a deterioration in travel time for respondents not receiving a travel cost reimbursement,
$WTP_{TCR}$ is the willingness to pay for an improvement in travel time for respondents receiving a travel cost reimbursement,
$WTP_{nTCR}$ is the willingness to pay for an improvement in travel time for respondents not receiving a travel cost reimbursement.

The reasoning is simple; as stated in the introduction a respondent who receives a travel cost reimbursement might very well assume that their costs are reimbursed up to their normal travel costs. Therefore, costs which are below the respondents' reference might be of less importance to them, causing an increase in the $VTT$. However, this increase in the $VTT$ will thus not be constant across all cost levels since cost levels above the respondent reference might normally not be reimbursed for that respondent. To understand Equation 7 we thus first ought to discuss; what an asymmetry in preferences entails, where it usually refers to, what the behavioural notion is and why it is relevant for the overall $VTT$.

Now that we established what the sign effect is and where it comes from we argue *why* a TCR could have an asymmetric impact on the $VTT$. There are two behavioural reasons which motivate the hypothesis in Equation 7.

Firstly, respondents with a TCR could potentially assume that up to their reference costs level is reimbursed. As their willingness to accept, $WTA_{TCR}$, is determined based on choices with travel costs below their *reference* travel costs, the $WTA_{TCR}$ could be higher as compared to respondents who do not receive a $TCR$. Along the same line of reasoning, we can argue that, under the assumption that respondents receive a $TCR$ up to their reference level, the $WTP_{TCR}$ should not be much different as compared to the willingness to pay for respondents not receive a travel cost reimbursement $WTP_{nTCR}$.

The second reason which motivates the hypothesis in Equation 7, is that there are types of travel cost reimbursement which are limited to a fixed amount. Applying microeconomic reasoning, the disutility of the travel costs is, therefore, lower below this fixed amount and higher above this fixed amount. Thus creating a discontinuity in the $VTT$ at that fixed amount. Regardless of the limit of the travel cost reimbursement, this will always result in a larger difference between the $WTA_{TCR}$ and the $WTP_{TCR}$ as compared to respondents who do not receive a TCR.

## III. METHOD

This chapter is devoted to the method applied to be able to answer the research question and the sub-research questions. Do note that the first sub-research question has been the topic of section subsection I-C. We first present the details of the datasets and methodology used in the research. It begins with a description of the dataset, including the choice tasks, reimbursement information retrieval, and filtering criteria for respondent instructions. We present a summary table, a table of reimbursement categories, and a reimbursement percentage distribution graph. Next, we assess data suitability and discuss limitations. The operationalization section covers coding of travel cost reimbursement variables, defining reimbursement bins, and presenting distribution tables and figures. The models applied include Multinomial Logit (MNL), MNL with heteroskedasticity, and Mixed Logit (ML). We describe the model development process, from initial analysis to evaluating reimbursement effects and adding interactions. We end this chapter by describing how we applied the developed model to assess policy implications. But, we begin this chapter with a general overview of model development.

### A. Coding and Definition of TCR

We now define the $D_{\text{TCR}}$ variables as denoted in Equation 8 and Equation 9. To do so we discuss how $TCR$ information is measured, how respondents were instructed to deal with an eventual $TCR$ and the corresponding model parameters.

Respondents self-reported whether they received a $TCR$. The options which were presented to the respondents are presented in Table I. In the NL2022 dataset, respondents were first instructed to think of a journey they recently made and to state e.g. with what type of mode and for what purpose they made this journey. Thereafter, respondents were told that the entire questionnaire would be about this journey they just described. Respondents were eventually asked to state if they received some form of reimbursement for this reference journey. Respondents could choose amongst the options presented in Table I. The UK2014 approach is, apart from specific wording, the same. Travel cost reimbursement information was retrieved similarly, by asking respondents to self-report the travel cost reimbursement of their reference trip, subsequently, it was stated that the questions in the survey would be about that reference trip. The descriptions are matched to those in the NL2022 dataset in Table I.

TABLE I: Options for travel expense reimbursement

| Option | NL2022 | UK2014 |
|---|---|---|
| 1 | Yes, in full | Employer/company paid |
| 2 | Yes, partially | - |
| 3 | Yes, travel costs were shared with fellow passengers | Shared with other travellers |
| 4 | No, no reimbursement | I paid all costs |
| 5 | There were no travel expenses for me | All paid by other travellers, Colleague /partner/family member paid |
| 6 | Other, namely: ___ | Other (please specify) |

In the NL2022 dataset, when respondents reported receiving a partial travel cost reimbursement a subsequent question was asked; "What percentage of the travel costs was reimbursed?" Which again is thus a self-reported metric. However one can argue that a respondent with a reimbursement of e.g. 10% has in fact paid most of the trip out of their own pocket. We therefore opted to redistribute the categorical reimbursement statuses. Respondents with a partial reimbursement were redistributed according to Table III. Thus respondents who reported receiving a partial travel cost reimbursement of e.g. $9\%$ were added to the *No reimbursement category*.

In the NL2022 dataset, only respondents with reimbursement options 1,2 and 4, in Table III, were selected for the analysis. This was done since for these options the reimbursement status is unambiguous and the number of observations was satisfactory. In the UK2014 dataset, only respondents with reimbursement options 1 and 4 were selected for the analysis. This was done based on the number of observations. Furthermore, respondents were also filtered based on several rules. For example, whenever the self-reported travel cost was unrealistic, respondents were removed from the dataset. It is outside the scope of this study to cover all these filtering procedures. Instead, we provide Table II with an overview of the final selection of observations. Thus denoting the number of completed choice tasks, not the number of respondents.

The meaning of the $TCR$ variables differs slightly between the NL2022 and the UK2014 datasets as the respondents got different instructions on how to deal with their $TCR$.

For the UK2014 dataset, the respondents were instructed to:

> "Please imagine that each situation is exactly the same as for the actual #MODE# journey at the time you made the journey, except for the travel costs and travel time. Choose the one you prefer, keeping in mind all of the circumstances of the #MODE# journey at the time you made the journey and assuming that the only things that vary are the costs and times shown. For each of the following pairs of options, carefully compare the two options, bearing in mind who paid for the actual journey and who else is travelling with you."[21]

For the NL2022 dataset respondents were instructed to:

> "Imagine that you can choose between two different routes. The travel time can vary, for example, because one trip is shorter or less congested, and not because you accelerate more quickly. The costs can also vary, for example,

TABLE II: NL2022 and UK2014, Observations distributed per mode

| **Mode** | **Commute** (1) | **Business** (2) | **Total** |
|---|---|---|---|
| Car (1) | 11680 | 5008 | 16688 |
| Train (2) | 7832 | 2456 | 10288 |
| BTM (3) | 6848 | 2024 | 8872 |
| **Total (NL2022)** | 26360 | 9488 | 35848 |
| Car (1) | 4835 | 8650 | 13485 |
| Train (2) | 4585 | 8790 | 13375 |
| LRT (3) | 2900 | 2300 | 5200 |
| Bus (UK2014, 4) | 1675 | 200 | 1875 |
| **Total (UK2014)** | 13995 | 19940 | 33935 |

because one trip requires less fuel. Each time, we ask which trip you prefer. For each choice, assume that both trips are feasible (even if they seem unrealistic), you travel in your own time, all costs are for your account, and all other characteristics are equal for both trips (equally safe, equally pretty, parking costs are equal, etc.). Additionally, all other circumstances, such as the weather and your appointments and activities that day, are the same as during the journey you have described."[7]

Note that the respondents of the UK2014 dataset were instructed **not** to ignore their travel cost reimbursement. We argue that this is an advantage because this difference might explain differences in the $TCR$ interactions.

To include the $TCR$ in the model specification, dummy coding was applied. The reason for applying dummy coding instead of effects coding has to do with the way the parameters can be interpreted. In effects coding the parameters should be interpreted relative to the sample average $VTT$, when dummy coding we interpret the parameters as absolute values for every subgroup. The absolute values for these subgroups are important in the calculation of national averages. Therefore dummy coding is preferred over effects coding.

TABLE III: Coding for reimbursement levels, $D_{\text{nTCR}}$, $D_{\text{pTCR}}$ and $D_{\text{fTCR}}$

| **Variable** | | **Coding** | | |
|---|---|---|---|---|
| **Reimbursement** | | $D_{\text{nTCR}}$ | $D_{\text{pTCR}}$ | $D_{\text{fTCR}}$ |
| 0 to 15% | No | 1 | 0 | 0 |
| 15 to 85% | Partial | 0 | 1 | 0 |
| 85 to 100% | Full | 0 | 0 | 1 |
| **Quadrants** | | $D_{\text{WTP}}$ | $D_{\text{WTA}}$ | |
| $WTP$ quadrant | | 1 | 0 | |
| $WTA$ quadrant | | 0 | 1 | |

with this definition of the $D_{TCR}$ variables, the corresponding parameters have to be interpreted accordingly. To illustrate; a $\delta_{fTCR}$ value of 0.09 and a t-ratio of 2.05, would indicate a 9% increase in the $VTT$ compared to the category which is fixed, $(0.09 \cdot 100\% = 9\%)$. The t-ratio of $2.05$ would indicate that we can reject the null hypothesis that the $VTT$ is equal between groups with 95% certainty.

### B. Model Development

*a) H1:* As we hypothesised that with an increase in travel cost reimbursement the $VTT$ would also increase we added an interaction with the $VTT$ and $TCR$, which is denoted by $\delta_{TCR}$. The definition and coding of the $TCR$ variable were discussed in subsection III-A. Furthermore, we added the $\delta_{TCR}$ parameters as relative interactions. The magnitude of the $\delta_{TCR}$ parameter must thus be interpreted as a relative change to the reference $VTT$. This specification allows for easy interpretation and quick comparison with different magnitudes for the $VTT$ itself. The interaction as described above is shown below in Equation 8

$$V_{ij} = \mu \cdot \{C_{ij} + [VTT \cdot T_{ij} \cdot (1 + \delta_{\text{TCR}} \cdot D_{\text{TCR}})]\} \quad \forall i \in I, j \in J \tag{8}$$

Where parameters to be estimated are denoted in blue, and $C_{ij}$ and $T_{ij}$ are respectively the cost and time for alternative $j$ for respondent $i$. Furthermore, $D_{\text{TCR}}$ denotes the variable which corresponds with the $TCR$. Further definition of this variable and the subsequent interpretation of the $\delta_{\text{TCR}}$ parameter is discussed in subsection III-A. We opted to use a significance level of 95% throughout this study, consequentially whenever the t-ratio $\geq 1.96$ [1] we can reject the Null hypothesis that the $VTT$ is not dependent on $TCR$, and therefore supporting the alternative hypothesis that the $VTT$ is dependent on $TCR$. This hypothesis can naturally only be tested for experiments where a $VTT$ can be accurately estimated, thus experiments which include a cost and a time attribute, and where the number of respondents is sufficient.

*b) H2:* We expected an increase in travel cost reimbursement would lead to more random choice behaviour which can be subsequently captured in the scale of the utility function as was discussed in subsection I-C. Hence, we incorporate $TCR$ interactions with the scale of the utility similarly to [26], [17], denoted in Equation 9

$$V_{ij} = \mu \cdot (1 + \mu_{TCR} \cdot D_{\text{TCR}}) \{C_{ij} + [VTT \cdot T_{ij} \cdot (1 + \delta_{\text{TCR}} \cdot D_{\text{TCR}})]\} \quad \forall i \in I, j \in J \tag{9}$$

Where again, the parameters to be estimated are denoted in blue, As preliminary estimates indicated significant interactions between the $VTT$ and $TCR$ we opted to not solely add the $\mu_{\text{TCR}}$ but instead also add the $\delta_{\text{TCR}}$ interactions as previously denoted in Equation 8.

Following the warnings given by [34], [35], simultaneously allowing for $VTT$ interaction circumvents the scale parameter from capturing differences in the $VTT$. With the specification as in Equation 5, this scale parameter interaction with $TCR$ was evaluated on experiments with two attributes, where only the $VTT$ is allowed to be a random variable. This again circumvents capturing other sources of correlation in the scale parameter [34], [35] This approach is very similar to the one applied in [17], where they aimed to explain scale heterogeneity through respondent-experienced time pressure. To explicate; the scale parameter $\mu$ is thus *not* specified as a random variable, again to circumvent capturing other sources of correlation.

This approach enabled us to evaluate the first *and* the second hypothesis with this specification since we allow for differences in the $VTT$ (Hypothesis 1) and differences in scale (Hypothesis 2). By comparing the estimates resulting from both specifications we can evaluate to what extent allowing for scale heterogeneity changes the $VTT$ estimates.

*c) H3:* Lastly, we hypothesised that a travel cost reimbursement effect on the $VTT$ would be more profound in the willingness-to-accept ($WTA$) compared to the willingness-to-pay ($WTP$). To test this hypothesis we could only make use of quadrant type questions. These are questions where one of the alternatives is exactly the same as the reference journey described by the respondent, and the other alternative is either an increase in time but a decrease in costs or vice-versa. Consequentially, the only sub-experiment which had these types of questions was the SP1A sub-experiment of the NL2022 dataset. These types of questions enabled estimating two separate values for the $VTT$, denoted by the $WTA$ and the $WTP$.[7] To evaluate our third hypothesis we added a similar interaction as denoted in Equation 8 but now separated on the $WTA$ and the $WTP$ parameters. This approach yielded the specification in Equation 10

$$\begin{aligned} V_{ij} = \mu \cdot \Bigg\{ C_{ij} + \Big[ WTP \cdot T_{ij} \cdot D_{\text{WTP}} \cdot (1 + \delta_{\text{TCR,WTP}} \cdot D_{\text{TCR}}) \Big] \\ + \Big[ WTA \cdot T_{ij} \cdot D_{\text{WTA}} \cdot (1 + \delta_{\text{TCR,WTA}} \cdot D_{\text{TCR}}) \Big] \Bigg\} \quad \forall i \in I, j \in J \end{aligned} \tag{10}$$

Where; $D_{\text{TCR,WTA}}$, $D_{\text{TCR,WTP}}$ are dummy variables to denote whether a choice set belongs to either the $WTA$ or the $WTP$ quadrant and $\delta_{\text{TCR,WTA}}$, $\delta_{\text{TCR,WTP}}$ are the corresponding parameters. These parameters should again be interpreted as a relative change to the reference $WTA$ and $WTP$. Testing the third hypothesis was done by comparing the magnitudes of the $\delta_{\text{TCR,WTA}}$, $\delta_{\text{TCR,WTP}}$ parameters. A higher $\delta_{\text{TCR,WTA}}$ than $\delta_{\text{TCR,WTP}}$ would support our hypothesis. We tested whether this difference was statistically different by (1), comparing it to a utility specification which allowed for $WTA$ and $WTP$ differences but restricted the $TCR$ interaction to be equal for both quadrants and evaluating whether the likelihood ratio statistic [36] yielded a significant improvement in model fit (2), evaluating whether the 95% confidence intervals of the $\delta_{\text{TCR,WTA}}$, $\delta_{\text{TCR,WTP}}$ parameters overlapped. The utility specification which assumes a homogenous interaction on both the $WTA$ and the $WTP$ is specified in Equation 11.

[1] Naturally under the condition that the degrees of freedom are sufficient.

$$V_{ij} = \mu \cdot \Big\{ C_{ij} + \Big( \Big[ WTP \cdot T_{ij} \cdot D_{\text{WTP}} \Big] + \Big[ WTA \cdot T_{ij} \cdot D_{\text{WTA}} \Big] \Big) \cdot (1 + \delta_{\text{TCR}} \cdot D_{\text{TCR}}) \Big\} \quad \forall i \in I, j \in J \tag{11}$$

Where the $\delta_{\text{TCR}}$ parameter interacts equally on both the $WTP$ and the $WTA$ in a relative sense, as it is defined as a multiplicative interaction. This specification is thus purely meant to provide a baseline where we do allow differences in the $WTA$ and the $WTP$, and subsequently evaluate what the average interaction is over the $WTA$ and the $WTP$ quadrants.

The limitations as discussed above defined the types of choice experiments which were eligible for testing. A summary is provided in Table V.

*1) Isolating TCR Effects:* We ensured that the parameters which interact with $TCR$ only capture $VTT$ differences due to a change in travel cost reimbursement. This was done because respondents with e.g. a high income are known to generally have a higher $VTT$. We observed in our sample that respondents with a high income also had a higher likelihood of receiving a full Travel Cost Reimbursement. Therefore we examined whether the distribution of $TCR$ was related to other socioeconomic or design-specific variables of which we *a-priori* knew from practical experience and or the body of scientific literature that they interacted with either the $VTT$ or the scale parameter ($\mu$).

Examining whether there was a significant association between receiving a $TCR$ and other variables, was done by performing either a Chi-square test of association for categorical variables or a KrusKal-Wallis test for non-categorical variables [37], [38]. Whenever the Chi-Square test or the Kruskal-Wallis test indicated a statistically significant relation the variable was added as an interaction. Instead of discussing all the test results we opted to provide them in Table IV. Do note that contrary to our a-priori expectations all of the variables tested yielded significant differences between receiving or not receiving a $TCR$.

Furthermore, whenever an association between the variable and the $TCR$ was significant we calculated the Cramers-V for the Chi-Squared test or the $\eta^2$ for the Kruskal-Wallis test to indicate the size of the effect. [2]

TABLE IV: Results of Chi-square and Kruskal-Wallis tests, to test whether a different number of reimbursed respondents across groups is statistically significant.

| Variable | Test | Statistic | p-value | D.O.F. | Cramér's V / $\eta^2$ | Effect Size |
|---|---|---|---|---|---|---|
| Gender | Chi-Square | 245.252*** | $< 0.001$ | 2 | 0.078 | Negligible |
| Income class | Chi-Square | 404.114*** | $< 0.001$ | 10 | 0.071 | Negligible |
| Age class | Chi-Square | 462.796*** | $< 0.001$ | 6 | 0.076 | Negligible |
| Education class | Chi-Square | 73.312*** | $< 0.001$ | 8 | 0.030 | Negligible |
| Reference time | Kruskal-Wallis | 1965.548*** | $< 0.001$ | - | 0.049 | Small |
| Reference costs | Kruskal-Wallis | 2186.181*** | $< 0.001$ | - | 0.055 | Small |
| Travel frequency | Kruskal-Wallis | 1863.306*** | $< 0.001$ | - | 0.047 | Small |
| Work situation | Chi-Square | 529.627*** | $< 0.001$ | 8 | 0.081 | Negligible |
| ZZP class | Chi-Square | 1000.340*** | $< 0.001$ | 2 | 0.158 | Small |
| Recruitment-type | Chi-Square | 155.648*** | $< 0.001$ | 2 | 0.062 | Negligible |
| Employment class | Chi-Square | 1959.111*** | $< 0.001$ | 4 | 0.157 | Small |

*Note.* *** p ¡ 0.001.

Subsequently, to ensure that there were no other respondent characteristics or survey-related aspects associated with a respondent receiving a $TCR$ or not, a correlation analysis was performed. Based on the outcome of this correlation analysis, employment status and ZZP were added as interaction variables with the $VTT$. Employment status refers to whether a respondent is currently in between jobs or not. ZZP refers to a variable which indicates whether a respondent is self-employed or not.

*2) Final Model Specification:* The model development procedure as discussed above yielded the utility specification in Equation 12. This utility specification was used for the $SP1A$ experiment in the NL2022 dataset presented in Table V. This specification was adopted where necessary to accommodate experiments with more attributes. We did this by simply adding the extra attribute multiplied by a parameter to the utility function. To accommodate the choice data from the UK2014 dataset only minor adjustments were made in the number of categories for socioeconomic interactions with the $VTT$.

[2]Effect size was not considered in the decision whether to include it as an interaction in the model, this decision was purely based on the statistical tests itself.

$$V_{ij} = \mu \tag{12}$$
$$\cdot \left(1 + \mu_{\text{nTCR}} \cdot D_{\text{nTCR}} + \mu_{\text{pTCR}} \cdot D_{\text{pTCR}} + \mu_{\text{fTCR}} \cdot D_{\text{fTCR}}\right) \tag{13}$$
$$\cdot \left(\left(\frac{T_i^{\text{ref}}}{60}\right)^{\lambda_{T\mu}} \cdot \left(\frac{C_i^{\text{ref}}}{5}\right)^{\lambda_{C\mu}}\right) \tag{14}$$
$$\cdot \{C_{ij} + [VTT \cdot T_{ij} \tag{15}$$
$$\cdot \left(1 + \delta_{\text{nTCR}} \cdot D_{\text{nTCR}} + \delta_{\text{pTCR}} \cdot D_{\text{pTCR}} + \delta_{\text{fTCR}} \cdot D_{\text{fTCR}}\right) \tag{16}$$
$$\cdot \left(1 + \sum_{k=1}^{2} \delta_{\text{employment},k} \cdot D_{\text{employment},k}\right) \tag{17}$$
$$\cdot \left(1 + \sum_{k=1}^{2} \delta_{\text{paidEmployment},k} \cdot D_{\text{paidEmployment},k}\right) \tag{18}$$
$$\cdot \left(1 + \delta_{\text{zzp}} \cdot D_{\text{zzp}}\right) \tag{19}$$
$$\cdot \left(1 + \sum_{k=1}^{4} \delta_{\text{ageClass},k} \cdot D_{\text{ageClass},k}\right) \tag{20}$$
$$\cdot \left(1 + \delta_{\text{panel}} \cdot D_{\text{panel}} + \delta_{\text{intercept}} \cdot D_{\text{intercept}}\right) \tag{21}$$
$$\cdot \left(1 + \sum_{k=1}^{5} \delta_{\text{edu},k} \cdot D_{\text{edu},k}\right) \tag{22}$$
$$\cdot \left(1 + \delta_{\text{male}} \cdot D_{\text{male}} + \delta_{\text{female}} \cdot D_{\text{female}}\right) \tag{23}$$
$$\cdot \left(1 + \sum_{k=1}^{6} \delta_{\text{freq},k} \cdot D_{\text{freq},k}\right) \tag{24}$$
$$\cdot \left(\frac{T_i^{\text{ref}}}{T_0^{\text{ref}}}\right)^{\lambda_{\text{T}}} \cdot \left(\frac{C_i^{\text{ref}}}{C_0^{\text{ref}}}\right)^{\lambda_{\text{C}}} \tag{25}$$
$$\cdot \left(1 + \sum_{k=1}^{6} \delta_{\text{inc},k} \cdot D_{\text{inc},k}\right)\Big]\Big\} \tag{26}$$

As mentioned earlier, the interactions added in Equation 12 were based on whether; (1) the variable of which was a-priori known to impact the $VTT$ or the scale had a statistically significant association with $TCR$ or (2) no a-priori interaction with the $VTT$ or the scale parameter was expected but the variable correlated with $TCR$. These variables were added as multiplicative interactions. Categorical variables were dummy-coded. The interactions with the respondent's reference time and reference costs were added as power law interactions since [7] found that these explained the data the best.

As denoted in Equation 14 this power law interaction was also added to the scale parameter. This was done because (1) respondents with different levels of $TCR$ also had statistically significant different levels of reference time and reference costs and (2), we knew a-priori that reference time and costs impact the scale parameter due to the use of a pivot type design. [3] In a pivot-type design, respondents are presented with tradeoffs which are based on their reference time and costs to ensure that they are faced with realistic choices. However, this thus changes the attribute levels which are presented to respondents and therefore induces scale differences. To accommodate these scale differences interactions with the reference time, reference costs and the scale parameter were added.

### C. The Sample

Both national stated preference surveys subdivided their experiments into sub-experiments. Not all of these sub-experiments are relevant to this study. For example, respondents who travelled by boat have by definition the travel purpose other, as this is typically a leisure activity. Only sub-experiments which were used to test the hypothesis are presented in Table V.

The relevant sub-experiments were all presented as unlabeled route-choice experiments, where respondents had to choose between two hypothetical alternatives, each described by the attributes as in Table V. Every respondent had to make 9 of these route choice decisions. One of these choice situations contained an alternative which dominated the other alternative. This was to ensure respondents were not randomly making decisions or e.g. always choosing the left option. During model estimation, this check question was not considered as it does not provide any information on the trade-offs respondents had to make.

[3]The reference time and the reference costs, denoted by $T_i^{\text{ref}}$ and $C_i^{\text{ref}}$ respectively, reflect the time and the cost of the self-reported reference journey. Respondents were instructed to assume that all conditions, except the journey time and the journey costs, were the same as their reference journey.

TABLE V: Overview of datasets, attributes, modes and hypotheses

| Dataset | Name | Attributes | Modes | Hypotheses |
|---|---|---|---|---|
| NL2022 | SP1A | time, cost | Car, Train, BTM | $H_1$, $H_2$, $H_3$ |
| | SP2A | time, cost, reliability | Car, Train, BTM | $H_1$ |
| | SP3A | time, cost, access/egress time, transfer time, number of transfers | Train, BTM | $H_1$ |
| | SP4A | time, cost, crowding, ability to sit, frequency | Train, BTM | $H_1$ |
| UK2014 | SP1 | time, cost | Car, Train, LRT, Bus | $H_1$, $H_2$ |
| | SP2 | time, cost, reliability | Car, Train, LRT, Bus | $H_1$ |

We now discuss to what extent the datasets are suitable to answer the research question and suitable to test the hypotheses. We do so by discussing a-priori limitations.

Firstly, the datasets from the 2022 Dutch VTT study and the 2014 UK VTT study are from different periods. Economic conditions, travel behaviours, and policies may have changed between these periods, potentially affecting the comparability of the datasets. However, as we are investigating a relative difference as compared to the VTT, absolute changes in the VTT can safely be ignored.

Secondly, the 2022 Dutch VTT study and the 2014 UK VTT study did not record the same socioeconomic data. Therefore, not all socioeconomic interactions were applied consistently. However, the most important socioeconomic interactions, interactions which could have explained away a potential reimbursement effect, are included across both datasets

Thirdly, as was mentioned earlier, the UK2014 respondents were not able to specify whether they received a *partial* travel cost reimbursement. Therefore it might have been possible that respondents with a partial travel cost reimbursement chose either the option; ”Employer/company paid” or ”I paid all costs.” However, we expect that whenever a respondent receives some form of travel cost reimbursement they are more likely to opt for the ”Employer/company paid” option.

### D. Model Estimation

The developed model was estimated using Apollo Choice modelling software [39]. For the Mixed Logit specifications, halton draws were used to simulate the randomly distributed $VTT$ [40]. Sufficiently stable parameters were retrieved for 500 draws. The $VTT$ was given a lognormal distribution, as [7] found that after extensive testing this fitted the data the best. Furthermore, using the lognormal distribution ensures that the $VTT$ can only attain positive values in line with microeconomic theory.

The developed model was estimated on the other choice tasks with minimal adjustments to prevent confirmation bias. This approach was taken to ensure that the model's performance was assessed objectively across different choice tasks, without being influenced by prior expectations. By limiting adjustments, we aimed to validate the model's generalizability and robustness based on the data itself, rather than fitting it to preconceived outcomes. Consequently, this philosophy was also used when estimating the $TCR$ interaction effects on the UK2014 data. Only necessary adjustments in, for example, the number of income classes and the number of age classes were accommodated. As a result, the model fit was not as good as it might have been able to, but we argue this approach increases the validity of the results.

### E. Model application

After the model was developed we applied it to the societal cost-benefit analysis (SCBA) for the *Ring Utrecht* project [41], aiming to give an example of how incorporating or neglecting TCR can impact a SCBA. This provides a tangible example of policy implications. In order words; by applying our findings we aimed to assess whether it actually matters, in the context of an SCBA, if travel cost reimbursement is taken into account or neglected.

## IV. Results

The goal of this chapter is to present the results and test the hypotheses. First, the effects on the VTT are analysed for models both including and excluding scale heterogeneity, thereafter we discuss the results regarding the scale parameter itself. Next, we discuss the results regarding the model which incorporates sign-effects and we end the results chapter with the findings from the *Ring Utrecht* SCBA.

### A. Hypothesis 1, VTT

We first hypothesised that respondents with any travel cost reimbursement have a higher $VTT$ as compared to respondents who do not receive any travel cost reimbursement. As described in the method section we added dummy coded multiplicative interactions with the $VTT$ to test this hypothesis. We can thus evaluate the first hypothesis by observing the statistical significance of these interaction parameters, a t-ratio greater than 1.96 means that we can reject the Null hypothesis that the $VTT$ is the same as the reference group, in our case the nTCR respondents. As was already mentioned in the method chapter, we also included $VTT$ interactions for the heteroskedastic model. This enables us to (1) evaluate the first hypothesis for the heteroscedastic model, and (2) observe to what extent incorporating scale heterogeneity affects the evaluation of the first Hypothesis. In this first section, we focus on the first Hypothesis and evaluate it on the models including and excluding scale heterogeneity. The results regarding the degree of scale heterogeneity, thus the results regarding the second hypothesis are discussed in the next section.

As mentioned above, we can reject the first Null hypothesis whenever the TCR interaction parameters can be estimated with a t-ratio exceeding 1.96. This reflects a significance level of $\alpha = 5\%$. This is equivalent to rejecting the Null hypothesis whenever the 95% confidence interval does not overlap with the VTT of respondents who do not receive any travel cost reimbursement. These equivalent methods are shown in Figure 2a, Figure 2b Figure 2c, Figure 2d. The coloured t-ratios correspond with the colours of the projected confidence intervals. The values correspond with the magnitude of the relative difference in $VTT$ and are both plotted graphically and presented in the table below the graph. All the values are presented as relative increases/decreases to the reference $VTT$, in our case the $VTT$ of respondents who did not receive any travel cost reimbursement. The values in the table below the graphs are the actual estimated values, thus they are reported as fractions. An increase in the $VTT$ of e.g. 31% is therefore reported in the table as 0.31. The figures for the NL2022 data report both the estimates for respondents receiving a full Travel Cost Reimbursement (fTCR) and a partial Travel Cost Reimbursement (pTCR). As the UK2014 data did not provide any information on what portion of the reference journey was reimbursed, the UK2014 figures only provide one reimbursement interaction per estimation. We added the number of observations to allow for a more nuanced interpretation of the results because an insignificant difference can sometimes be attributed to a lack of observations.

As shown in Figure 2a and in Figure 2b the results are mixed. We found that for a simple time-money tradeoff, denoted by SP1A in Figure 2a and in Figure 2b commuter respondents with a full Travel Cost Reimbursement (fTCR) had a significantly higher VTT as compared to respondents with no Travel Cost reimbursement (nTCR) thus supporting our first Hypothesis that respondents with a travel cost reimbursement have a higher $VTT$ than respondents who do not receive a travel cost reimbursement. This is the case for models with and without scale heterogeneity. The only exception is the Light Rail Transit (LRT) commute estimation in Figure 2b. However, do observe that the number of observed choices for respondents with a fTCR was 180. Therefore a likely explanation is that these respondents coincidentally had a VTT which closely matched that of the nTCR respondents.

The business estimates of the time-money tradeoff (Figure 2a & Figure 2b) were less convincing than the commute estimates. Three out of the six business estimates across the NL2022 and the UK2014 datasets showed statistically significant fTCR interactions at a 5% confidence level. Interestingly, the estimates for Business Light Rail Transit (LRT) in and for Business Bus Tram Metro (BTM) do indicate statistically significant differences in VTT compared to the reference situation of not receiving a Travel Cost Reimbursement.

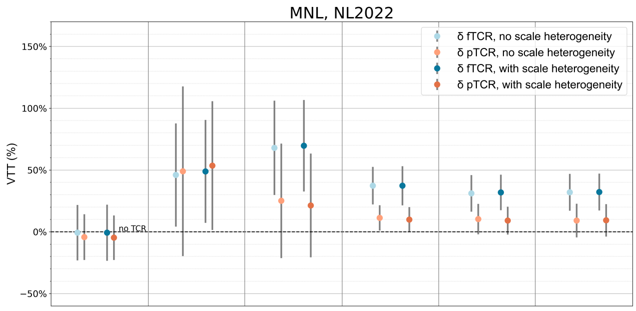
For the more complex choice experiments to obtain plausible and significant results, the estimates for the more complex experiments, SP3A & SP4A, were not disaggregated across modes and purposes. Instead, we opted to create a model which included a mode- and purpose-specific constant to account for potential VTT differences. None of the results yielded significant effects. The only exceptions are Car Business and Car Commute for SP2A, where respondents receiving a full Travel Cost Reimbursement (fTCR) had a significantly higher VTT as compared to respondents who received no Travel Cost Reimbursement (nTCR).

The mixed Logit specification generally widened the confidence intervals of the $TCR$ interaction. However, this deterioration in statistical significance did not cause any $fTCR$ interactions to become insignificant at a 95% confidence level. Conversely, when examining the Commute estimates in Figure 2a and Figure 2c, we can observe that the mixed logit specification did cause the interactions with the $VTT$ and the partial travel cost reimbursement, denoted by $pTCR$, to decrease in magnitude and as a consequence, the difference in $VTT$ for $pTCR$ respondents is no longer statistically significant w.r.t. respondents not receiving a travel cost reimbursement.

Interestingly, when comparing the Mixed Logit and the MNL estimates for Train Commute (Figure 2a, Figure 2c) we can observe that the magnitude of the $\delta_{\text{fTCR}}$ parameter increased when comparing the MNL and the Mixed Logit specifications. For the model excluding scale heterogeneity, this was an increase from 32% (t-ratio, 4.36) to 45% (t-ratio, 4.04) and for the heteroscedastic model, this was an increase from 32% (t-ratio, 4.11) to 43% (t-ratio, 3.79).

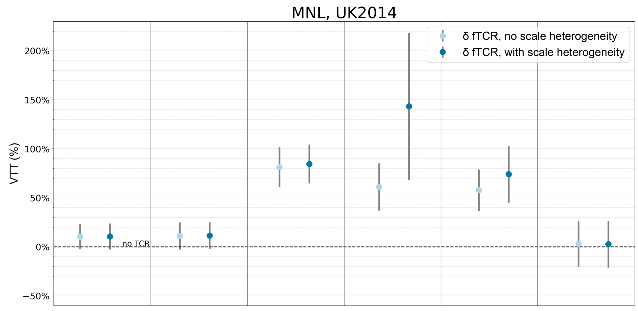
The findings, for commuter respondents with a full Travel Cost Reimbursement across the UK2014 and the NL2022 datasets, between incorporating and neglecting scale heterogeneity and for Mixed- and Multinomial Logit, supported our hypothesis that respondents with a travel cost reimbursement have a higher $VTT$. The results indicated that the difference in $VTT$ is somewhere between 29% and 143%. The lowest reported t-ratio was 2.48 whenever we excluded the LRT commute estimation since it lacked observations.

The other findings strongly suggested that TCR affects the $VTT$ but were either inconsistent or insignificant.



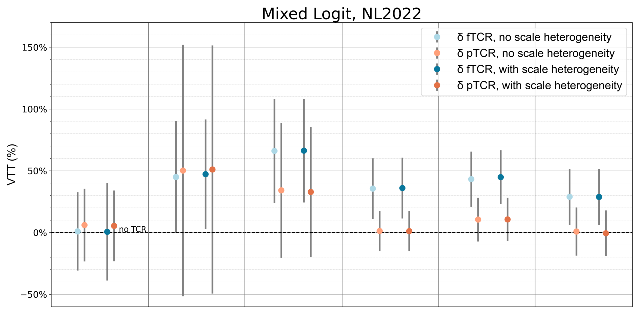

| | | | | | | |
|---|---|---|---|---|---|---|
| δ_fTCR. | -0.01, (-0.06) | 0.46, (2.16) | 0.68, (3.49) | 0.37, (4.83) | 0.31, (4.11) | 0.32, (4.21) |
| δ_pTCR. | -0.04, (-0.45) | 0.49, (1.4) | 0.25, (1.06) | 0.11, (2.17) | 0.1, (1.64) | 0.09, (1.32) |
| δ_fTCR | -0.01, (-0.06) | 0.49, (2.3) | 0.7, (3.69) | 0.37, (4.62) | 0.32, (4.36) | 0.32, (4.23) |
| δ_pTCR | -0.05, (-0.52) | 0.54, (2.02) | 0.21, (1.0) | 0.1, (1.93) | 0.09, (1.57) | 0.09, (1.39) |
| Δ(μ) fTCR | 0.08, (0.71) | -0.07, (-0.4) | -0.24, (-1.29) | -0.09, (-1.17) | -0.16, (-1.84) | -0.09, (-1.14) |
| Δ(μ) pTCR | 0.25, (1.91) | 0.96, (0.98) | -0.04, (-0.12) | 0.29, (3.05) | 0.09, (0.73) | 0.12, (1.05) |
| Modes | Car | Train | BTM | Car | Train | BTM |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 1376 | 1192 | 752 | 2120 | 3136 | 1936 |
| #obs pTCR | 624 | 48 | 152 | 2168 | 752 | 872 |
| #obs nTCR | 504 | 248 | 328 | 1552 | 872 | 1352 |

(a) NL2022, time-money tradeoff results.



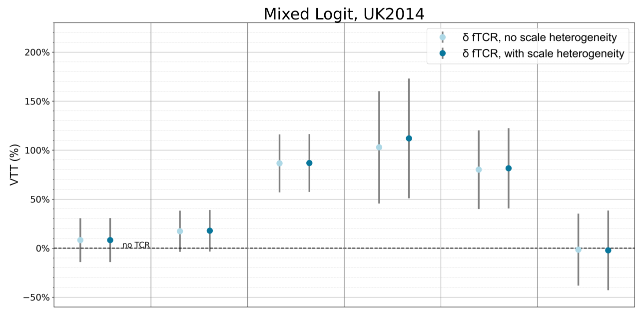

| | | | | | | |
|---|---|---|---|---|---|---|
| δ_fTCR. | 0.1, (1.59) | 0.11, (1.57) | 0.81, (7.87) | 0.61, (4.97) | 0.58, (5.39) | 0.03, (0.26) |
| δ_fTCR | 0.11, (1.57) | 0.11, (1.65) | 0.85, (8.34) | 1.43, (3.75) | 0.74, (5.01) | 0.03, (0.22) |
| Δ(μ) fTCR | -0.01, (-0.1) | -0.06, (-0.61) | -0.28, (-1.73) | -0.62, (-7.34) | -0.26, (-2.9) | 0.02, (0.14) |
| Modes | Car | Train | LRT | Car | Train | LRT |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 3360 | 3455 | 495 | 895 | 810 | 180 |
| #obs pTCR | 0 | 0 | 0 | 0 | 0 | 0 |
| #obs nTCR | 965 | 940 | 655 | 3940 | 3775 | 2720 |

(b) UK2014, time-money tradeoff results.




| | | | | | | |
|---|---|---|---|---|---|---|
| δ_fTCR. | 0.01, (0.06) | 0.45, (1.94) | 0.66, (3.08) | 0.36, (2.85) | 0.43, (3.79) | 0.29, (2.5) |
| δ_pTCR. | 0.06, (0.4) | 0.5, (0.97) | 0.34, (1.23) | 0.01, (0.15) | 0.11, (1.16) | 0.01, (0.09) |
| δ_fTCR | 0.01, (0.03) | 0.47, (2.09) | 0.66, (3.1) | 0.36, (2.87) | 0.45, (4.04) | 0.29, (2.48) |
| δ_pTCR | 0.05, (0.37) | 0.51, (1.0) | 0.33, (1.22) | 0.01, (0.14) | 0.11, (1.2) | -0.01, (-0.06) |
| Δ(μ) fTCR | 0.24, (1.73) | -0.24, (-1.02) | -0.07, (-0.34) | -0.11, (-1.2) | -0.24, (-2.24) | -0.11, (-1.18) |
| Δ(μ) pTCR | 0.59, (2.92) | -0.25, (-0.48) | 0.13, (0.37) | 0.1, (0.97) | -0.18, (-1.5) | 0.43, (2.46) |
| Modes | Car | Train | BTM | Car | Train | BTM |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 1376 | 1192 | 752 | 2120 | 3136 | 1936 |
| #obs pTCR | 624 | 48 | 152 | 2168 | 752 | 872 |
| #obs nTCR | 504 | 248 | 328 | 1552 | 872 | 1352 |

(c) NL2022, time-money tradeoff results.




| | | | | | | |
|---|---|---|---|---|---|---|
| δ_fTCR. | 0.08, (0.72) | 0.17, (1.61) | 0.87, (5.74) | 1.03, (3.52) | 0.8, (3.91) | -0.01, (-0.08) |
| δ_fTCR | 0.08, (0.72) | 0.18, (1.64) | 0.87, (5.77) | 1.12, (3.59) | 0.82, (3.9) | -0.02, (-0.11) |
| Δ(μ) fTCR | 0.11, (1.23) | 0.16, (1.56) | -0.04, (-0.22) | -0.35, (-3.98) | -0.07, (-0.55) | -0.2, (-1.34) |
| Modes | Car | Train | LRT | Car | Train | LRT |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 3360 | 3455 | 495 | 895 | 810 | 180 |
| #obs pTCR | 0 | 0 | 0 | 0 | 0 | 0 |
| #obs nTCR | 965 | 940 | 655 | 3940 | 3775 | 2720 |

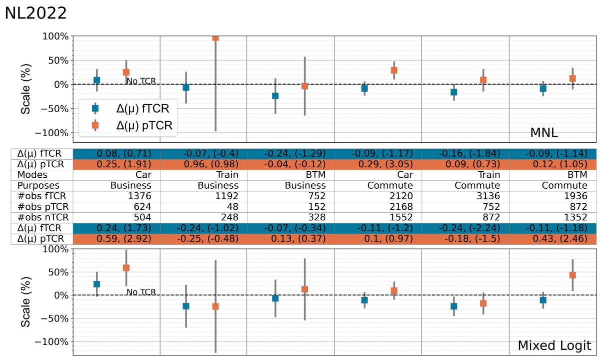
(d) UK2014, time-money tradeoff results.

Fig. 2: Time-money tradeoff results, including 95% confidence intervals for both NL2022 and UK2014 data. $\Delta(\mu)$ estimates are reported as fractional differences with respect to the nTCR observations including their t-statistic in parentheses.

## *B. Hypothesis 2, Scale Heterogeneity*

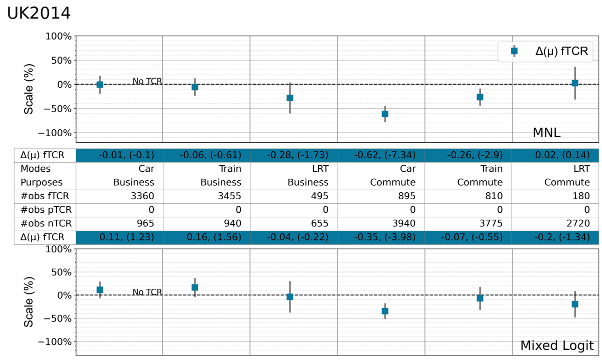
As a second hypothesis, we expected that respondents with a travel cost reimbursement make less consistent choices as compared to respondents who did not receive any travel cost reimbursement. We tested this hypothesis by incorporating interactions with the scale of the utility function. Our findings did generally *not* support our hypothesis at a significance level of 95%. In other words; we did not find convincing evidence which supported our hypothesis that respondents with a travel cost reimbursement make less consistent choices.

The way we reported our results is similar to the style of reporting for the first hypothesis, only now we report the difference in scale as opposed to the difference in $VTT$. The results are presented in Figure 3a and Figure 3b. The colours in the table again match those in the graph. In contrast to the figures for the first hypothesis we now included both the Mixed Logit and the MNL estimates in one figure. Whenever a $\Delta(\mu)$ estimate is lower than 0, it indicates that respondents with a travel cost reimbursement had a lower scale than respondents without a travel cost reimbursement. A $\Delta(\mu)$ estimate higher than 0 naturally implies the opposite. Whenever the t-ratio ¿ 1.96 or similarly whenever the 95% confidence interval does not overlap with the horizontal 0% line, we reject the Null hypothesis for the respective estimate at $\alpha = 0.05$, therefore supporting our hypothesis that the scale for TCR respondents is lower than for nTCR respondents.




| | | | | | | |
|---|---|---|---|---|---|---|
| Δ(μ) fTCR | 0.08, (0.71) | -0.07, (-0.4) | -0.24, (-1.29) | -0.09, (-1.17) | -0.16, (-1.84) | -0.09, (-1.14) |
| Δ(μ) pTCR | 0.25, (1.91) | 0.96, (0.98) | -0.04, (-0.12) | 0.29, (3.05) | 0.09, (0.73) | 0.12, (1.05) |
| Modes | Car | Train | BTM | Car | Train | BTM |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 1376 | 1192 | 752 | 2120 | 3136 | 1936 |
| #obs pTCR | 624 | 48 | 152 | 2168 | 752 | 872 |
| #obs nTCR | 504 | 248 | 328 | 1552 | 872 | 1352 |
| Δ(μ) fTCR | 0.24, (1.73) | -0.24, (-1.02) | -0.07, (-0.34) | -0.11, (-1.2) | -0.24, (-2.24) | -0.11, (-1.18) |
| Δ(μ) pTCR | 0.59, (2.92) | -0.25, (-0.48) | 0.13, (0.37) | 0.1, (0.97) | -0.18, (-1.5) | 0.43, (2.46) |

(a) NL2022 time-money tradeoff results.




| | | | | | | |
|---|---|---|---|---|---|---|
| Δ(μ) fTCR | -0.01, (-0.1) | -0.06, (-0.61) | -0.28, (-1.73) | -0.62, (-7.34) | -0.26, (-2.9) | 0.02, (0.14) |
| Modes | Car | Train | LRT | Car | Train | LRT |
| Purposes | Business | Business | Business | Commute | Commute | Commute |
| #obs fTCR | 3360 | 3455 | 495 | 895 | 810 | 180 |
| #obs pTCR | 0 | 0 | 0 | 0 | 0 | 0 |
| #obs nTCR | 965 | 940 | 655 | 3940 | 3775 | 2720 |
| Δ(μ) fTCR | 0.11, (1.23) | 0.16, (1.56) | -0.04, (-0.22) | -0.35, (-3.98) | -0.07, (-0.55) | -0.2, (-1.34) |

(b) UK2014 time-money tradeoff results.

Fig. 3: Time-money tradeoff results for NL2022 and UK2014 models with scale heterogeneity. The difference in scale is presented in the graph and reported in the table. The graph includes the 95% confidence interval whilst the table includes the t-statistic in parentheses.

Examining Figure 3a, we found that, across respondents with a full Travel Cost Reimbursement for both the Mixed- and Multinomial Logit models the $\Delta(\mu)$ estimates are of the expected sign, except for the car business estimate, hinting that respondents with a fTCR might have a lower scale parameter as compared to respondents without a TCR. However, none of these interactions were statistically significant at $\alpha = 0.05$, except for the Mixed Logit estimation for Train Commute.

Interestingly for the MNL Car Commute estimate and the Mixed Logit BTM commute estimate the interactions regarding a partial Travel Cost Reimbursement indicated a higher scale compared to nTCR respondents. This is not in line with our second hypothesis which stated that an increase in $TCR$ would lead to a decrease in scale. These estimations thus indicate that respondents with a partial travel cost reimbursement make more consistent choices compared to nTCR respondents.

The scale interaction estimates, denoted by $\Delta(\mu)$fTCR, for UK2014 in Figure 2b are quite similar. All of the interactions are of the expected sign, hinting that respondents with a full Travel Cost Reimbursement (fTCR) had on average a higher scale compared to respondents who had to pay for the travel costs themselves but again these interactions were not significant. The only estimates with significant scale interactions were Car Commute and Train Commute; -0.62 t-ratio (7.34) and -0.26 t-ratio (2.9) respectively. The $\Delta(\mu)$fTCR UK2014 Car Commute estimate was the largest in magnitude indicating a scale reduction of 62% compared to respondents who received no Travel Cost Reimbursement. Additionally, this relatively large difference in scale also reduced the VTT interaction for this particular estimate as can be observed in Figure 2b. Whenever scale heterogeneity was not taken into account the VTT difference was 61% (t-ratio, 4.97). Conversely, taking scale heterogeneity into account yielded a VTT difference of, on average, 143% (t-ratio, 3.75).

The difference in scales for the Car Commute estimate as was observed in the MNL specification, was not as apparent in the Mixed Logit specification, as can be observed in Figure 3b. In the MNL specification, the scale of the utility for $fTCR$ respondents was 62% lower relative to $nTCR$ respondents, whereas for the Mixed Logit specification the difference in scale was reduced to just 35%, (t-ratio, 3.98). Simultaneously, the indirect effect of the scale difference on the $VTT$ is also greatly reduced as can be observed by revisiting Figure 2d

Generally speaking, the differences in scales when comparing MNL and Mixed Logit related to scale heterogeneity do not seem to follow a particular pattern. The Mixed Logit specification, does not cause scale heterogeneity between nTCR and fTCR respondents to increase or decrease.

## *C. Hypothesis 3, Sign Effects*

To achieve statistically significant parameters and test whether or not there is a disparity between the willingness to pay and the willingness to accept, the model incorporated all modes and purposes in one estimation. This was a necessary aggregation to achieve significant results. However, the model did include parameters to account for apriori known differences between the VTT across modes and purposes. The results of this estimation are presented in Table VI. Interpretation of these results is less trivial than before. Firstly, both models, including a TCR sign effect and excluding a TCR sign effect *both*, allowed us to estimate a VTT in the WTP quadrant and the WTA quadrant. We can observe that for both estimations, the VTT in the WTP quadrant is higher as compared to the VTT in the WTA quadrant, denoted by the *VTT(WTP)* and the *VTT(WTA)* parameters. Conceptually this means respondents were willing to pay more to avert a deterioration in travel time than they were willing to pay to improve their travel time. This aligns with empirical evidence [7], [18] and with the behavioural notion of loss aversion [42]. The fact that these were different from one another thus implies that the iso-utility curve is no longer a straight line, but has a discontinuity at the reference time and cost of the respondent. We originally hypothesised that this difference in willingness to accept (WTA) and willingness to pay (WTP) might be the sole reason that there is a difference in the VTT to begin with. To test this we estimated a travel cost reimbursement (TCR) interaction on both sides of the discontinuity, on the VTT(WTP) and the VTT(WTA), denoted by the estimate on the left in Table VI. We expected that the TCR interaction with the VTT would be explained solely by the interaction with the VTT(WTA). Our results supported this hypothesis. The nTCR interaction with the VTT(WTA), denoted by nTCR(WTA), indicated a significant decrease of 25% in the VTT (t-ratio, 3.23), whilst the nTCR(WTP) parameter indicated an insignificant decrease of 10% (t-ratio, 1.01). This indicated that, in this sample, the VTT difference can be mostly attributed to a difference in the WTA quadrant. However, when comparing both models in terms of their LogLikelyhood the Ben-Akiva and Swait test indicated that the model improvement is not statistically significant at a 5% level. Thus, the model including a TCR sign effect did not outperform the model excluding a TCR sign effect in terms of model fit. Therefore, it could be due to random chance that we observed a slight improvement in LogLikelihood.

TABLE VI: Estimation results Including and Excluding TCR Sign Effect. Both models account for a discontinuity in the VTT at the reference time and costs of every individual respondent.

| | | (1) Including TCR sign effect | | (2) Excluding TCR sign-effect | |
|---|---|---|---|---|---|
| Modelled outcomes | | 3690 | | 3690 | |
| LL(final) | | -1653.62 | | -1654.28 | |
| Adj.Rho-square (C) | | 0.3404 | | 0.3405 | |
| Parameters | Unit | Estimate | t-ratio(0) | Estimate | t-ratio(0) |
| $\mu$ | - | -0.89 | -21.88 | -0.89 | -21.95 |
| VTT(wtp) | €$/h$ | 8.77 | 8.61 | 8.97 | 8.77 |
| VTT(WTA) | €$/h$ | 11.79 | 9.04 | 11.62 | 9.07 |
| fTCR(WTP) | % | 0.00 | NA | - | - |
| nTCR(WTP) | % | -10% | -1.01 | - | - |
| fTCR(WTA) | % | 0.00 | NA | - | - |
| nTCR(WTA) | % | -25% | -3.23 | - | - |
| fTCR VTT | % | - | - | 0.00 | NA |
| nTCR VTT | % | - | - | -19% | -3.05 |

## V. Application: Ring Utrecht SCBA

This section describes the results regarding the calculation of the weighted average as denoted in subsection III-E and the application to the SCBA of the *Ring Utrecht* project.

*a) Amenability:* As we were unable to prove that travel cost reimbursement has a significant effect on the $VTT$ for business respondents using the car, we evaluated the *Ring Utrecht* SCBA assuming that the $VTT$ for business travellers is unaffected by travel cost reimbursement. For Car Commute, significant effects were found for both the MNL- and the Mixed Logit models. In the NL2022 dataset estimated $VTT$ differences were significant for both model specifications; without scale effects $\delta_{\text{TCR}}$ = 36%, t-ratio (2.85) and with scale effects $\delta_{\text{TCR}}$ = 36%, t-ratio (2.87) for the Mixed-Logit estimates. Therefore we opted to only vary the $VTT$ in the *Ring Utrecht* SCBA for commuters. We opted for the Mixed-Logit estimations as these fitted the data significantly better than the MNL estimations.

*b) Interpretation:* We present our findings in Table VII. It should be interpreted as follows; the five scenarios as described in the procedure of subsection III-E are listed as the columns of the table. The base scenario, where travel cost reimbursement interactions are not taken into account, is included in the first column. The other columns present the scenarios where TCR is taken into account, with and without scale heterogeneity and with either using $\delta_{\text{fTCR}}$ or $\delta_{\text{pTCR}}$ as inputs for the sample enumeration.

Going from top to bottom in the table; model estimates are presented first, they are the same as the model estimates presented earlier in Figure 2a. These model estimates were used as input for the sample enumeration procedure. The complete model estimations used in the *Ring Utrecht* case study can be found in **??**. Further down, the weighted values of travel time as calculated through the sample enumeration procedure are presented. These weighted values are denoted by $\overline{VTT}$.

Lastly, the different cases of the *Ring Utrecht* SCBA are presented. The cases are defined by; (1) different road layouts denoted by *Selecteren* and *Compact* and (2) under different economic scenarios, denoted by *GE** for Global Economies and *RC** for Regional Communities. The economic scenarios are similar to the Dutch WLO scenarios, which are defined by the CPB, the Netherlands Bureau for Economic Policy Analysis, and the PBL, the Netherlands Environmental Assessment Agency. A recent version of these scenarios can be found at [43]. The cases and their background are discussed in detail in the SCBA itself [41]. The *Value of Travel Time Savings* in the first scenario (presented in the first column) are exactly the same as in the *Ring Utrecht* SCBA [41]. For every scenario the Travel Time Savings in terms of the number of hours is constant. However, the monetary valuation of these hours differs according to the increase of the $\overline{VTT}$ compared to the $\overline{VTT}$ of the Baseline scenario. Table VII ends with the monetary impact on respectively the yearly and the perpetually discounted commuter and total values of travel time savings.

*c) Synthesis:* The $VTT$ effects correspond with the model estimations which were presented earlier in Figure 2c. The presented t-ratios correspond with the confidence intervals presented earlier in Figure 2d. The VTT values are very similar between neglecting and including scale heterogeneity. Both indicated a significant difference of approximately 36% in $VTT$ between receiving- and not receiving a travel cost reimbursement. Incorporating TCR caused the $VTT_{\text{nTCR}}$ and the $VTT_{\text{pTCR}}$ to reduce and the $VTT_{\text{fTCR}}$ to increase relative to the baseline scenario. This is to be expected as when we do not consider travel cost reimbursement the $VTT$ is estimated on respondents who receive a fTCR, on respondents who receive a $pTCR$ and on respondents who do not receive a travel cost reimbursement.

Next, the average weights are presented. They are simply calculated as;

$$\overline{v_{\text{TCR}}} = \left( \sum_{i \in I_{\text{TCR}}} \sum_{t \in T_i} v_{it} \right) / \left( \sum_{i \in I} |T_i| \right) \tag{27}$$

Where, $I_{\text{TCR}}$ denotes the set of respondents who receive any Travel Cost Reimbursement, $|T_i|$ denotes the number of trips in the set of trips for respondent $i$ and $v_{it}$ denotes the weight for trip $j$ made by respondent $i$. Thus in the *ODIN2023* survey, 78% of Car Commute trips were made by respondents who reported receiving some form of travel cost reimbursement, whereas 22% of trips were made without any form of travel cost reimbursement. This distribution is very similar to the 80% reported by [8]. A small difference was to be expected as [8] did not only cover the car as a mode of transportation.

Thereafter, the $\overline{VTT}$'s resulting from the sample enumeration procedure are presented. They are approximately €1.42/hour (+15%) higher and €0.95/hour (-10%) lower than the baseline $\overline{VTT}$ for models using the fTCR and the pTCR estimates respectively. There were negligible differences between the models including and excluding scale heterogeneity. Thus accounting for travel cost reimbursement does change the $\overline{VTT}$, also after applying representative weights for the Dutch population. However, we can observe that the impact on the $\overline{VTT}$ is heavily dependent on whether we adopt the fTCR or the pTCR estimates. The actual effect will be somewhere in between as a portion of the population receives a fTCR and another portion a pTCR.

Evaluating the commute $VTTS$ with the increases to the baseline sketches the impact on the *Ring Utrecht* SCBA. For instance, in the "RC Selecteren" case, the baseline value of travel time savings was €5.20M, which increased to €6.00M with the $VTT_{\text{fTCR}}$ estimate adopted and reduced to €4.67M with the $VTT_{\text{pTCR}}$ estimate adopted. Similarly, the "GE Compact" case showed an increase and decrease from €9.0M baseline to €10.38M and €8.07M. Which in turn reflects a value of $+$€32.32M and $-$€21.61M on the total travel time benefits of the *Ring Utrecht* project.

TABLE VII: Comparison of VTT with and without accounting for TCR. The Baseline is the model which does not take TCR into account. $\overline{VTT}$ is the weighted VTT as calculated through the sample enumeration procedure. VTTS denotes the Value of Travel Time Savings. The grey VTT values are not used in the sample enumeration.

| **Scenario** | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** |
|---|---|---|---|---|---|
| | | **Excl. heterskedasticty** | | **Incl. heterskedasticty** | |
| | **Baseline** | **fTCR** | **pTCR** | **fTCR** | **pTCR** |
| | | **Mixed Logit Model Estimations** | | | |
| $VTT_{\text{fTCR}}$ (€/hour) | 14.82 | 16.99 | 16.99 | 17.06 | 17.06 |
| $VTT_{\text{pTCR}}$ (€/hour) | 14.82 | 12.74 | 12.74 | 12.75 | 12.75 |
| $VTT_{\text{nTCR}}$ (€/hour) | 14.82 | 12.53 | 12.53 | 12.55 | 12.55 |
| | | **Sample Enumeration on ODIN2023** | | | |
| ODIN weight $\overline{v_{\text{TCR}}}$ | | | 0.78 | | |
| ODIN weight $\overline{v_{\text{nTCR}}}$ | | | 0.22 | | |
| $\overline{VTT}$ (€/hour) | 9.24 | 10.66 | 8.29 | 10.69 | 8.28 |
| $\overline{VTT}$ increase | 0.00% | 15.38% | -10.28% | 15.65% | -10.40% |
| | | ***Ring Utrecht* SCBA** | | | |
| **RC* selecteren** | | | | | |
| **Yearly** | | | | | |
| TTS Car commute | | | 454k [hours] | | |
| VTTS car commute | €5.20M | €6.00M | €4.67M | €6.01M | €4.66M |
| | | 15.38% | -10.28% | 15.65% | -10.40% |
| VTTS total | €16.90M | €17.70M | €16.37M | €17.71M | €16.36M |
| | | 4.73% | -3.16% | 4.82% | -3.20% |
| Absolute Increase | €- | €0.80M | -€0.53M | €0.81M | -€0.54M |
| **Discounted perpetuity** | | | | | |
| VTTS total | €187.00M | €195.85M | €181.08M | €196.00M | €181.01M |
| | | 4.73% | -3.16% | 4.82% | -3.20% |
| All benefits − All costs | -€485.00M | -€476.15M | -€490.92M | -€475.99M | -€490.99M |
| | | 1.82% | -1.22% | 1.86% | -1.23% |
| Absolute Increase | €- | €8.85M | -€5.92M | €9.00M | -€5.99M |
| **GE* Selecteren Compact** | | | | | |
| **Yearly** | | | | | |
| TTS Car commute | | | 712k [hours] | | |
| VTTS car commute | €9.00M | €10.38M | €8.07M | €10.41M | €8.06M |
| | | 15.38% | -10.28% | 15.65% | -10.40% |
| VTTS total | €37.90M | €39.28M | €36.97M | €39.31M | €36.96M |
| | | 3.65% | -2.44% | 3.72% | -2.47% |
| Absolute Increase | €- | €1.38M | -€0.93M | €1.41M | -€0.94M |
| **Discounted perpetuity** | | | | | |
| VTTS total | €885.00M | €917.32M | €863.39M | €917.89M | €863.13M |
| | | 3.65% | -2.44% | 3.72% | -2.47% |
| All benefits − All costs | €589.00M | €621.32M | €567.39M | €621.89M | €567.13M |
| | | 5.49% | -3.67% | 5.58% | -3.71% |
| Absolute Increase | €- | €32.32M | -€21.61M | €32.89M | -€21.87M |

## VI. Conclusion & Discussion

*Method Summary*

The method applied in this study investigated the impact of full- and partial travel cost reimbursement ($TCR$) on the value of travel time ($VTT$) using both the NL2022 (Netherlands) and UK2014 (United Kingdom) datasets [7], [21]. We tested three hypotheses: whether $TCR$ influenced the $VTT$, if respondents with $TCR$ made less consistent choices, and if these respondents exhibited a higher willingness to accept ($WTA$) compared to those without $TCR$. Utilising Multinomial Logit ($MNL$) and Mixed Logit ($ML$) models, the analysis incorporated evaluating the interaction terms between $TCR$ and $VTT$, while allowing for discrete scale heterogeneity. Furthermore, by considering the respondents' reference situation, we separated a $TCR$ interaction with $VTT$ into willingness to accept ($WTA$) and willingness to pay ($WTP$) components. Lastly, we evaluated to what extent accounting for $TCR$ impacted the *Ring Utrecht* case study.

*Evaluation Hypotheses*

*a) When does a travel cost reimbursement impact the VTT?:* Our 2-attribute results indicated that for a 2-attribute choice experiment commuter respondents with a full Travel Cost Reimbursement (fTCR) had a significantly higher $VTT$ compared to those without a TCR, supported by; (1) data from both the UK2014 and the NL2022 national value of travel time study, (2) MNL and Mixed Logit models, (3) including and excluding scale heterogeneity. This effect was significant and consistent for five out of six 2-attribute experiments. The likely cause for the insignificant effect of this one estimation was likely due to the lack of the number of respondents receiving a $TCR$. The Mixed Logit specifications reduced the statistical significance of the difference in $VTT$ between the $fTCR$ and the $nTCR$ groups, but this did not result in any of the abovementioned observations becoming statistically insignificant at a 5% significance level. This might be attributed to the Mixed Logit model's ability to account for unobserved heterogeneity in respondents' preferences. Unlike the Multinomial Logit (MNL) model, which only allows for observed heterogeneity in the VTT, the Mixed Logit model allows for inter-individual variations in the $VTT$. Thus some of the reimbursement effects captured in the interaction parameters for the MNL specificion might now have been captured in the distribution of the VTT.

Interestingly, as the number of attributes in the experiments increased the difference in $VTT$ between the $nTCR$ group and the $fTCR$ group was no longer significant. In practice, this is similar to what happens to interactions with socioeconomic variables, whenever the number of attributes in the experiment increases, these interactions typically also become less significant.

We thus found a difference in the $VTT$ between commuter respondents who receive a full travel cost reimbursement ($fTCR$) and those who do not receive a $TCR$ in a 2-attribute discrete choice experiment. The difference ranged from 30% to 50% for the NL2022 data and from 80% to 140% for the UK2014 data. This supports our hypothesis that respondents who receive a $TCR$ have a higher value of travel time.

The estimates for business respondents were generally not significant enough to support our hypothesis that respondents receiving a $fTCR$ have a higher $VTT$. Only the 2-attribute estimates for the closely related Bus, Tram & Metro (BTM) and Light Rail Transit (LRT) modes showed significant and consistent differences in $VTT$ between respondents who receive a full travel cost reimbursement $fTCR$ and those who do not receive a $TCR$. These effects were consistent for the MNL and Mixed Logit models both excluding and including scale heterogeneity. The differences in $VTT$ for BTM and LRT were between 65% & 70% for the NL2022 dataset and between 80% & 85% for the UK2014 dataset.

The interaction between the $VTT$ and partial Travel Cost Reimbursement (pTCR) was generally lower in magnitude and results were more often than not insignificant. Thus the observed difference in $VTT$ between respondents not receiving a TCR and respondents receiving a pTCR was lower as compared to the effects observed for respondents receiving a fTCR. The signs of the effects were generally in line with our hypothesis that travel cost reimbursement increases the $VTT$. However, we cannot formally prove a relationship between pTCR and the $VTT$ at a satisfactory significance level.

Thus to answer the first research question; a full travel cost reimbursement significantly increased the $VTT$ for commute respondents in a 2-attribute discrete choice experiment supported by data from both the Dutch national $VTT$ study and the UK national $VTT$ study. It did not significantly affect the $VTT$ for 3 or more attribute experiments or business respondents.

*b) Do respondents with a travel cost reimbursement make less consistent choices?:* The direction of the scale interaction effects is as expected for respondents receiving a full Travel Cost Reimbursement ($fTCR$), their scale parameter estimate is generally lower compared to respondents not receiving a travel cost reimbursement. The signs of the scale parameter interactions for $pTCR$ respondents were not as expected, 8/12 estimations yielded *higher* scale parameters for the $pTCR$ group compared to the $nTCR$ group. However, these abovementioned scale interactions were only significant in 5/24 estimations. In other words; we were unable to find convincing evidence which supports the notion that the scale parameter might be statistically different for respondents receiving any form of $TCR$. Thus we were unable to find convincing evidence to support our hypothesis that respondents who receive a travel cost reimbursement make less consistent choices.

*c) Does a travel cost reimbursement impact the WTP and the WTA differently?:* We were able to find evidence which supports our third hypothesis. Namely, the interaction with the $VTT$ was more predominant in the $WTA$ quadrant compared to the $WTP$ quadrant. However, the model which took this into account did not perform statistically better in terms of model fit compared to a model which only allowed for equal $TCR$ interaction on both the $WTA$ and the $WTP$ quadrant. This suggests that, although there may be a nuanced difference in how TCR influences WTA and WTP, this distinction might not be substantial enough to justify a more complex model that separately accounts for these interactions. Therefore, for practical purposes and model simplicity, treating the TCR effect as uniform across both WTA and WTP might be sufficient, without sacrificing the accuracy or predictive power of the model.

*d) To what extent do sign effects and choice consistency explain an effect of TCR on the VTT?:* Our results generally indicated that taking scale differences into account between respondents receiving a TCR and not receiving a TCR did not significantly alter the $VTT$ estimates. This is probably because the scale differences were not too large between groups.

Only one estimation yielded greatly different $VTT$ estimates due to a scale decrease of approximately 62% for $TCR$ observations as compared to $nTCR$ observations. However, the confidence intervals of the $VTT$ taking scale heterogeneity into account and the $VTT$ when neglecting scale heterogeneity still overlapped. Thus stating that the scale for this estimation was indeed statistically different cannot be concluded. When we assumed the $VTT$ to be randomly distributed following a lognormal distribution, the scale difference was reduced, and simultaneously the $VTT$ difference was no longer apparent. We are unsure about the exact underlying mechanism, but it might be possible that the randomly distributed $VTT$ in the Mixed Logit specification captures effects which were previously captured in the scale parameter in the MNL specification.

Nonetheless, we did observe hints of what we refer to as a *dampening effect*. Thus whenever the scale for the $TCR$ group was lower than for the $nTCR$ group, incorporating scale heterogeneity slightly increased the $VTT$ difference between the two groups. This was in line with our theory, as a decrease in the scale of the overall utility might be misattributed to a reduction in the $VTT$.

*e) What is the impact of the findings on cost-benefit analysis?:* We evaluated this research question by applying our findings to the *Ring Utrecht* societal cost-benefit analysis (SCBA) and assessed the extent to which accounting for $TCR$ impacted the travel time reduction benefits of this project. Based on the conclusions of the other research questions, we considered $TCR$ effects only for commuters. Given that the time benefits for the *Ring Utrecht* project are primarily associated with car transportation, we approximated the time benefits if $TCR$ for car-using commuters had been incorporated in the calculation for national averages. When calculating these national averages we separately adopted the pTCR and the fTCR estimates as the ODIN2023 database used for these calculations did not include any information on the magnitude of the travel cost reimbursement. We also opted to evaluate the *Ring Utrecht* SCBA for the model which accounts for scale heterogeneity even though we concluded earlier that there are no significant scale effects. By applying this more sophisticated model, we aimed to confirm that the conclusions drawn from the SCBA are robust across different modelling approaches. We however found no substantial differences in the total travel time benefits when using the model which allows for scale heterogeneity.

We did find that adopting the models which account for travel cost reimbursement significantly affect the national $VTT$, with this value for car commuters increasing by approximately 15.38% to 15.65% when using the $VTT_{\text{fTCR}}$ estimate, and decreasing by 10.28% to 10.40% when using the $VTT_{\text{pTCR}}$ estimate. After mapping these changes onto the travel time savings for commuting traffic in the *Ring Utrecht* SCBA, the total travel time savings of the entire project either increased by up to 5.58% or decreased by up to 3.71%, corresponding to a monetary impact ranging from an increase of €32.32M to a decrease of €21.61M in the total travel time benefits of the project. These findings underscore the importance of considering travel cost reimbursement when calculating the value of travel time, as its impact on an SCBA can be substantial. Other SCBAs might be affected to an even greater extent whenever the share of commuters is larger.

*Discussion*

We made a distinct contribution by focusing on the interaction between the $VTT$ and $TCR$ within the context of stated preference experiments. While prior research, including studies by [10], [9], [12] investigated the effects of financial incentives and penalties on travel decisions, they did not specifically address how $TCR$ impacts the $VTT$. Furthermore, our method of incorporating discrete scale heterogeneity to assess choice consistency provided a nuanced understanding of how $TCR$ influences decision-making processes. In contrast to a large body of scientific literature, who conclude that marginal willingness to pay can be significantly distorted when not taking scale heterogeneity into account, [32], [25], [44], [15], [23], [29] our results generally indicated that discrete scale differences between mutually exclusive and assumed homoskedastic subsets did not significantly alter the $VTT$. Significant effects on the $VTT$ were observed only in one case where the scale of the observed utility for respondents who received a travel cost reimbursement was reduced by as much as 62% compared to the $nTCR$ group. The reason that other studies have found more sizable differences whenever allowing for scale heterogeneity might be because it is very easy to confound scale heterogeneity with other sources of correlation. Scale heterogeneity is in essence a special form of correlation where the parameters for all the attributes increase or decrease simultaneously in magnitude across subsets. This is exactly the reason why we only allowed for scale heterogeneity in 2-attribute experiments, as the only correlation is the correlation between the parameters of the 2-attributes. Otherwise, an increase in the correlation between for example the parameter for the number of transfers and the $VTT$ can be misattributed to an increase in the scale parameter.

Furthermore, a potential limitation in the interpretation of travel cost reimbursement among business respondents, particularly those using a car, should be considered. It is possible that some business respondents who indicated that they do not receive a travel cost reimbursement might actually be driving a company car. These respondents may have interpreted the term ”travel cost reimbursement" differently, potentially excluding the value of having a company car from their response. This could have led to an underestimation of the true impact of travel cost reimbursement on their VTT, as the benefit of a company car effectively reduces their out-of-pocket travel costs, even if they do not receive a direct monetary reimbursement. Such a mechanism could explain the reduced effects we observed for business respondents using the car. Future research could address this by clarifying the definition of travel cost reimbursement in surveys or by specifically asking about the use of company cars to ensure that all forms of cost compensation are adequately captured.

For future research, we suggest a methodological improvement regarding the evaluation of socioeconomic interactions. In this study, we added and evaluated each socioeconomic interaction step by step, assessing them individually before incorporating them into the model. In hindsight, it may have been more efficient to include all potential socioeconomic interactions simultaneously and then evaluate whether the TCR interactions remain significant.

On a last note; we had two options for the sample enumeration procedure; either enumerate over trips which were made by car for commuting purposes or over respondents who frequently commute by car. Travel cost reimbursement information is stored as a personal characteristic, thus not as a trip characteristic. Therefore enumerating based on persons would perhaps make sense. However, we do not know whether the $TCR$ effect on the $VTT$ also applies to trips which the employer does not reimburse. We therefore opted to enumerate over the trips, thus assuming the reimbursement effects *only* apply to commuting trips.

*Methodological Implications*

The car commute estimate for the UK2014 dataset was the only estimate which showed a sizeable change in $VTT$ whenever we allowed for scale heterogeneity between reimbursed and non-reimbursed respondents. This showed that it is worthwhile to check the homoscedasticity assumption for MNL models to ensure unbiased parameter estimates. However one must do so very carefully because it is very easy to misattribute taste heterogeneity to scale heterogeneity.

On another note, as we found significant differences in the $VTT$ for reimbursed and non-reimbursed respondents in the UK2014 dataset, future $VTT$ studies might want to incorporate this effect when calculating national averages. However, the UK National Travel Survey (NTS), the UK counterpart of the *ODIN2023* dataset does not include information on travel cost reimbursement. Therefore, if the British authorities wish to incorporate travel cost reimbursement in the calculation of national averages, they must (1) ensure that their NTS includes travel cost reimbursement data, and (2) integrate travel cost reimbursement into their models.

A similar but slightly different argument can be made for the *ODIN2023* database. This national travel survey does collect information on travel cost reimbursement and even asks respondents what type of reimbursement they receive, for example; fixed amount, distance-based or any other type. However the magnitude of the reimbursement is not documented, thus to what extent this reimbursement covers their trip expenses is not known. Conversely, in the choice experiments information on the magnitude of the reimbursement was collected, but information on the type of reimbursement was lacking. Thus for future $VTT$ studies, we recommend aligning the travel cost reimbursement questions with the questions asked for the *ODIN2023* survey. As for the *ODIN2023* survey itself, we recommend also asking respondents about what portion of their trip is reimbursed. We would advise making this a trip-specific characteristic as a reimbursement of a fixed amount means that the proportion of the trip which is reimbursement changes with the trip costs.

One last methodological implication is related to the type of instructions respondents are given, recall; in the NL2022 dataset commuting respondents were generally instructed to make decisions as if they have to pay for the travel costs themselves and in the UK2014 dataset commuting respondents were instructed to make choices taking into account who paid for their actual reference journey. As there are more differences between the studies other than the type of instructions we cannot claim that any difference in the extent to which $TCR$ affects the $VTT$ can be attributed to the instruction. Nonetheless, our results do indicate that the $VTT$ was affected more radically in the UK2014 dataset than in the NL2022 dataset. Future $VTT$ studies could validate this by assigning for example half of the sample with a UK2014-type instruction and the other half of the sample with a NL2022-type instruction. Through such an approach one would be able to make more thorough claims on the impact of the instruction as the instruction would be the only variable which is varied. Any differences in the results aside; As we want to know what choices respondents would make in real life it is arguably better to follow the UK2014 approach, as in real life people presumably do take $TCR$ into account when making choices. However, an approach like that could potentially lead to more lexicographic choice behaviour as more people could be prone to always choose the fastest alternative. As lexicographic choice behaviour does not provide any information on the time-cost tradeoff a respondent makes, common practice is to discard these observations from the sample. Therefore instructing respondents with a UK2014-type instruction could effectively reduce the number of usable observations and therefore reduce the statistical significance of the model estimates. But whether a $TCR$ causes lexicographic choice behaviour was not the topic of our study, thus this is purely speculative.

*Policy Implications*

We found that respondents with a TCR have a higher $VTT$ in 2-attribute choice experiments, particularly among commuters. If the goal is to ascertain what individuals are willing to pay out-of-pocket for travel time, it is more appropriate to base the $VTT$ on respondents who paid for their reference journey themselves.

However, when using the $VTT$ in national cost-benefit analysis or when determining network effects as a result of an infrastructural change we argue that travel cost reimbursement should be taken into account as in practice the national average $VTT$ is often based on 2-attribute experiments. Thus one should weigh the $VTT$ according to the number of people who receive a TCR in the general population. The reason is, that the utilisation of e.g. a new piece of road is also dependent on people who receive a travel cost reimbursement. In real life, these respondents are not instructed to ignore the fact that they receive a reimbursement, and their subsequent choices will thus be made under the influence of them receiving a travel cost reimbursement.

We evaluated such a cost-benefit analysis by approximating to what extent the SCBA of the *Ring Utrecht* project would change as a consequence of taking travel cost reimbursement into account. We found substantial differences in this specific case where the total travel time benefits could differ as much as 3.72%, and we found a heavy dependence on whether the estimates for a partial- or a full Travel Cost Reimbursement were adopted. This further strengthens the argument that future cost-benefit analysis should be based on models which incorporate travel cost reimbursement effects on the $VTT$ as the results could be even more dramatic for other projects due to; (1) the ratio of commuters to other road users changing, (2) the mode of transportation not being limited to cars.

When further zooming out, we observed in our samples that the number of commuter respondents who received a travel cost reimbursement is substantially higher in the NL2022 dataset compared to the UK2014 dataset (73% versus 19% respectively for Car Commute). We argue that this difference in the number of reimbursement respondents across both datasets is probably because there is no tax-free travel cost reimbursement in the UK as we were unable to find any other reasonable explanations. Uncovering to what extent a tax-free travel cost reimbursement impacts the number of employers who receive a $TCR$ was naturally not the goal of either the Dutch [7] or the UK [21] study but it does provide an interesting perspective. Since the $VTT$ is significantly higher for reimbursed respondents, increasing the number of people in the population who receive a travel cost reimbursement also increases the average $VTT$ in the population. Which in turn tips the scale of cost-benefit analysis more often to the benefit side as travel time savings are valued more. This subsequently increases the overall efficiency of a transport network as for more projects the benefits outweigh the costs. This is an interesting policy effect as a consequence of the number of people who receive a travel cost reimbursement in the population, which is likely determined by there being any tax benefits.

One another note; as TCR seems to affect the $VTT$ predominantly for commuters, evaluating a SCBA in which the share of commuters is larger than that of the *Ring Utrecht* SCBA will make the impact of TCR even greater. A typical policy where the ratio of commuters in proportion to the total traffic flow is very high is during rush hour. Thus an SCBA to evaluate a policy aimed at for example reducing traffic flow during rush hour can be affected to a great extent. A controversial policy idea focussed on exactly that is that of congestion charging or ”rekeningrijden” in Dutch. It aims to reduce the number of vehicles in congested areas or congested periods by imposing time or location-based fees on drivers. As we found that TCR significantly affects the VTT for commuters the perceived benefits of congestion charging (e.g., reduced travel times) might be heavily undervalued if TCR is not considered. Policymakers should therefore adjust weighted $VTT$'s in such an SCBA to reflect the proportion of reimbursed commuters, ensuring that the benefits of congestion charging are accurately assessed. When we assess the congestion charging policy from a more broader perspective we can uncover more mechanisms where our results are relevant. If a substantial portion of drivers are reimbursed for their travel costs, their higher VTT might make them less sensitive to congestion charges, as they perceive their time as more valuable and may be willing to pay the charge. Thus we might see less reduction in traffic volumes from congestion charging than anticipated as the effectiveness of the policy could be moderated by the extent to which commuters receive TCR. Besides, employers might opt to also reimburse the congestion charge further reducing the effectiveness of the policy. Policymakers could consider varying congestion charges based on whether or not drivers are reimbursed for their travel costs or implement complementary measures to ensure that commuters with a travel cost reimbursement also ”feel” the financial stimulus to change their travel behaviour. Our findings also suggest a potential approach to ensuring fairness in congestion charging. By offering a form of travel cost reimbursement from the government to low-income road users, we could increase their $VTT$ to align more closely with that of high-income road users. This adjustment would effectively level the playing field, ensuring that all individuals experience the congestion charging stimulus on equal terms, thereby enhancing the policy's equity.

### *Limitations*

One limitation of the study was the definition of reimbursement categories, where the percentages were chosen through reasoning semantically about what percentage of reimbursement can be defined as a full Travel Cost Reimbursement. To recall; respondents with a reimbursement of [0%, 15%] were set as respondents receiving no travel cost reimbursement, (15%, 85%) were set as respondents receiving a partial travel cost reimbursement and [85%, 100%] was defined as a full travel cost reimbursement. It would be interesting to explore how the results might have changed as a function of these definitions, potentially providing more nuanced insights. One such alteration was briefly evaluated where a comparison was made between the bin as defined in this study and the bins defined in a more trivial manner, where any fraction of reimbursement other than 0% or 100% is denoted as a partial travel cost reimbursement. We observed that the estimated interactions did not differ dramatically. Nonetheless, a robustness analysis which evaluates the reimbursement effect on the $VTT$ whilst varying the reimbursement categories would be a valuable contribution.

Furthermore, as our research question was about the difference in $VTT$ between nTCR and TCR respondents, we focused on the relative difference in $VTT$ between the two groups. However, we later observed that when allowing for scale heterogeneity the reference $VTT$, thus the $VTT$ for the nTCR group sometimes also changed slightly. This effect was implicitly taken into account for the sample enumeration procedure as the estimated $VTT$ values themselves are used and not solely the difference thus it will not have affected the results for the *Ring Utrecht* SCBA. Nonetheless, it would be an interesting avenue for further analysis from a methodological viewpoint.

Moreover, as mentioned earlier ODIN2023 does not provide information on the magnitude of the travel cost reimbursement. Thus for the evaluation of the textitRing Utrecht project, we used the parameter estimates for a full Travel Cost Reimbursement and applied the corresponding $VTT$ values to the respondents who reported receiving *any* travel cost reimbursement in ODIN2023. Whenever the yearly ODIN survey also reports on the magnitude of the reimbursement it will be possible to distinguish between a partial and a full travel cost reimbursement which might alter the impact on the *Ring Utrecht* SCBA.

### *Further Research*

This study opens several avenues for further research, all aimed at deepening our understanding of how travel cost reimbursement (TCR) influences the Value of Travel Time (VTT) across different contexts and populations.

An intriguing yet unanswered question is whether a travel cost reimbursement increases the $VTT$ to the same extent for low-income individuals as it does for high-income individuals. This could be studied through a similar analysis as was performed in this study, only now allowing for a heterogeneous interaction across income groups. This also leads to the methodological implication listed earlier; namely that for future national $VTT$ surveys it could be beneficial to retrieve more specific information on what type of travel cost reimbursement respondents receive. One can imagine that reimbursement effects on the $VTT$ might be very similar for low- and high-income respondents whenever their travel costs are *directly* and in full paid by the employer. However, when people first have to pay for the travel cost themselves and are only thereafter compensated a fixed amount, a financial incentive to make cheaper choices remains, and a different interaction for low- and high-income groups might show.

Another avenue for further research would be to ensure whether the observed difference in the $VTT$ is a reflection of true preferences and not limited to a stated preference study, we suggest performing a revealed preference study with the same goal as this study. This might even be possible on existing data, e.g. by exploring smart-card data. However, we expect that finding a significant interaction between the $VTT$ and travel cost reimbursement will be very difficult if not impossible due to multicollinearity. In other words; the number of occurrences in real life where one can choose between a slower but cheaper route is very limited.

Furthermore, another avenue could be to confirm whether partial travel cost reimbursement indeed affects the $VTT$ to a lesser extent than a full travel cost reimbursement. To do so in a future study it would be very valuable to also record the *type* of reimbursement respondents receive to ensure the difference between a partial- and a full travel cost reimbursement is not due to another confounding variable; the type of travel cost reimbursement.

Additionally, further research could be to ask respondents whether they usually receive a travel cost reimbursement for their day-to-day commute but ask them to keep a trip in mind which is *not* reimbursed. Thus asking respondents with a reference journey which is not reimbursed if they receive a TCR for commuting. This approach would enable us to evaluate whether *usually* receiving a TCR affects the $VTT$ even if the respondent thinks of another reference journey. In other words; this would enable us to evaluate whether a travel cost reimbursement changes a person's $VTT$ for all trips or if a person's $VTT$ is only affected for the commute trip to which the reimbursement applies. Because even if respondents are thinking of a non-reimbursed trip, the habit of receiving reimbursement could influence their perception of costs and benefits, potentially altering their VTT.

An additional evaluation could consider whether a reference-based design impacts the extent to which travel cost reimbursement effects occur. [4] Specifically, it is plausible that a choice set which is based on the reference (reimbursed) trip influences the extent to which individuals with travel cost reimbursement take their reimbursement into account as such a choice set more closely resembles their reimbursed reference trip.

Another interesting avenue for further research would be to investigate whether respondents with a lower-scale parameter are also respondents who make faster choices. Put simply; does scale correlate with response time? Such a relation was already found by [17] for externally imposed response time limitations but to our knowledge not yet for response time without externally imposed limits. If a strong and consistent relation is found this could be a nice check for researchers when assuming homoscedasticity of their data. A strong correlation between the two would also be in line with our expectation that task motivation might be one of the causes of scale differences as we would expect lower motivated respondents to make choices more quickly. Furthermore, following the same arguments, respondents with a $TCR$ could also show more non-trading

[4]A reference-based design in short, is where the reference cost- and time values of the respondent are included in either one or both of the alternatives. Thus there is always some value in the choice set which is the same as the reference journey which was described by the respondent.

behaviour. Non-trading behaviour in this case would mean that respondents always choose the alternative with the lowest travel time.

One last potential avenue for further research is the exploration of the interaction parameters between reimbursement bins and the $VTT$ by assigning them a random distribution and estimating a mixed Logit Model. This approach may allow for the analysis of heterogeneity within these interaction parameters since it might be possible that some respondents assume that their reimbursement may not apply to the tradeoffs presented and therefore do not show an increased $VTT$. However, applying this procedure is guaranteed to result in a higher model fit, as a more flexible utility specification is set. Thus evaluating whether this method yields greater predictive power will be difficult.

## ACKNOWLEDGMENT

First and foremost, I would like to thank my parents, who have given me a strong emotional foundation, enabling me to cope with setbacks and celebrate successes. Additionally, I would like to thank Roelof Akkermans for sparking my enthusiasm for Technology and Physics, Bert Rijpkema for his support during my final year at the Linde College, allowing me to add Mathematics B to my final list at the last minute and start my Bachelor's in Mechanical Engineering at TU Delft, Julie for her support and love, which helped me endure long days, my housemates, who were always willing to listen to my musings and thoughts and provided critical yet constructive feedback, my friends for remaining friends despite the limited contact we may have had during my graduation, of course, my graduation supervisors from TU Delft: Dr. Eric Molin and Prof. Oded Cats, for their guidance, positive approach, and constructive feedback, Jeroen Muller and Marco Kouwenhoven from Significance, for providing a fantastic graduation spot with an abundance of expertise, and Jeroen in particular for his consistently critical feedback.

# REFERENCES


[1] J. Dupuit, "De la mesure de l'utilité des travaux publics," *Annales des ponts et chaussées*, vol. 8, no. 2, pp. 332–375, 1844, translated by R.H. Barback as "On the Measurement of the Utility of Public Works", in International Economic Papers, 1952, no. 2, pp. 83–110.

[2] G. Romijn and G. G. Renes, *Algemene leidraad voor maatschappelijke kosten-batenanalyse*, 2013.

[3] U. Liebe, J. Meyerhoff, M. Kroesen, C. Chorus, and K. Glenk, "From welcome culture to welcome limits? uncovering preference changes over time for sheltering refugees in germany," *PloS one*, vol. 13, no. 8, p. e0199923, 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0199923

[4] D. Mcfadden, "Computing willingness-to-pay in random utility models," *University of California at Berkeley, Econometrics Laboratory Software Archive, Working Papers*, 3 1996.

[5] J. J. Louviere and G. Woodworth, "Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data," *Journal of Marketing Research*, vol. 20, no. 4, p. 350, 1983.

[6] J. J. Louviere and D. Hensher, "Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event," *Journal of Consumer Research*, vol. 10, no. 3, pp. 348–361, 1983.

[7] M. Kouwenhoven, J. Muller, S. Thoen, J. Willigers, and G. de Jong, "Values of time, reliability and comfort in the netherlands 2022 new values for passenger travel and freight transport," 2022.

[8] A. Bachaus, "Evaluatie onbelaste reiskostenvergoeding," 2023. [Online]. Available: www.ce.nl

[9] P. C. Bueno, J. Gomez, J. R. Peters, and J. M. Vassallo, "Understanding the effects of transit benefits on employees' travel behavior: Evidence from the new york-new jersey region," *Transportation Research Part A: Policy and Practice*, vol. 99, pp. 1–13, 5 2017.

[10] M. Baldassare, S. Ryan, and C. Katz, "Suburban attitudes toward policies aimed at reducing solo driving," 1998.

[11] C. Raux, Y. Croissant, and D. Pons, "Would personal carbon trading reduce travel emissions more effectively than a carbon tax?" *Transportation Research Part D: Transport and Environment*, vol. 35, pp. 72–83, 3 2015.

[12] C. Evangelinos, S. Tscharaktschiew, E. Marcucci, and V. Gatta, "Pricing workplace parking via cash-out: Effects on modal choice and implications for transport policy," *Transportation Research Part A: Policy and Practice*, vol. 113, pp. 369–380, 7 2018.

[13] E. Zawojska, A. Bartczak, and M. Czajkowski, "Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences," *Journal of Environmental Economics and Management*, vol. 93, pp. 63–84, 1 2019.

[14] M. Ding, R. Grewal, and J. Liechty, "Incentive-aligned conjoint analysis," *Journal of Marketing Research*, vol. 42, pp. 67–82, 2005. [Online]. Available: https://doi.org/10.1509/jmkr.42.1.67.56890

[15] J. Swait and W. Adamowicz, "Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice," *Organizational Behavior and Human Decision Processes*, vol. 86, pp. 141–167, 2001.

[16] S. Jang, S. Rasouli, and H. Timmermans, "The effect of task complexity on stated choice processes: The moderating role of cognitive ability," *Transportation Research Part E: Logistics and Transportation Review*, vol. 163, 7 2022.

[17] C. Chen, C. Chorus, E. Molin, and B. van Wee, "Effects of task complexity and time pressure on activity-travel choices: heteroscedastic logit model and activity-travel simulator experiment," *Transportation*, vol. 43, no. 3, pp. 455–472, 2016.

[18] S. Hess, J. M. Rose, and D. A. Hensher, "Asymmetric preference formation in willingness to pay estimates in discrete choice models," *Transportation Research Part E: Logistics and Transportation Review*, vol. 44, pp. 847–863, 2007.

[19] A. Tversky and D. Kahneman, "Loss aversion in riskless choice: A reference-dependent model," pp. 1039–1061, 1991. [Online]. Available: https://about.jstor.org/terms

[20] S. Hess and K. Train, "Correlation and scale in mixed logit models," *Journal of Choice Modelling*, vol. 23, pp. 1–8, 6 2017.

[21] ARUP, "Provision of market research for value of travel time savings and reliability: Phase 2 report," Department for Transport, London, United Kingdom, Tech. Rep. Issue 14 August 2015, August 2015, in partnership with Ove Arup & Partners Ltd, Institute for Transport Studies (ITS), University of Leeds, and Accent. Contributors: Stefan Sanders, Efie Drivyla, Jeremy Bourdon, Richard Batley, Manuel Ojeda Cabral, Phani Kumar Chintakayala, Charisma Choudhury, Andrew Daly, Thijs Dekker, Anthony Fowkes, Stephane Hess, Daniel Johnson, James Laird, Peter Mackie, Mark Wardman, John Bates, Michiel Bliemer, Maria Börjesson, John Parkin, Tom Worsley, Chris Heywood, Miranda Mayes, Rachel Risely, Nancy Curzon, Rob Sheldon.

[22] D. McFadden and K. Train, "Mixed mnl models for discrete response," *Journal of Applied Econometrics*, vol. 15, no. 5, pp. 447–470, 2000.

[23] J. Zhang and W. L. Adamowicz, "Unraveling the choice format effect: A context-dependent random utility model," 2011.

[24] D. A. Hensher, "Towards a practical method to establish comparable values of travel time savings from stated choice experiments with differing design dimensions," *Transportation Research Part A: Policy and Practice*, vol. 40, pp. 829–840, 2006.

[25] J. R. Deshazo, "Designing choice sets for stated preference methods: The effects of complexity on choice consistency," *Journal of Environmental Economics and Management*, vol. 44, 2002.

[26] B. G. C. Dellaert, J. D. Brazell, and J. J. Louviere, "The effect of attribute variation on consumer choice consistency," pp. 139–147, 1999.

[27] K. Train and M. Weeks, "In preference space and willingness-to-pay space," in *The Economics of Non-Market Goods and Resources*, ser. The Economics of Non-Market Goods and Resources. Berlin, Heidelberg: Springer, 2015, vol. 6. [Online]. Available: https://link.springer.com/chapter/10.1007/1-4020-3684-1_1

[28] M. Bradly and A. Daly, "Estimation of logit choice models using mixed stated-preference and revealed-preference information see profile," 1996. [Online]. Available: https://www.researchgate.net/publication/230663774

[29] D. G. Fiebig, M. P. Keane, J. Louviere, and N. Wasi, "The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity," vol. 29, pp. 393–421, 2010.

[30] R. H. Thaler, "Mental accounting matters," pp. 183–206, 1999.

[31] K. Brosnan, N. Babakhani, and S. Dolnicar, ""i know what you're going to ask me" why respondents don't read survey questions," *International Journal of Market Research*, vol. 61, no. 4, pp. 366–379, 2019. [Online]. Available: https://doi.org/10.1177/1470785318821025

[32] A. L. Alter, D. M. Oppenheimer, N. Epley, and R. N. Eyre, "Overcoming intuition: Metacognitive difficulty activates analytic reasoning," *Journal of Experimental Psychology: General*, vol. 136, pp. 569–576, 11 2007.

[33] D. Kahneman, *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.

[34] S. Hess and J. M. Rose, "Can scale and coefficient heterogeneity be separated in random coefficients models?" *Transportation*, vol. 39, no. 6, pp. 1225–1239, 2012.

[35] S. Hess and K. Train, "Correlation and scale in mixed logit models," *Journal of Choice Modelling*, vol. 23, pp. 1–8, 2017.

[36] J. Neyman and E. S. Pearson, "On the use and interpretation of certain test criteria for purposes of statistical inference: Part i," *Biometrika*, vol. 20A, no. 1/2, pp. 175–240, 1928. [Online]. Available: https://academic.oup.com/biomet/article/20A/1-2/175/284408

[37] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.

[38] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[39] S. Hess and D. Palma, "Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application," *Journal of Choice Modelling*, vol. 32, 2019.

[40] J. H. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numerische Mathematik*, vol. 2, no. 1, pp. 84–90, 1960.

[41] IenW, "Mkba ring utrecht," Ministerie van Infrastructuur en Milieu, Tech. Rep., Aug. 2016, accessed: 2024-07-18. [Online]. Available: https://www.rwseconomie.nl/onderwerpen/veel-gestelde-vragen/documenten/rapporten/2016/augustus/augustus/mkba-ring-utrecht

[42] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.185.4157.1124

[43] PBL and CPB, "Nederland in 2030 en 2050: twee referentiescenario's-wlo," 2020, https://www.pbl.nl/sites/default/files/downloads/PBL-CPB-2020-Nederland-in-2030-en-2050-twee-referentiescenario's-WLO.pdf. [Online]. Available: https://www.pbl.nl/sites/default/files/downloads/PBL-CPB-2020-Nederland-in-2030-en-2050-twee-referentiescenario's-WLO.pdf

[44] M. Oehlmann, J. Meyerhoff, P. Mariel, and P. Weller, "Uncovering context-induced status quo effects in choice experiments," *Journal of Environmental Economics and Management*, vol. 81, pp. 59–73, 1 2017.