Helsinki University of Technology Transportation Engineering Publication 106
Teknillinen korkeakoulu Liikennetekniikka Julkaisu 106
Espoo 2003

THE TRANSFERABILITY OF TRAVEL DEMAND MODELS

An analysis of transfer methods, data quality and model estimation

Nina Karasmaa

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Civil and Environmental Engineering for public examination and debate in Auditorium R1 at Helsinki University of Technology (Espoo, Finland) on the 17th of December 2003, at 12 noon.

Helsinki University of Technology
Department of Civil and Environmental Engineering
Laboratory of Transportation Engineering

Teknillinen korkeakoulu
Rakennus- ja ympäristötekniikan osasto
Liikennelaboratorio

Inquiries and orders:
Helsinki University of Technology
Laboratory of Transportation Engineering
P.O.Box 2100
FIN-02015 HUT
Tel. +358-9-451 3791
Fax. +358-9-451 5019
E-mail: anneli.fogel@hut.fi

ISBN 951-22-6856-6
ISSN 0781-5816

Picaset Oy
Helsinki 2003

HELSINKI UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

ABSTRACT OF DOCTORAL THESIS

Author:	Nina Karasmaa		
Title:	The transferability of travel demand models, An analysis of transfer methods, data quality and model estimation		
Date:	17.12.2003	Pages: 214	Code: Yhd-71
Supervisor:	Professori Matti Pursula		Professuuri: Liikennetekniikka

The main goal of this study was to compare alternative methods of spatial transfer as a function of sample size, and identify the factors affecting the models' quality and the impreciseness of the model parameters. In addition, different test measures for studying model transferability were compared and the applicability of the traditional statistical tests, with respect to those based on the prediction accuracy of sample enumeration tests and forecasts, were assessed. The research primarily concerned the transferability of mode and destination models; however, the preciseness of the trip generation level was considered as well.

The study was mainly based on the mobility surveys conducted in the Helsinki Metropolitan Area (HMA) in 1995 and in the Turku region in 1997. The transferring procedures examined were Bayesian updating, combined transfer estimation, transfer scaling, and joint context estimation procedures. The trip groups studied were home-based work trips and other home-based trips. The studied modes were walk and bicycle, car and public transport. To explore the impact of sample size on transferring performance, model transferability was tested using three to four different sample sizes. Thus, all the transferability tests were made by using 100 bootstrap samples (resampled from the Turku 1997 dataset) for each trip group, transfer method and sample size category.

The results indicated that joint context estimation gives the best prediction performance in almost all cases. In particular, the method is useful if the transfer bias is large or only some of the coefficients are precise. The applicability of joint context estimation can be improved by viewing the coefficients as variable-oriented and emphasizing precise and imprecise coefficients differently. The models transferred by using combined transfer estimation or transfer scaling were most sensitive to the sample size and their use, therefore, requires much larger samples than the Bayesian approach or joint context estimation. In addition, note that due to repeated measurements the results based on the Bayesian method and combined transfer estimation may be strongly biased. When defining the sample size required the fact that defining mode shares precisely may require more observations than the transferring mode and the destination choice models must be taken into account.

The results also showed that statistical tests are not able to evaluate the goodness of transferred models with a high enough degree of versatility. For example two models that have totally different values for coefficients may have the same TTS. As a result, their ability to predict the effect of changes in a transportation system may differ greatly. On the whole, the differences between the best transfer methods are, in some cases, rather small, and the errors caused by the factors connected to the modelling and sample size seem to be larger than the errors caused by the model transfer itself.

TEKNISKA HÖGSKOLAN
AVDELNINGEN FÖR BYGGNADS-
OCH MILJÖTEKNIK

SAMMANDRAG AV
DOKTORSAVHANDLING

Utfört av:	Nina Karasmaa		
Ämne:	The transferability of travel demand models, An analysis of transfer methods, data quality and model estimation		
Datum:	17.12.2003	Sidoantal: 214	Kod: Yhd-71
Övervakare:	Professor Matti Pursula		Professur: Trafikteknik

Avsikten med detta arbete var att utreda olika överföringsmetoders tillämplighet för olika situationer. Överföring av modeller har traditionellt uppskattats genom att jämföra parametrar för modeller gjorda för två olika orter eller vid två olika tidpunkter med varandra. Eftersom en målsättning för överföring av modeller är att göra det möjligt att uppgöra modeller utgående från rätt små observationsmaterial, koncentrerar sig detta arbete speciellt på sampelstorlekens inverkan på modellers överföring.

I detta forskningsarbete jämfördes olika överföringsmetoder med hjälp av material insamlat i Helsingforsregionen år 1995 och Åboregionen år 1997. Dessutom uppskattades sampelstorlekens inverkan på modellernas överförbarhet. Överföringen av modellerna undersöktes för bostads-baserade arbetsresor och övriga bostadsbaserade resor. De undersökta färdmedlen var gång- och cykeltrafik, bil och kollektivtrafik. Modellernas överföring undersöktes skilt utgående från trafikalstring, färdmedelsvalsmodeller och destinationsvalsmodeller samt utgående från hela trafik-prognossystemet. De undersökta överföringsmetoderna var skalfaktormetoden, Bayes metod, den generaliserade Bayes metoden samt estimering från det kombinerade materialet från Helsingforsregionen år 1995 och Åboregionen år 1997. Dessutom användes 1997 års material från Åboregionen för att för varje sampelstorlek estimera nya modeller genom att använda modell-definitioner som tillämpats på Helsingforsregionens 1995 års material. För varje resärendegrupp testades 100 slumpmässigt valda bootstrap-sampel. De undersökta sampelstorlekarna varierade från 400 till 13900 reseobservationer. Trafikalstringen överfördes egentligen inte, utan frekvenstalen beräknades som enkla trafikalstringstabeller för varje sampel.

Resultaten visar, att estimering av det kombinerade materialet är den bästa överföringsmetoden i nästan alla situationer, i synnerhet om de tillförlitligaste variablerna estimeras skilt för varje material. Modeller estimerade med den generaliserade Bayes metoden och skalfaktormetoden är mest känsliga för variationer i sampelstorlek, vilket innebär att dessa metoder förutsätter större observationsmaterial än Bayes metod och estimering från det kombinerade materialet. Beträffande Bayes metod och i viss mån också den generaliserade Bayes metoden bör man ta i betraktande, att modellernas estimater kan vara sneda, om observationsmaterialet innefattar flera resor gjorda av samma person, vilket ofta är fallet i material insamlat i Finland. Man bör också lägga märke till att bestämmandet av färmedelsfördelningen kan kräva större sampelstorlek än överförandet av färdmedelsvals- och destinationsvalsmodellerna.

Resultaten visar, att enbart statistisk prövning inte räcker till för att uppskatta modellernas överförbarhet tillräckligt mångsidigt. Enligt de testningar som baserades på log-likelihoodmått kan modeller som uppskattats vara nästan lika bra skilja sig märkbart från varandra vad beträffar enskilda variabler, varvid också ifrågavarande modellers tidsvärden och elasticitetsverkningar avviker från varandra. Allt som allt kan man konstatera, att skillnaderna mellan olika överföringsmetoder ofta är mindre än skillnaderna förorsakade av olika sampelstorlek.

TEKNILLINEN KORKEAKOULU VÄITÖSKIRJAN
RAKENNUS- JA YMPÄRISTÖTEKNIIKAN OSASTO TIIVISTELMÄ

Tekijä: Nina Karasmaa
Aihe: The transferability of travel demand models, An analysis of transfer methods, data quality and model estimation
Päivämäärä: 17.12.2003 Sivumäärä: 214 Koodi: Yhd-71
Valvoja: Professori Matti Pursula Professuuri: Liikennetekniikka

Tämän työn tavoitteena oli selvittää eri siirtotapojen soveltuvuutta eri tilanteisiin. Mallien siirtoa on perinteisesti arvioitu vertaamalla kahdessa eri paikassa tai kahtena eri ajankohtana tehtyjen mallien parametreja toisiinsa. Koska mallien siirron yhtenä tavoitteena on mahdollistaa mallien teko kohtalaisen pienillä havaintoaineistoilla, on tässä työssä tarkasteltu erityisesti otoskoon vaikutusta mallien siirtoon.

Tutkimuksessa vertailtiin pääkaupunkiseudulla vuona 1995 ja Turun seudulla vuonna 1997 kerättyjen aineistojen avulla eri siirtotapoja sekä arvioitiin otoskoon vaikutusta mallien siirrettävyyteen. Mallien siirtoa tutkittiin kotiperäisillä työmatkoilla ja muilla kotiperäisillä matkoilla. Tarkasteltavat kulkutavat olivat kevytliikenne, auto ja joukkoliikenne. Mallien siirtoa tutkittiin erikseen matkatuotosten, kulkutapa- ja suuntautumismallien sekä koko ennusteen kannalta. Tarkasteltavat siirtotavat olivat tasokorjausmenetelmä, Bayesin menetelmä, yleistetty Bayesin menetelmä sekä yhdistetystä pääkaupunkiseudun 1995 ja Turun seudun 1997 aineistoista estimointi. Lisäksi Turun seudun 1997 aineistosta estimoitiin kullakin otoskoolla uudet mallit käyttäen pääkaupunkiseudun 1995 aineistoon sovellettuja mallimäärittelyjä. Kullakin matkaryhmällä ja otoskoolla testattiin 100 satunnaisesti valittua bootstrap-otosta. Tutkitut otoskoot vaihtelivat 400:sta 13900:aan matkahavaintoon. Matkatuotoksia ei varsinaisesti siirretty, vaan tuotosluvut laskettiin yksinkertaisina tuotostaulukkoina kullekin 100:lle otokselle.

Tulokset osoittivat, että yhdistetyn aineiston estimointi on lähes kaikissa tilanteissa paras siirtotapa varsinkin, jos luotettavimmat muuttujat estimoidaan aineistokohtaisina. Yleistetyllä Bayesin menetelmällä ja tasokorjausmenetelmällä estimoidut mallit ovat herkimpiä otoskoon vaihtelulle, joten näiden menetelmien käyttö edellyttää huomattavasti suurempia havaintomääriä kuin Bayesin menetelmä ja yhdistetystä aineistosta estimointi. Bayesin ja pienemmässä määrin myös yleistetyn Bayesin menetelmien osalta on huomattava, että mallien estimaatit voivat olla harhaisia, jos malliaineistossa on samalta henkilöltä useita matkoja, kuten Suomessa kerätyissä aineistoissa yleensä on tilanne. On myös huomattava, että kulkutapaosuuksien määrittäminen saattaa vaatia suurempaa otoskokoa kuin kulkutapa- ja suuntautumismallien siirtäminen.

Tulokset osoittivat, että pelkillä tilastollisilla testeillä ei pystytä arvioimaan mallien siirrettävyyttä riittävän monipuolisesti. Suurimman uskottavuuden arvoihin perustuvien testien perusteella lähes yhtä hyviksi arvioidut mallit saattavat yksittäisten muuttujien osalta poiketa toisistaan hyvinkin paljon, jolloin myös kyseisten mallien ajanarvot ja joustovaikutukset poikkeavat toisistaan. Kaikenkaikkiaan voitiin todeta, että erot eri siirtotapojen välillä ovat usein pienempiä kuin otoskoosta aiheutuvat erot.

PREFACE

The main goal of the thesis was to study the spatial transferability of the Helsinki Metropolitan Area model system. The initial inspiration came from Hannu Kangas, M.Sc, who had worked earlier on the same issue. The original idea of the work was to compare different transfer methods in practice. However, because the greatest advance in model transfer could have been achieved by reducing the size of the data to be collected in the new situation, we started to study the effect of sample size in model transfer as well.

The doctoral thesis was carried out at Helsinki University of Technology, Laboratory of Transportation Engineering. The supervisor for this thesis was Professor Matti Pursula. I would like to express my deepest thanks to him for his support and advices during my studies. Without him this research would not have been possible. The financial support of the Academy of Finland and the Nordic Academy for Advanced Study made it possible for me to visit the Royal Institute of Technology in Stockholm and the University of Leeds. Special thanks go to Dr. Farideh Ramjerdi for her valuable criticism and to Professor Lars-Göran Mattson for his support and encouragement during my visit. The visit to Leeds University gave me the opportunity to learn more about the repeated measurement issue. Special thanks go to Dr. David Watling and to professor Andrew Daly for their help during the visit. I would also like to thank Professor Takashi Uchida for his help during the year he visited HUT.

Professor Pertti Laininen's comments on the Bootstrap technique and on the statistical and mathematical aspects of my work have been extremely helpful. I would also like to express my thanks to Jari Kurri M.Sc for his co-operation on the model projects carried out in the Helsinki metropolitan area and in the Turku region and the many comments he made on my reports.

Thanks should also go to Helsinki metropolitan area council, the city of Turku and to Ministry of Transport and Communications Finland, whose projects made it possible for me to finish this research. The work was also supported by The Foundation for Financial Aid at Helsinki University of Technology, the Henry Ford Foundation, TES (Stiftelsen för teknikens främjande), The Academy of Finland and the Nordic Academy for Advanced Study.

My thanks also go to Professor Otto A. Nielsen and Dr. Karin Brundell-Freij for their thorough prereview of my work. The English language of my work has been checked by Ruth Vilmi, M.Sc and Anna Reiter, B.Sc. and I would like to thank them for this.

Finally, my deepest thanks go to my husband Risto and to my daughter Jonna for their continuous support during my work. I would also like to thank all my friends and relatives who have supported me over the last few years.

Espoo, Finland
November 26, 2003 Nina Karasmaa

ALKUSANAT

Tämän väitöskirjan tavoitteena on ollut selvittää pääkaupunkiseudulla vuonna 1995 tehdyn sisäisen liikenteen mallijärjestelmän alueellista siirrettävyyttä. Idean aiheeseen sain DI Hannu Kankaalta, joka itse aikanaan tutki samaa aihepiiriä. Työ lähti liikkeelle erilaisten siirtomenetelmien vertailusta. Koska mallien siirron tavoitteena on nimenomaan vähentää aineiston keräämisestä aiheutuvia kuluja, työn toiseksi pääteemaksi muodostui otoskoon vaikutusten tutkiminen mallien siirrossa.

Työ on tehty Teknillisen korkeakoulun liikennelaboratoriossa vuosina 1996-2003, ja sitä on valvonut ja ohjannut professori Matti Pursula. Kiitän lämpimästi professori Pursulaa saamistani neuvoista ja opastuksesta työni eri vaiheissa. Ilman häntä työ tuskin olisi valmistunut koskaan. Suomen Akatemian ja NorFa:n tuen ansiosta minulla oli työni aikana mahdollisuus vierailla myös Tukholman teknillisessä korkeakoulussa (KTH) sekä Leedsin yliopistossa. KTH:lta haluaisin erityisesti kiittää TkT Farideh Ramjerdiä hänen antamastaan arvokkaasta välikritiikistä sekä professori Lars-Göran Mattsonia hänen antamastaan tuesta ja kannustuksesta vierailuni aikana. Vierailu Leedsin yliopistossa toi tärkeää lisätietoa aineiston luotettavuuden analysointiin liittyvissä kysymyksissä. Erityisesti haluan kiittää PhD David Watlingia ja professori Andrew Dalyä heidän avustaan vierailuni aikana. Lisäksi haluan kiittää professori Takashi Uchidaa hänen Suomen vierailunsa aikana antamastaan avusta.

Professori Pertti Lainista kiitän hänen Bootstrap-otantaa ja tilastollista testausta koskevista kommenteistaan. Haluan myös kiittää DI Jari Kurria yhteistyöstä sekä pääkaupunkiseudun että Turun mallien laadinnassa sekä hänen julkaisuihini esittämistään kommenteista.

Kiitokset myös Pääkaupunkiseudun yhteistyövaltuuskunnalle, Turun kaupungille ja liikenne- ja viestintäministeriölle, joiden toimeksiannot mahdollistivat työn tekemisen. Työtä ovat tukeneet myös Teknillisen korkeakoulun tukisäätiö, Henry Fordin säätiö, Tekniikan Edistämissäätiö, sekä Suomen Akatemia ja NorFa (Nordisk Forsker-utdanningakademi).

Kiitokset työn esitarkastuksesta kuuluvat professori Otto A. Nielsenille Tanskan teknillisestä korkeakoulusta sekä TkT Karin Brundell-Freij'lle Lundin teknillisestä korkeakoulusta. Työn englanninkielisen kieliasun ovat tarkastaneet maisteri Anna Reiter sekä maisteri Ruth Vilmi. Kiitokset heille siitä.

Haluan myös tässä yhteydessä kiittää perhettäni, sukulaisiani ja kaikkia niitä ystäviä ja työtovereita, jotka ovat edesauttaneet työn valmistumista.

Espoossa
26.11.2003 Nina Karasmaa

CONTENTS

1 INTRODUCTION

1.1 Background

Travel forecasting is usually based on the presumption that travel demand is based on firm and predictable behaviour patterns. Analogously, the expectation that travel demand models are transferable is based on the idea that an individual's travel behaviour can be described by the individual's personal characteristics, their socio-economic backround, the transportation system, and that the behavioural process is relatively constant over time and space.

Transferability has usually been described in terms of temporal or spatial transferability. Temporal transfer is the application of a model estimated at one point in time for the prediction of behaviour in the same spatial environment at another point in time. Temporal transfer is employed and temporal transferability is implicitly invoked whenever models estimated on historic data are used to predict the future. Spatial transfer is the application of a model estimated in one location (estimation context) for prediction of behaviour in a different spatial environment (application context). Spatial transferability is explicitly invoked when models developed in one geographic region are used to make predictions in another *(Koppelman and Rose 1983)*.

Transferability has been a matter of considerable practical and theoretical interest because the transfer of a previously estimated model to a new application context can reduce or eliminate the need for large data collection and model development effort in the application context. Thus, several studies have been conducted to assess the effectiveness of model transfer from one context to another. Some of these studies have examined spatial transfer *(Atherton and Ben-Akiva 1976, Galbraith and Hensher 1982, Koppelman et al. 1985, McCoomb 1986, Gunn and Pol 1986, Tretvik 1989, Abdelwahab 1991, Algers et al. 1994, Sermons 2000)*, while others have examined the temporal transfer of these models *(Talvitie and Kirshner 1978, Train 1978, McCarthy 1982, Hague Consulting Group 1990, Badoe 1994, Badoe and Miller 1995a, Badoe and Miller 1995b, Badoe and Miller 1998, Walker et al. 1998, Elmi et al. 1999)*.

The transferability of travel demand models is commonly perceived as an issue of the equality of the model coefficients in two contexts. However, assessing model transferability only on the basis of the set of model parameters being equal in the two contexts is a very stringent test and it is one that is unlikely to be met since no model is perfectly specified. Consequently, all models are in principle context dependent. A more pragmatic evaluation of transferability is achieved by assessing the extent of useful information provided in an application context by transferred models *(Badoe 1994)*.

In Finland, model transferability was not studied much before the 1990's, when the first transferability project started at Helsinki University of Technology *(Karasmaa 1995, Karasmaa and Pursula 1995)*. However, travel demand models have been produced since 1967, when the first models were estimated for the Helsinki Metropolitan Area, and two years later also for the Lahti and Tampere regions. Since then travel demand models have been estimated in approximately 15 municipalities. Although, these models may be precise and well made, the existence of a uniform modelling system (for local purposes) has been lacking, thereby making it difficult to compare and evaluate results.

The goal of our model transfer project, started in 1994, has been to examine how the model system developed in the Helsinki Metropolitan area is to be applied in other areas in Finland. During the project, different transfer methods have been studied, and the key issues causing differences in modelling and data collection have been investigated. In connection with the study the regional models have been re-estimated to the Helsinki Metropolitan area, and totally new models have been estimated to the Turku and Vaasa regions. HUT has also been involved with projects concerning data quality and harmonizing both in Finland (*Kurri and Karasmaa 1999, Kivari et al. 2000*) and at the European level *(Dateline 2003)*. All this research has contributed to the implementation of this study even though the results are mainly based on the experiences gained from the Helsinki Metropolitan Area (HMA) and from the Turku region.

1.2 History and problems

There are two main issues, which affect the results of tests of model transferability. The first one is the difference between the coefficients estimated from the estimation and application context data. This difference is traditionally known as transfer bias. The other is the statistical preciseness of model parameters, which refers to the repeatability of the estimation results *(Richardson et al. 1995)*. Thus the final results depend on both the quality of the models in the estimation and application context and the differences in the traffic system, and on the travel behaviour pattern.

In literature, four different transfer methods are presented. These are:

- transfer scaling,
- Bayesian method,
- combined transfer estimation and,
- the joint context estimation.

The transfer methods differ from each other in emphasizing the estimation and the application context data. Errors expressed in model transfer can be caused by sampling errors or errors that are caused by differences between the coefficients in the initial and the final stages. The problem in model transfer is that we do not know how much the difference between the coefficients estimated in estimation and application contexts is caused by the random variation of the model parameters, and how much is caused by the difference in travel behaviour.

Model transferability has been extensively studied, particularly in the Nordic countries, the USA, and Canada. The starting point has typically been the testing of the model parameter's equality in two contexts. What this means, is that the conclusions have been based on the comparison of the model coefficients in two contexts and no actual transfer has been made in these studies. The most popular transfer methods have been transfer scaling and the Bayesian method. In 1987, Ben-Akiva and Bolduc presented a new method, known as the combined transfer estimation method. This method can be viewed as an extension of the Bayesian method. When transferring using any of these three methods, only application context data is required (and the coefficients based on the estimation context). The first time both estimation and application context datasets were used together in the transfer process, was in 1994, when Badoe and Miller presented their work *(Badoe 1994, Badoe and Miller 1995a, Badoe and Miller 1995b).* In their research, the authors compared the joint context estimation method with the other three transfer methods.

On the whole, the results relating to model transfer have been controversial. Many studies have shown that model transfer is possible when the estimation context models are well defined and the data quality is good. On the other hand, especially when studying spatial transferability, the opposite results have often been reached. One reason for this may be that the conclusions, in most cases, have been drawn based on the testing of the model parameter's equality in two contexts. Thus the results cannot be fully generalized, but rather are dependent on the similarities of the datasets in two contexts. The basic assumption, in many cases, has been that model transfer is only possible if the coefficients in the estimation and application contexts are quite similar. The possibilities of transfer bias control (the difference in the true parameters between the two contexts) have rarely been explored. In addition, the importance of the impreciseness of the coefficients has not been analyzed. Thus, conclusions have, in most cases, been based on statistical tests only rather than any reference to the performance of models in practice.

The first studies, in which researchers consciously committed themselves to how the difference between estimation and application contexts should be considered, were not published until the end of the 1980’s *(Ben-Akiva and Bolduc 1987, Badoe 1994, Badoe and Miller 1995a, Badoe and Miller 1995b, Badoe and Miller 1998)*. Prior to this methods (e.g. transfer scaling) which took into account the transfer bias were used; however, the main thrust of the research had been to study the model parameters similarity in two contexts. As a result of this, almost all the previous research has been conducted using the whole set of mobility survey data without studying the effects of the sample size. This has been done despite the knowledge that the greatest advance in model transfer could have been reached by reducing the size of the data to be collected in the new situation.

1.3 Model transfer experiments in Finland

The main reason for the lack of Finnish experiments with model transfer has been a lack of appropriate models and suitable data. The study of temporal or spatial transferability requires an appropriate model structure based on reliable data for the base situation. In addition, reliable data is required for the transfer of this model structure in time or space. The appropriate data for model transfer in Finland now exists for the Helsinki Metropolitan Area. In 1988, a comprehensive travel survey was conducted in the area *(YTV 1990b)*. Based on this survey travel demand models *(Pursula and Kanner 1992)*, that were used in traffic forecasting and evaluating the effects of land use, traffic system and policy making were estimated. In addition, a new travel survey was conducted in 1995 *(Kaartokallio 1997)* and new travel demand models for the Helsinki Metropolitan Area were estimated based on this survey *(Karasmaa et al. 1997)*. Similar travel surveys were conducted in the Turku, Tampere, Jyväskylä, Vaasa, Oulu and Rovaniemi regions in the 1990's.

The current model system used in the HMA and in most cities in Finland is the traditional four step model, in which the travel modelling has been divided into four different sub models, namely: trip generation, destination choice, mode choice and route choice. The process includes feedback between the last three steps *(YTV and Liikenneministeriö 1990a, Pursula and Kanner 1992, Karasmaa et al. 1997)*. Steps two and three have been modelled using discrete choice models. This means that every individual chosen in the sample has a limited number of alternatives and from these alternatives he or she will choose the one that will maximize his or her utility. Travel forecasting is based on the presumption that the distribution of travel demand is based on firm and predictable behaviour. The most common model type for individual choices is a logit formulation, which has also been used in this thesis to model the mode and destination choices.

Research into the transferability of the travel demand models began in Helsinki University of Technology in 1994. The study was started with a literature review *(Karasmaa and*

Pursula 1995). In addition, the spatial transferability of the Helsinki Metropolitan Area models was studied. The empirical study was based on the data collected in the Oulu region in 1989 and examined the transferability of mode choice models by using the transfer scaling method *(Karasmaa 1995).* Later on, spatial transferability was also studied in the Turku region *(Karasmaa 2001).* Research into the temporal transferability of the Helsinki Metropolitan Area models was started for the first time in 1995. The research was based on the data collected in the Helsinki Metropolitan Area in 1981 and 1988 *(Karasmaa 1996a, Karasmaa and Pursula 1997)*. Temporal transferability was next studied in 1996 based on the data collected in the Helsinki Metropolitan Area in 1988 and 1995 *(Karasmaa 1998, Kurri and Karasmaa 1999)*. In both cases the transferability of the whole four-step model system was studied and different transfer methods were compared to one another. A licentiate thesis on model transferability was submitted to HUT in 1996 *(Karasmaa 1996b)*.

1.4 Goals of the thesis

The purpose of this thesis is to compare alternative methods of spatial transfer as a function of sample size and develope a new approach to joint context estimation. The purpose is also to identify the factors affecting the models' quality and the impreciseness of the model's parameters. The research primarily concerns the transferability of mode and destination models; however, the preciseness of the trip generation level is considered as well.

Specifically, the main issues are:

- the transferability of the mode and destination choice models and the corrections that are needed for the transfer,
- the amount of data required in the application context,
- the application areas of different transfer methods,
- the improvement of the efficiency of transfer methods, and
- ways in which model transferability should be tested.

The results presented here are mainly based on information obtained from the Helsinki Metropolitan Area (HMA) and the Turku region in Finland. The HMA travel database is used for the estimation of the models which are to be transferred. The database for the Turku region represents the application context to which the HMA models are transferred. The model transfer is studied using two different trip groups, namely home-based work trips (HBW) and other home-based trips (OHB). This makes it possible to study different kinds of combinations related to the model parameters' preciseness and the transfer bias. Some special issues are also considered through the use of simulated data.

1.5 Summary of contents

The dissertation is divided into seven main chapters. Chapter 2 reviews the history and problems surrounding model transferability. The issues associated with data quality are also considered in this chapter. Chapter 3 introduces the theory of transferability and the logit models. Four different transfer methods are presented in this chapter as well as the goodness of fit measures. Chapter 4 presents the methodology, and the data and models used in this study. The factors affecting the models' quality and the impreciseness of model parameters are discussed in Chapter 5.

Chapter 6 presents the case study of the spatial transferability of the Helsinki Metropolitan Area models. Different transfer methods are compared to each other, and the effect of the sample size is examined. In addition, one special issue, which may affect the generalization of the results, is studied. This is the problem caused by the correlation between the answers provided by the same individual (the problem of repeated measurements). General conclusions are drawn in Chapter 7.

2 LITERATURE REVIEW OF THE TRANSFERABILITY OF TRAVEL DEMAND MODELS

2.1 Temporal and spatial transferability

Disaggregate traffic forecast models are based on the idea that an individual's travel behaviour can be described by the individual's preferences, and this behavioural process is relatively constant over time and space. In model transfer the issue is the general validity of that description of individual's behaviour. Model structure and accuracy is determined by the purpose of the models and by the availability of data. To be transferable from the point of view of estimation, model parameters should be well specified and as precise as possible. In addition, the model theory and the model structure must be consistent with this premise.

There are two different approaches when looking at model transfer. In the first one, model transferability is studied statistically by comparing models made in different contexts to each other (e.g., the Helsinki Metropolitan Area and the Turku region). The hypothesis is that the behavioural parameters of all models are identical. The basic idea is to study whether this hypothesis holds true.

The second approach is to actively accept that there are some differences in treating longer periods, or models estimated in different areas. These differences can be caused by the differences in the traffic system or differences in travel behaviour, i.e., how individuals evaluate different things. This viewpoint of assessing model transferability is adopted in this thesis. The basic idea of the research is to study how the transfer bias caused by these differences can be controlled, without estimating an entirely new model system, or collecting large amounts of new data (which are needed to estimate "perfect" new models).

According to the *Hansen (1981)* and *Brand and Cheslow (1981)* the model transferability can be classified as follows:

- the transferability of broad behavioural postulates such as utility maximization,
- the transferability of a mathematical model class,
- the transferability of model specifications,
- the transferability of model coefficients.

By studying the transferability of logit models, it is assumed that the first two conditions are fulfilled and only the last two conditions have to be tested.

Regardless of whether existing models are to be transferred, or new models are to be estimated, the study of model transferability requires similar travel behaviour surveys to be made in both the application context area and the estimation context area (excluding naive transfer). This similarity means that the same variables are to be measured for model purposes. In addition, similar survey methods are also recommended. This issue is discussed in Chapter 5. The difference in the process in estimating new models, or transferred models, is the requirement imposed by the quality and quantity of data. Pursula and *Widlert (1990)* have assessed that the estimation of new models requires two to four times more observations than the transfer of models, as long as the transfer is done with great care *(Pursula and Widlert 1990).*

2.2 The importance of source data to model transfer

Mobility surveys are usually performed for two reasons: monitoring of the travel behaviour of the population and estimating models that are used to understand the past and present, and to predict the future. Travel behaviour is commonly studied via mobility surveys in which a trip diary and personal background are observed.

When testing model transferability, it is important to be able to control how well the survey methods correspond to each other in the estimation and application context. The data should be collected in a standardized way, and consistent model structures should be used in all phases. In addition, the questionnaires should be designed in such a way that the questions can be interpreted unambiguously.

Three popular forms of survey relating to the collection of information are travel diaries where we only record trips, activity diaries where we mostly record out-of-home activities, and time-use diaries where we record almost everything we do in a day inside and outside the home, with the exception of certain highly private activities. In this thesis we are looking at model transferability based on the data collected through the use of simple one and two-day travel diaries. We have collected repeated cross-sectional data rather than panel-data, which is very popular at the moment.

There are two basic methods of collecting data relating to the individual preferences which form the basis of behaviour. The method used in this study is known as the Revealed Preference method (RP). In this method information about travel behaviour is collected in real situations. The other method that can be used is to collect Stated preference (SP) data, a method which is based on responses to hypothetical travel situations in a survey context. SP-data is normally used in cases where observed choice behaviour is not in itself adequate enough to model the context of interest. SP data may be used, for example, when one is interested in types of travel alternatives or characteristics which do not yet exist, or in qualitative attributes which are very difficult to measure in real situations. The statistical efficiency can also be improved by using SP data. Combined RP and SP data have been used in order to exploit the relative advantages of the different data sources and obtain more reliable parameter estimates than those estimated from a single data source *(Ben-Akiva and Morikawa 1990).*

Both RP and SP studies can be carried out as a household or as a individual interview. In a household interview the sample is taken from the households in the area and all the members of the households are interviewed. By using personal interviews all the people living in the area generate the population set of data. The interview can be carried out by using different methods, with the most common of these being postal questionnaires and telephone interviews, although the seldom used personal interview is the most accurate way of gathering information.

Postal and telephone interviews have been analysed in several studies. For example, *Brög and Meyburg (1981)* indicate, that the use of postal questionnaires increased the response rate of mobile individuals, beyond that of telephone interviews. In contrast, participation in both personal interview and a telephone interview is largely determined by accessibility. More mobile people are more difficult to reach in personal and telephone surveys and as a consequence they show a higher non-response rate. On the other hand, less mobile people may be reluctant to answer, if they have not made any trips.

In a study made in the Oslo, Norway in 1990, there was a 40 percent increase in the reported number of trip generations due to the use of telephone interviews as opposed to postal questionnaires. It was thought to be due to the fact, that walk and bicycle trips are usually better reported by telephone interviews *(Algers et al. 1994).*

In principle, surveys are never capable of providing an exact replication of reality. Thus, at least four different kinds of errors can be identified: sampling error, coverage error, measurement error and, non-response error *(Kalfs et al. 2000*).

Sampling error is not a property of the data itself, but rather a consequence of the lack of data. Any sampling error does not affect the expected values of the means of the estimated parameters; it only affects the variability around them, thus determining the degree of confidence that may be associated with the means. It is basically a function of sample size and of the inherent variability of the parameter under investigation. This uncertainty can be reduced by taking larger samples or by using sample designs, such as stratified samples that yield less variance in the statistics of interest.

Another important source of error is the one resulting from **coverage,** for instance, the failure to locate or visit some units in the sample. Well-known examples of problems with coverage are found in telephone surveys. Telephone coverage is partly a problem because the coverage is not uniformly distributed among the population. Access to telephones is related to specific socio-economic background variables, such as age and income.

Measurement errors occur when the observed or reported value is different from the true value. Measurement error is generally described by the term “observational errors”, whether they arise from the interviewer, the respondent, the questionnaire or the data collection mode. Three types of consequences of measurement error can be identified: non-reporting of trips, incomplete information and incorrect information. Certain data items, such as income levels or trip departure and arrival times that require estimates on the respondent’s part can be a source of error. Such errors can produce non-random biases when respondents consciously or unconsciously distort their answers.

Non-response is a phenomenon in which units that belong to the selected sample provide no information at all. Partial non-response is present when only part of the data is collected. Non-response has two effects on the results of sample surveys. First, it reduces the sample size and hence increases sampling error. A more negative consequence is that non- response may bias results and lead to under- or overestimation of variances. This occurs when non-respondents systematically differ from respondents. A lot of research has focused on the characteristics of non-respondents. It seems to be common knowledge that non- and less mobile households in particular, and people with relatively simple activity patterns and those with complex activity patterns as well, tend to respond less.

The most problematic sources of errors are those, whose influence cannot be corrected by socio-demographic or other kinds of weighting. Although mode shares, among other factors, could at least be partially corrected by applying socio-demographic weighting (see Section 5.3.2.1), it must be noted that the mode choice models are usually estimated using unweighted data. In model transfer, it is also very important to take seasonal variations into account. If different pieces of research are compared to each other, the minimum requirement is that the mobility surveys have been carried out in the same season.

It should also be noted, that all the trips made during the report period are used as the basis of the Finnish travel demand models. Consequently, there will be correlation between the answers provided by the same individual, and therefore in the error terms in the utility function. This violates the assumption that the error terms are independent of each other

and means that we can no longer rely on the estimated variance estimates *(Cirillo et.al. 2000)*. The problem is known as the **repeated measurement issue** and it is considered more specifically in Section 5.3.3.3.

2.3 Literature review

2.3.1 The transferability of model structure and model parameters

There is no concensus on the model transferability in the literature. However, results have mainly shown that model transfer is possible when the estimation context models are well defined and the data quality is good *(Karasmaa and Pursula 1995)*. According to *Algers, Colliander and Widlert (1987)* the biggest problems are caused by the low quality of the original data. If the models estimated in the estimation context situation are not good, then the transferred models cannot be good.

According to *Ben-Akiva (1981)* specification errors of the explanatory variables are major contributors to failures of transferability. Similarly, *Louviere (1981)* argued that transferability is frequently linked to validation of travel-choice models. Because each study area has a different covariance structure, each represents only one of the possible parameter estimates for regions; and the nature of random observations results in inefficient parameter estimates.

Koppelman and Pas (1986) have compared the transferability of joint choice and sequential (or nested) choice models. The hypothesis that joint models would be more transferable than sequential models was based on the work of *Ben-Akiva (1974)* in which he argues for the adaptation of joint choice models in preference to sequential models. The contradictory hypothesis that sequential models would be more transferable was based on the increased flexibility of sequential models, which allow them to take account of similarity among some of the joint choice alternatives *(McFadden 1978)*. In their study, *Koppelman and Pas (1986)* found only a small difference between the transferability of joint and sequential models. However, the authors suggested that the conclusion appeared to be dependent on the similarity of the estimation results for the joint and sequential models in this specific case.

Chester (1983) looked at the potential of market segmentation in improving the transferability of discrete choice models. If the population can be segmented into subpopulations in which travel tastes are reasonably homogeneous and transfer is made between groups of similar taste, then there will be no transfer error resulting from the difference in the distribution of tastes between the populations. The transferability of models between segmented and unsegmented data was tested empirically with data from Minneapolis-St.Paul, Baltimore and Washington D.C. The analysis involved transfers within each area and between the areas. The segmentation according to the gender, income level, workers per household and cars per household were tested. The main finding was, that transferability was consistently improved by using segment models instead of a single pooled model. However, the degree of improvement observed was small in absolute terms. According to the author, one reason for this could be that there was relatively little difference in the distribution of the travel tastes between transfer contexts in this study.

Model complexity can be assumed to have different effects on model quality and transferability. *Sermons (2000)* found, in his study of residential location choice models, that improvements in the specification aimed at better representing household-level systematic taste variation did not necessarily result in improvements in model transferability because the more complex models are more difficult to transfer.

Dehghani and Talvitie (1983) tested the effect of model specification (simple versus complex) on the transferability of mode choice models. They also found that the potential of transferability of a model did not increase at the same rate as did the complexity of the model specification.

According to *Talvitie and Kirchner (1978)* model specifications strongly affect the coefficients and the models' degree of fit. For this reason, the walking and waiting times, among others, should always be treated separately. On the other hand, the Finnish experiences have shown that travel time components cannot be produced reliably by the use of the Emme/2 assignment program, which has been the basic network related analysis tool in most Finnish research *(Karasmaa 1995)*.

Some research, such as those by *Parody (1977)* and *Train (1978),* indicate that the prediction efficiency can be improved by adding socio-economic and demographic variables to the models.

Tardiff (1979) has shown that the omitted variables can strongly affect the values of alternative-specific constants and increase their variability. When comparing two similarly specified models to each other, the difference between the average values of random components is usually quite large, the difference between the estimated variances of coefficients is smaller, and the difference between the parameter values is the smallest. This suggests the importance of the re-estimation of alternative-specific constants in model transfer.

Next, a short summary of the previous studies relating to temporal and spatial transferability is presented. The results of these studies have been more widely discussed in a study by *Karasmaa and Pursula (1995)*. The transfer methods presented in this section are described in more detail in Section 3.7.

2.3.2 Spatial transferability

The earliest of the transferability studies was carried out by *Watson and Westin (1975),* who studied the transferability of mode-choice models among different subareas within a single urban area. Their data were for the Edinburgh-Glasgow area of Scotland. The data were grouped into six categories according to whether the trip origins and destinations were in the city centre, the suburbs, or the area peripheral to the urban area. Each of the six models was then used to predict the mode choices of the other five samples. Watson and Westin concluded that the predictive ability of the model for the city centre was fairly favourable to transferability, but that the results for the other groups indicated a need to refine the models for locational differences.

Atherton and Ben-Akiva (1976) investigated the transferability of a work trip mode choice model estimated on 1968 Washington D.C. data. The transferred model was applied to data sets representative of New Bedford, Massachusetts from 1963 and Los Angeles, California from 1967. In addition, the authors developed a hierarchy of model transfer methods for empirical testing. Despite the Washington D.C. data set representing significantly different socio-economic and demographic distributions from what existed in Los Angeles and New Bedford, the authors, on the basis of transfer evaluation measures, came to the conclusion that the model was transferable. Of the several approaches for transferring that were developed, the Bayesian approach based on combining the existing model coefficients with the estimation results from a new sample gave the best overall performance.

Koppelman et al. (1985) investigated the intraurban and interurban transferability of work trip mode choice models. The intraurban transferability was studied across three geographic areas in Washington D.C. The interregional transferability analysis was undertaken in Minneapolis St. Paul, Baltimore, and Washington, D.C. According to the results, transfer effectiveness improved with the updating of alternative-specific constants and improved further with the updating of the parameter scale for both intraurban and interurban transfers.

The research done in Norway, in 1989, further supports the view that alternative-specific constants should be re-estimated *(Tretvik 1989)*. However, the research shows that in predicting the effects of changes in the travel system, the correction of alternative-specific constants already gave better performance than might have been the case had the transfer scaling for some variables also been used. *Algers, Colliander and Widlert (1987)* obtained similar results in their studies relating to the transferability of mode and destination choice models in Sweden.

Galbraith and Hensher (1982) investigated work trip mode choice model transferability between two suburban areas of Sydney, Australia. Both transfer scaling and the Bayesian approach were examined. Transferability was rejected using model parameter equality tests as well as the comparison of "proper fit" measures.

The Bayesian approach was also studied by *Abdelwahab (1991).* In this study the comparison of mode choice models estimated on the data sets from two regions in Canada yielded inconclusive results in relation to model transferability. In general, transferred models were found to be 18-23 percent less accurate than local models in predicting modal shares.

McCoomb (1986) studied the transferability of mode choice models for the journey to work in Canada's ten largest cities. Data for all cities were collected on a specific day, common to all, by Statistics Canada. Model specifications for the ten cities were identical. McCoomb's analysis did not result in a single transferable model to all the cities. However, some models were found to be "reasonably" comparable in terms of model coefficients, which led the author to conclude that a model from one city can be transferred and used to forecast modal split in another city when cities are reasonably similar in size, structure, transportation system, and so forth.

Sermons (2000) also studied the transferability of residential location models estimated for the San Francisco metropolitan area and the Portland metropolitan area and assessed the effectiveness of variance scaling, weighting, and systematic taste variation on model transferability. The results showed that scaling to account for variance and using the weighted likelihood function vastly improved the transferability of the model. However, the more complex models including separate time parameters for male and female workers, did not yield good results in comparison with the more simple models.

Most studies have focused on the transferability of the mode choice models; however, transferring the entire model system was examined in the Netherlands where the model system made for the Rotterdam and The Hague region was transferred to Utrecht *(Gunn and Pol 1986)*. The model system differed from the "standard four-stage transportation planning models", being more similar to the activity-based models than the traditional trip-based models. The model transferability was evaluated in terms of relevance, adequacy and validity. This research showed that the model transfer is also possible for the entire model system, when the alternative-specific constants are re-estimated and transfer-scaling factors are used. On the other hand, transfer scaling separately for travel standards and socio-economic variables did not give remarkably better transfer effectiveness.

In Sweden, the model combination estimated in Göteborg, Jönköping and the Netherlands (known as the Zuidvleugel-study 1977-81) was transferred to Helsingborg in 1989. The studied trip groups were work, business, shopping, social and recreational trips. Transfer effectiveness was measured in terms of transfer scaling factors, which deviated significantly from zero and one. The transfer effectiveness of the level of service variables seemed to be good. However, the naive transfer, without any transfer scale factors, or socio-economic variables, was not adequate. Some attraction variables for destination choice levels were required. However, there were quite large-scale factors, due to the population density differences between these countries and also different opinions, regarding which kind of society is better. *(Widlert 1990, Algers et al. 1994.*)

In Norway, the transferring of work and shopping trip models was tried in the Oslo-Akerhus region in 1990. The re-estimation of alternative-specific constants of work trip models indicated the largest improvements with respect to the naive model. In the transfer of the shopping trips, the best results were obtained by using an inverse model structure for mode- and destination choice models, and by using scale-factors for the level of service variables. According to the study, transfer scaling can lower the models' ability to predict changes in the travel system if the relationships of the coefficients do not stay the same between the estimation and application context *(Algers et al. 1994).*

In 1992, in Stockholm, the work trip model system was partially transferred to Trondheim. Although the travel behaviour between these two regions was quite different, the model transfer, which was carried out using the transfer scaling method succeeded quite well *(Algers et al. 1994).*

2.3.3 Temporal transferability

Most transferability studies focus on spatial dimension; the transfer scaling approach is the most common approach when corrections are made. In principle, the problem is the same when considering temporal transferability. However, in some cases, the viewpoint can differ from that used in the context of spatial transfer (e.g. the use of combined RP and SP data).

One of the first studies of temporal transferability was carried out by *Talvitie and Kirshner (1978),* who investigated the transferability of model coefficients in a temporal context for San Francisco, using pre-BART (Bay Area Rapid Transit) data (1973) and post-BART data (1975) and in a spatial context on Washington D.C. data 1968 and Minneapolis St. Paul data sets 1970. Transferability was evaluated by a statistical test of equality of the entire set of estimated model parameters including the modal alternative-specific constants for the local and application context. The authors found that the model coefficients were highly sensitive to model specification, thus having a significant impact on explanatory (and hence transferability) power. Moreover, based on statistical criteria, they found “little ground to claim that the coefficients of the work mode-choice models are transferable.”

Talvitie’s and Kirshner’s findings were supported by *Train (1978)* who investigated model specification and its relationship to temporal transfer effectiveness using the San Francisco pre-BART and post-BART data sets, which Talvitie and Kirshner also used in their study. Transferability was evaluated by tests of parameter equality in the two contexts and by forecasting accuracy.

Train found the coefficients of the level-of service attributes to be comparable in the two contexts but different for those of the socio-economic attributes. Large forecast errors were also obtained with the transferred model. In predicting in the application context, Train attempted to address the problem of defining the coefficient’s estimates for the attributes of the introduced new BART alternatives. The poor performance of his model is partly a reflection of the conjecture that was part of this process.

Finally, *McCarthy (1982)* investigated short term temporal stability using the previously mentioned BART data sets in addition to a third data set defined as early-BART. McCarthy could not reject temporal stability of the entire set of estimated model coefficients in the short run.

The studies by *Badoe (1994), Badoe and Miller (1995a), Badoe abd Miller (1995b)* compare different transfer methods. In their study, Badoe and Miller examined the long-term temporal transfer of work trip logit mode choice models, estimated using 1964 data for the Greater Toronto Area (GTA), to represent 1986 work trip mode choice in the GTA. Three updating procedures, which were previously presented in the literature (Bayesian updating, combined transfer estimation and transfer scaling) were examined plus a fourth new procedure, joint context estimation. The results indicated that the combined transfer estimation yielded, in almost all cases, the best predictive performance in the 1986 application context, based on the disaggregate full-sample log-likelihood measure used. However, as the authors stated, this was largely the result of the dominance of the transfer-scaling component of the procedure. This effectively resulted in the procedure being equivalent to a simple re-estimation of the model using the application context data set. Due to the impreciseness of the coefficients, the use of the smaller sample sizes reduces the predictive performance.

The joint context estimation yielded results which were generally comparable to the combined transfer procedure, but with a significantly more parsimonious parameter structure. Thus, the authors recommended that the estimation context data set should be available to support joint context estimation. On the whole, the differences between the methods were found to be rather small, possibly due to the use of statistical tests instead of sensitivity analysis, such as the calculation of elasticity.

One remarkable finding was that improving the model specification yielded far greater improvements in model performance than either "optimising" the transfer procedure or increasing the application context sample size. Thus, the authors concluded that model specification is extremely important in the transfer process.

Furthermore, *Badoe and Miller (1998)* investigated the additional complexity brought about by introducing scale parameters for the random utilities of the various modes in the application context in addition to the period-specific alternative-specific constants. This appeared to be justified by the statistically significant improvement in fit to the data at a disaggregate level compared to similarly specified models, but without the alternative-specific scales.

Elmi et al. (1999) reported from a study in which the temporal transferability of entropy-type trip distribution models was examined. Data for the study was drawn from three travel surveys conducted in 1964, 1986 and 1996 within the Toronto area. The study results showed that the travel-time parameter was not temporally stable. However, the transferred models were found to provide forecasts that were very comparable to those generated by locally estimated models. Stratification of the data by occupation-category of worker resulted in models with the best fit to estimation data as well as to forecasts.

Different ways of using RP and SP data to update the RP models were examined in the Netherlands, in 1989 *(Hague Consulting Group 1990)*. The results showed that in the case of RP models updated by RP data the updated mode-choice models for the commuter and business trips gave a good level of transfer effectiveness. The transferability of the models estimated for other trips was weaker, due to the different trip specifications in 1982 and 1989. In the case of RP models, which were updated using both the RP and SP data, the use of SP data did not improve the transferability of commuter and business trips models but it did improve the transferability of the models estimated for other trips. The use of the adjusted network RP-data, that is the network impedances supplemented by the additional information (public transport ticket type, parking type) from the SP study improved models. On the other hand, defining the travel time and cost coefficients by using the SP model and using these results to estimate RP models, did not improve results.

Finnish reseach relating to the model updating of the Helsinki metropolitan area mode choice models *(Kurri et al. 2001)* also showed that the combined use of revealed and stated preference information turned out to be quite difficult. This was because it is nearly impossible to define the variables of a stated preference exercise in the same way as can be done with the variables used in mode choice models based on actual choices.

Walker et al. (1998) examined the parameter scaling method to update the Delaware Valley Planning Commission's existing travel simulation models. The parameter scaling method was applied to update the mode choice models. For updating purposes two different surveys were conducted: A regional cordon line traffic survey to collect information on traffic volumes and patterns generated by vehicles entering or exiting the region and a small-sample home interview survey stratified by automobile availability and county was performed in 1988. The estimation of the parameter scaling factors (from current survey and secondary source data) took the form of an iterative parameter adjustment procedure. Trial adjustments were made, the travel simulation model or models were rerun, and the errors were calculated on the basis of survey or secondary source data, or both. Revised parameter estimates were then prepared using a proportional scaling technique, and the process was run through another iteration. The process ended when the current parameter set achieved an acceptable level of accuracy. The results showed, that when the required changes are small, the use of parameter scaling method yields good results.

2.4 Conclusions from the previous transferability studies

On the whole, the results relating to model transfer have been controversial. Many studies have shown that model transfer is possible when the estimation context models are well defined and the data quality is good. On the other hand, especially when studying spatial transferability, the opposite results have often been achieved.

Generally, by comparing the models estimated for urban and sub-urban areas it has been found that urban models apply well to other urban areas, but transferring models between urban and suburban areas has not succeeded as well. Additionally, most research has shown that under similar circumstances people in different cities and countries react in quite similar ways to changes in travel time and costs. However, variation in the distribution of the socio-demographic attributes varies greatly, and these have been taken into account when transferring models.

The literature shows that the segmentation, or added variables, may improve the model quality up to a certain point. Nevertheless, improvements in the specifications do not necessarily result in improvements in model transferability because more complex models are more difficult to transfer. Neither is there any broad evidence as to how the model structure, or estimation method affects model transferability.

The earliest studies have been mainly concerned with model comparison in two contexts. Also in many later studies, the basic assumption has been that model transfer is only possible if the true coefficients in the estimation and application contexts are quite similar. The possibilities for control transfer bias (the difference in the true parameters between the two contexts) have rarely been explored. When they have been looked at, the importance of the impreciseness of the coefficients has been ignored.

In literature, four different transfer methods are presented. From these, the Bayesian approach and transfer scaling are used in most studies in the 70's, and 80's. The Bayesian approach emphasizes the coefficients with respect to the inverse of the variances of each coefficient. The transfer scaling approach uses the new data purely to correct the transfer bias (the difference between the coefficients in the estimation and application context). In 1987, Ben-Akiva and Bolduc presented a method in which both the transfer bias and the model parameter variance were taken into account. In the 1990s, Badoe and Miller compared the three transfer methods presented earlier and applied an additional new approach, namely the joint context estimation as well. In their study, the combined transfer estimation yielded, in nearly all cases, the best predictive performance. However, the joint context estimation yielded results which were generally comparable to the combined transfer procedure.

Model transferability has traditionally been evaluated on the basis of how well transferred models replicate existing behaviour rather than on their ability to forecast adequately changes in travel demand. Such analyses where different sets of data are used in the model calibration and validation are severely limited. If the primary function of a mode-choice model is to predict the impact of changes in the transportation system on travel behaviour, then an essential characteristic of such a model has to be its ability to predict accurately.

Examining only one element (mainly the transfer of trip generation, or mode choice level) instead of the whole four-step model system can also be regarded as a weakness of the many previous studies.

2.5 Implications for this study

Earlier transferability studies have mainly focused on the study of model transferability by using only one method and one sample size. The assumption has been that model transfer is only possible if the coefficients in the estimation and application contexts are quite similar. As a large part of the difference in coefficients and predictions is due to the discrepancies in the formulation of the initial data, or random variation, and only partly due to the real differences in the estimation and application context, the main emphasis in this study is to investigate the relationship between the transfer bias and the impreciseness caused by the sample size. Different transfer methods are compared to each other, as well as the amounts of data needed to estimate mode and destination choice models.

Thus, the intention of this thesis is to compare different transfer methods as a function of sample size, and identify the factors affecting the models' quality and the impreciseness of model parameters.

More specifically, the aims of the research are:

- to investigate the importance of transfer bias and the preciseness of model coefficients to the applicability of different transfer methods, and to further develope the joint context estimation method to control the effects of transfer bias and coefficient impreciseness,

- to investigate the application context sample size requirements when transferring models or estimating new models as well, and to analyse how different quality criteria affect this sample size requirement,

- to compare different test measures used for studying model transferability and in particular to evaluate the applicability and goodness of the traditional statistical tests, with respect to those based on the prediction accuracy of forecasts.

The study is divided into two main parts. The first part of the thesis deals with the factors which decrease model quality and cause the apparent differences between the estimation and application context. For example, the importance of data gathering methods (telephone interview versus postal questionnaire), and sample size are studied. The second part analyses the spatial transferability of the HMA models.

3 OVERVIEW OF THE THEORY OF TRAVEL DEMAND MODELS AND MODEL TRANSFER

3.1 Random utility theory

The concept of random utility was first introduced by *Thurstone (1927),* who was a researcher in the field of mathematical psychology. Experiments showed that individual choices were, in many respects "irrational" and unpredictable. At the same time there was also considerable practical interest in being able to predict behaviour, and there was also empirical evidence that certain aspects of behaviour were in fact "predictable" on a more aggregate level *(Brundell-Freij 1995).*

Generally, discrete choice models, like the logit model, postulate that the probability of individuals choosing a given option is a function of their socio-economic characteristics and the relative attractiveness of the option. The basic assumption of choice models is that each individual is attempting to maximize his or her utility. In other words, when a traveller has to make a decision in selecting an alternative from the available choice set he/she chooses the one which is the most attractive (or provides the maximum utility) for him/her *(Ben-Akiva and Lerman 1985).*

The attraction of available alternatives *i* for each individual *n* can be measured with utility function. In a real situation there are attributes that can be observed by the modeller and attributes that remain unobserved. In Equation 1 the first term of the total utility U_{in} stands for the observed utility which is derived from the observed attributes. This portion of the utility is usually marked with the letters V_{in} and it is called deterministic utility. Additionally, the random portion for person *n* and alternative *i* is presented in term ε_{in}.

$$U_{in} = V_{in} + \varepsilon_{in}. \tag{1}$$

The deterministic component of utility function is usually expressed as a linear function of the attributes $x_1,\ldots,x_k$.

$$V = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k, \tag{2}$$

where $\beta_1,...,\beta_k$ are the coefficients to be estimated. Variables can be included as either generic or alternative specific *(Stopher and Meyburg 1979, Ben-Akiva and Lerman 1985)*. Introducing a variable in a generic form means that when considering the different choices travellers have equal perception toward this variable across all alternatives. Whereas, an alternative specific variable suggests that the perception of users toward the same variable varies from one alternative to another.

In logit models it is assumed that the unobserved part of the utility (labelled ε_{in}) is distributed independently and identically following Gumbel distribution with a zero mean *(Ben-Akiva and Lerman 1985).* The value of random terms can, by definition, vary across different decision-makers and groups of persons as well as alternatives. There are a number of reasons why there has to be a random component in a utility function *(Lerman 1984):*

(a) **Measurement errors**. There are often errors in measuring the attributes of the alternatives.

(b) **Proxy variables.** The modeller is often forced to use proxy variables if real variables are difficult to measure. The differences between the proxies and the actual attributes are source of randomness.

(c) **Omitted attributes.** The decision-maker (traveller/individual) often uses attributes that the modeller does not know, cannot measure or chooses to omit for some reason.

(d) **Unobserved taste variations.** There are undoubtedly taste variations between individuals. Although the modeller can use socio-economic attributes to take this into account, there is probably always some residual variation that remains unexplained.

3.2 Multinomial Logit Model (MNL)

The multinomial logit model is the most widely used discrete choice model. It is easy to use and an estimation of its parameters is inexpensive. The logit model gives a probability of a specific alternative to be chosen by an individual. The selected mode has to be available in the choice set and the selection is done using the differencies in the utilities of alternatives. Under the assumptions made in Section 3.1, the choice probability for alternative *i* and individual *n* is given by *(Ben-Akiva and Lerman 1985):*

$$P_n(i) = \frac{e^{\mu V_{in}}}{\sum_{i=1}^{J} e^{\mu V_{jn}}} = \frac{1}{1 + e^{\mu(V_{2n} - V_{in})} + e^{\mu(V_{3n} - V_{in})} + \ldots + e^{\mu(V_{jn} - V_{in})}}. \tag{3}$$

In Equations 3 and 4 the value of the scale parameter μ depends on the variance of the random term in such a way, that the smaller the variance σ, the greater the value of μ, and the more sensitive the model to the changes of variable values. The scale parameter μ is not identifiable (if V_{in} is linear in response to its parameters, the scale parameter μ can not be separately estimated). Thus in model estimation the scale parameter μ has to be fixed. Usually the scale parameter is assumed to be 1. It implies that the variance of the disturbances is fixed and homoscedasticity (i.e the error term has a constant variance) is assumed among choices *(Ben-Akiva and Lerman 1985).*

$$\mu^2 = \frac{\pi^2}{6\mathrm{Var}(\varepsilon)}. \tag{4}$$

Equation (3) shows that the choice probabilities depend only on the absolute difference in utilities between alternatives, not the absolute values of utilities.

The principal form of the relationship between P_i and V_i, given by the multinomial logit model, is S-shaped, in the same way as binomial model which is illustrated by Figure 1.

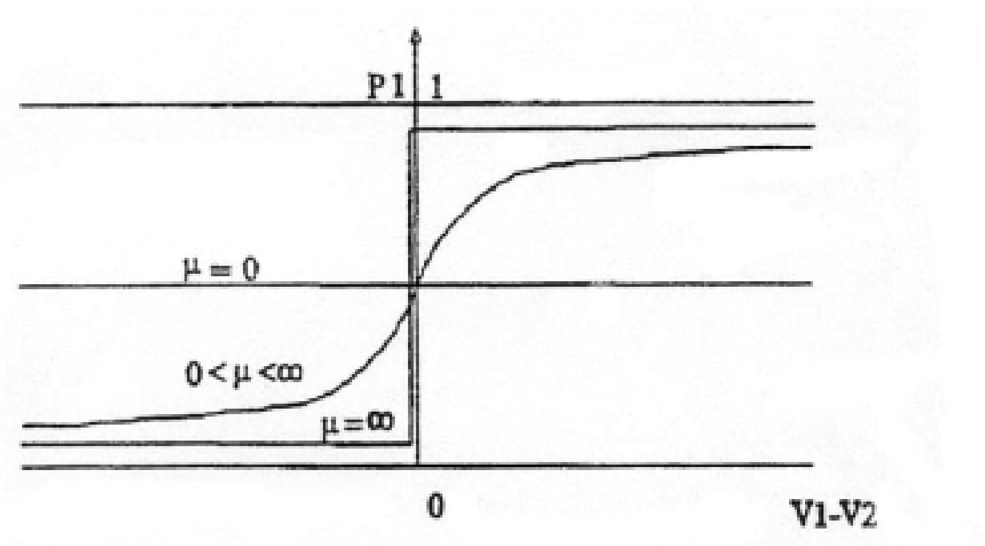


Figure 1: The principal form of the binomial logit model relationship with different values of μ.

The change in the probability is largest and the curve is steepest when the choice probability is near 0.5, and becomes smaller as it approaches zero or one. If the same alternative is chosen by almost all persons in the same situation, the variance of random factors approaches zero and the value of scale parameter μ approaches infinity.

When modelling several dimensions of choice, e.g. destination and mode choice, simultaneously, the multinomial logit model is expressed as *(Ben-Akiva and Lerman 1985)*:

$$P_{md} = \frac{e^{V_{md}}}{\sum_{m'd' \epsilon C} e^{V_{m'd'}}}. \tag{5}$$

where subscriptions m and d denote mode and destination choice, respectively.The formulation is called to joint logit model and it is illustrated in Figure 2.

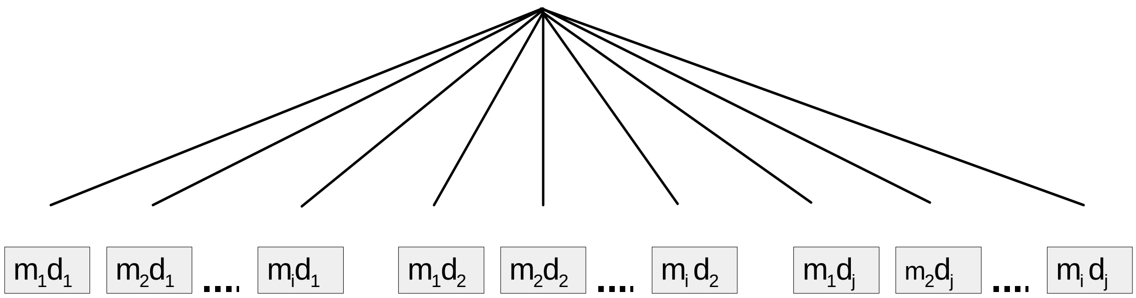


Figure 2: A joint logit model for mode (m) and destination choice (d).

The multinomial logit model has the property that the relative choice probability of any two of the alternatives is independent of the presence of a third alternative. This feature is commonly denoted as IIA or **indepence from irrelevant alternatives** and it can be easily shown to hold in the case of MNL as follows *(Luce and Suppes 1965).* If a ratio of the probabilities for two alternatives, *i* and *k* for the individual *n*, is considered, it can be seen that:

$$\frac{P_n(i)}{P_n(k)} = \frac{e^{V_{in}} / \sum_{j \in C_n} e^{V_{jn}}}{e^{V_{kn}} / \sum_{j \in C_n} e^{V_{jn}}} = \frac{e^{V_{in}}}{e^{V_{kn}}} = e^{V_{in} - V_{kn}}. \tag{6}$$

This ratio is indeed only dependent on the difference between the utilities for alternatives *i* and *k*. This allows a new alternative to be easily added into the model. The IIA property was originally regarded as a great advantage of the logit model but there has also been a lot of concern about the problems caused by this property.

However, one of the IIA's advantages is that if there are too many alternatives, such as in the case of destination choice, it can be shown *(McFadden 1978)* that the model parameters can be estimated consistently on a subset of available alternatives to the decision-maker. Thus, if there are 100 alternatives, the modeller can estimate the logit model for each sampled person on a subset of 10 alternatives which include the actual choice and 9 other randomly selected alternatives. Due to the IIA property, the relations within the subset are unaffected by the exclusion of the alternatives that are not in the subset *(Ben-Akiva and Lerman 1985, Ortúzar and Willumsen 1994).*

3.3 Nested Logit Model

The nested logit model, first derived by *Ben-Akiva* (*1973, 1974*), is an extension of the multinomial logit model designed to capture correlations among alternatives. In applications where utilities of some alternatives are correlated, or have different variances attached, the logit model may substantially overpredict or underpredict the shifts in the share of each alternative. Some of these problems can be avoided by using nested logit model. The nested logit model can be used when the choice set can be partitioned into subsets according to their properties. Alternatives that are correlated and share the same source of variance of error are placed into the same nest, which partly removes the IIA property. In the tree structure the IIA property holds within a subset but not across subsets. Hence, the ratio of probabilities of any two alternatives within a subset is independent of the existence of other alternatives. However, the ratio of two alternatives in different subsets is dependent of existence of the other alternatives *(Train 1986).*

The nested logit model is illustrated in Figure 3. The introduction of lower nests in their immediate superiors is done by means of the utilities of the composite alternatives which, in general, have two components: one which consists of the expected maximum utility (EMU) of the lower nest options, and another which considers the vector of attributes which are common to all members of the nest *(Ortúzar and Willumsen 1994).*

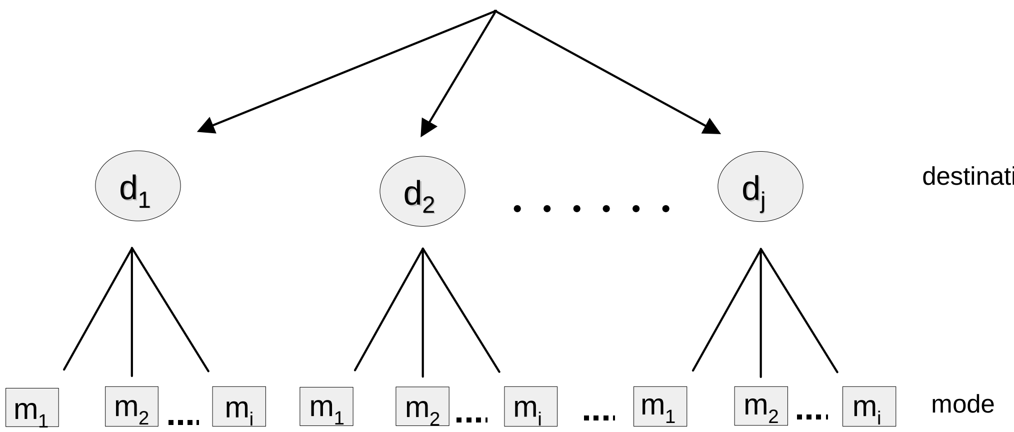


Figure 3: The nested logit model for mode (m) and destination (d) choice.

Let us assume that the mode choice probability, when the destination d has been chosen, is $P_{m|d}$ and the destination choice probability is P_d . Thus, the nested logit model can be stated as

$$P_{md} = P_{m|d} * P_d \ ,$$

$$P_{m|d} = \frac{e^{V_{m|d}}}{\sum_{m' \in M_d} e^{V_{m'|d}}}, \quad P_d = \frac{e^{V_d + W \ln \sum_{m' \in M_d} e^{(V_{m'|d})}}}{\sum_{d' \in D} e^{V_{d'} + W \ln \sum_{m' \in M_{d'}} e^{(V_{m'|d'})}}}, \tag{7}$$

where M_d is the set of all possible modes to destination d and D the set of all possible destinations.

In Equation 7 the log of the denominator of $P_{m|d}$ has been used. This term, labelled logsum or expected maximum utility, EMU (*Ortúzar and Willumsen 1994*), is the connection between the mode and destination choice (Equation 8). The parameter W estimated for logsum variable describes the similarity of the alternatives and it is estimated as an additional parameter of the model.

$$logsum = \ln \sum_{m' \in M_d} e^{(V_{m'|d})}. \tag{8}$$

The value of W depends on scale parameters μ_d and μ_m and variance of the random terms (assuming $\varepsilon_m \simeq 0$):

$$W = \sqrt{\frac{\mathrm{Var}(\epsilon_{dm})}{\mathrm{Var}(\epsilon_d) + \mathrm{Var}(\epsilon_{dm})}} = \frac{\mu_d}{\mu_m}. \tag{9}$$

If W=1, the choice probabilities become multinomial logit. If $0 \leq W \leq 1$, the IIA property holds within a subset but not across subsets. The situation, in which W>1, is not consistent with theory of utility maximization. This means that an increase in the utility of an alternative in the nest would tend to increase not only its selection probability but also that of the rest of options in the nest.

The variance of the random utilities is thus the smallest at the lowest level of the tree, and it can not decrease as we move to a higher level. It also means that choices whose variance is greater are usually placed to the upper level of structure. The model structure also describes how well the choices can be explained. The choices of model structure used do not determine in which order the decisions are made in real word, only consider similarities between alternatives. In some cases, if W>1, the models for mode and destination choices are to be estimated in inverse order, so that the destination choice is placed in the upper level of structure to have $W \leq 1$.

The nested logit model may be estimated sequentially or simultaneously. The benefit of simultaneous estimation is that the model can use all information at the same time. Some information can be lost in sequential estimation, when the lower level variables are studied only in relation to the chosen alternatives without the simultaneous connection to the unchosen altenatives on the upper level *(Ortúzar and Willumsen 1994).* Moreover *Amemiya (1978)* has shown that by using standard multinomial logit estimation programs and sequential estimation procedure, the variance-covariance matrices of the estimates obtained for the marginal probabilities of higher level models are incorrect and too small. On the other hand, simultaneous estimation may have problems if W will be over 1 (see Section 4.5).

3.4 Elasticities of logit model

One useful property of econometric demand models is the concept of an elasticity. An elasticity is the relative change in one variable that is associated with a relative change in another variable. The simplest case is the elasticity of the probability of an individual n choosing alternative i with respect to a change in some attribute k that is an independent variable in the model, namely one of the x_{ink}'s. In this case the direct elasticity of logit is given by *(Ben-Akiva and Lerman 1985):*

$$E_{x_{ink}}^{P_n(i)} = \frac{\partial P_n(i)}{\partial x_{ink}} * \frac{x_{ink}}{P_n(i)} = \frac{\partial \ln P_n(i)}{\partial \ln x_{ink}} = [1 - P_n(i)] x_{ink} \beta_k. \tag{10}$$

Similarly the disaggregate cross elasticity of the probability that alternative i is selected with respect to an attribute of alternative j is

$$E_{x_{jnk}}^{P_n(i)} = \frac{\partial \ln P_n(i)}{\partial \ln x_{jnk}} = -P_n(j) x_{jnk} \beta_k, \; for \;\; j \neq i. \tag{11}$$

The change in the probability of choosing alternative i given a change in an observed factor, x_{ink}, entering the representative utility of alternative i is stated as *(Train 1986)*:

$$\frac{\partial P_n(i)}{\partial x_{ink}} = \frac{\partial V_n(i)}{\partial x_{ink}} P_n(i)[1 - P_n(i)]. \tag{12}$$

Usually $V_n(i)$ is linear in the observed variables, with parameters as coefficients. If the coefficient of x_{ink} is the scalar β_x, then $\partial V_n(i)/\partial x_{ink}=\beta_x$. Note that, since β_k is constant, the derivate is largest when $P_n(i)=1-P_n(i)$, which occurs when $P_n(i)=1/2$, and becomes smaller as $P_n(i)$ approaches zero or one (Figure 4). Stated intuitively, the effect of a change in an observed variable is highest when the choice probabilities indicate a high degree of uncertainty regarding the choice; as the choice becomes more certain (i.e., the probabilities approach zero or one), the effect of a given change in an observed variable decreases (*Train 1986*).

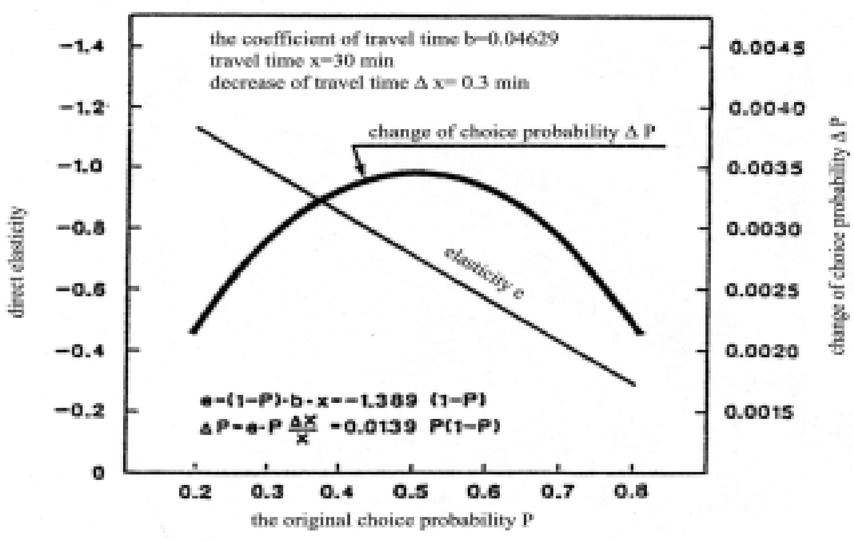


Figure 4: The effect of original choice probability to the direct elasticity of travel time and to the change of choice probability (YTV 1989).

A logically necessary aspect of derivates of choice probabilities is also that, when an observed variable changes, the changes in the choice probabilities sum to zero. Hence, to increase the probability of one alternative necessitates decreasing the probability of other alternatives.

Usually a researcher is interested in the average probability or average response within a population, rather than the response of any one individual. For example, suppose the researcher has a random or stratified random sample of individuals drawn from a population. Aggregate, or population, choice probabilities are calculated by taking the weighted average of the choice probabilities calculated for each individual. The average probability for alternative i is estimated as *(Train 1986):*

$$\overline{P_i} = \sum_n w_n P_n(i), \tag{13}$$

where w_n is sampling weight associated with individual n, and the summation is over all sampled individuals. If the sample is purely random, then w_n is the same for all sampled individuals and equals $1/N$, where N is the sample size. For stratified random samples, w_n varies over strata.

The number of individuals in the population predicted to choose alternative i is estimated as the average probability for alternative i times the population size (M):

$$N_i = M\overline{P_i}, \tag{14}$$

where M is the number of decision makers in the population and N_i is the estimated number that will choose alternative i. Average derivates and elasticities are calculated similarly as the weighted average of individual derivates and elasticities.

An alternative method of estimating average probabilities and responses is common but not consistent. Instead of calculating the probabilities and responses for a sample of decisionmakers and then taking averages, one possibility is to calculate probabilities and responses for an average decision maker and consider these to be in some way representative of average population behaviour. The inconsistency of this approach results from the fact that the choice probabilities, derivates, and elasticities are nonlinear functions of the observed data and, the average value of a nonlinear function over a range of data is not equal to the value of the function evaluated at the average of the data. This error is called aggregation error and it concerns all the situations in which the zonal values are used to estimate models or by making aggregate forecasts.

Let us consider Figure 5, which gives the probabilities of choosing a particular alternative for two individuals with representative utility for this alternative of a and b assuming the representative utility of other alternatives is the same for the two individuals. The average probability is the average of the probabilities for the two individuals, namely, $(P_a+P_b)/2$. The probability evaluated at the average representative utility is given by the point on the logit curve above $(a+b)/2$. As shown for this case, the average probability is above the probability at the average representative utility. In general, the probability evaluated at the average utility underestimates the average probability when the individuals' choice probabilities are low and overestimates it when they are high.

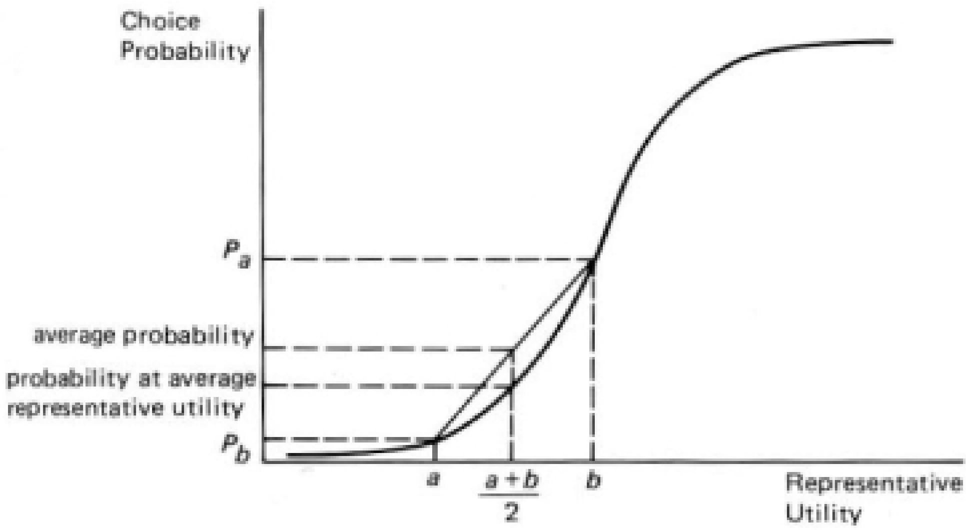


Figure 5: Error due to average individual aggregation (Train 1986).

3.5 Maximum likelihood estimation

The logit model is normally estimated using maximum likelihood method. Maximum likelihood estimation is based on an idea to estimate parameters, which will reproduce as high probability of observed choices as possible with the observed data. This is done by calculating the predicted probabilities of the observed choices *(Pn(i))* and maximising the product of those probabilities. Equation (15) shows the formula to be maximised *(McFadden 1974).*

$$L^* = \prod_{n=1}^{N} \prod_{i \in C_n} P_n(i)^{y_{in}}, \tag{15}$$

where y_{in} = 1, if alternative *i* from choice set C_n is chosen by individual *n*
= 0, otherwise
C_n = the set of alternatives for the individual *n* (universal choice set)
N = number of observations.

Normally *L** itself is not maximised, because the natural logarithm of *L**, labelled *L*, is more manageable. The modified formula is shown in Equation (16).

$$L = \sum_{n=1}^{N} \sum_{i \in C_n} y_{in} (\beta^T x_{in} - \ln \sum_{j \in C_n} e^{\beta^T x_{jn}}), \tag{16}$$

where $\beta^T x_{in} = \sum \beta_k x_{ink}$ is the deterministic component of the utility for the k'th independent variable for alternative i and individual n.

Setting the first derivates of L with respect to the coefficients equal to zero, the necessary first-order conditions can be obtained. The estimation of the parameters can be done iteratively using Newton-Raphson's method. The likelihood values are positive and the values of the log-likelihood are negative. The closer to zero the log-likelihood value is, the better is the model.

The maximum likelihood estimator $\hat{\beta}_n$ is not inevitable unbiased or efficient; however, under certain assumptions *(Vasama and Vartia 1979, Rao 1973)* the estimator is consistent, asymptotically normal and asymptotically efficient (no other consistent and asymptotically normally distributed estimator has a smaller asymptotic covariance matrix). The lack of bias and efficiency do not only depend on the sample size but on the distribution of the estimator. This distribution is often unknown. Nevertheless, it is often possible to calculate whether the estimator approaches the true value, when the sample size is infinite.

Due to the asymptotic efficiency, the variance-covariance matrix of the estimator will asymptotically approach the Cramér-Rao bound. The Cramér-Rao theorem can be written as *(Ben-Akiva and Lerman 1985, Rao 1973)*:

$$\mathrm{Var}(\hat{\beta}_N) \geq B^{-1}, \text{ where}$$

$$B = \begin{pmatrix} -E\left(\frac{\partial^2 L}{\partial \beta_1^2}\right) & -E\left(\frac{\partial^2 L}{\partial \beta_1 \partial \beta_2}\right) & \cdots & -E\left(\frac{\partial^2 L}{\partial \beta_1 \partial \beta_K}\right) \\ -E\left(\frac{\partial^2 L}{\partial \beta_2 \partial \beta_1}\right) & -E\left(\frac{\partial^2 L}{\partial \beta_2^2}\right) & \cdots & -E\left(\frac{\partial^2 L}{\partial \beta_2 \partial \beta_K}\right) \\ -E\left(\frac{\partial^2 L}{\partial \beta_K \partial \beta_1}\right) & -E\left(\frac{\partial^2 L}{\partial \beta_K \partial \beta_2}\right) & \cdots & -E\left(\frac{\partial^2 L}{\partial \beta_K^2}\right) \end{pmatrix} \tag{17}$$

It can be difficult; even impossible, to calculate the expected values of the matrix elements analytically. Matrix B can also include unknown parameters β. In many cases, the β must be replaced by the estimator $\hat{\beta}_N$.

3.6 The mixed logit or error components model

In a significant generalization of multinomial logit *McFadden (1978)* has derived the generalized extreme value (GEV) model. This is actually a large class of models which includes MNL and NL. All these models typically maintains homogeneity in responsiveness to attributes of alternatives across individuals. Recently, also some more advanced techniques have been developed, such as the mixed logit formulation. The mixed logit model allows for the variation in preferences between individuals. Let us assume the utility function of option A_i for an individual *n* in a choice situation *t* is given by *(Ortúzar and Willumsen 2001)*:

$$U_{int} = \boldsymbol{\theta}^{\boldsymbol{T}}{}_{\boldsymbol{n}} \mathbf{X}_{\boldsymbol{int}} + \varepsilon_{int}, \tag{18}$$

where $\mathbf{X}_{int}$ is a vector of observable variables, as usual, but now $\boldsymbol{\theta}_{\boldsymbol{n}}$ is a vector of unknown coefficients that vary randomly according to the individual tastes. Finally ε_{int} is a random error term which distributes IID Gumbel, independently of both $\boldsymbol{\theta}_n$ and $\mathbf{X}_{int}$. Note that this specification is not completely general because the parameters $\boldsymbol{\theta}_n$ do not depend on *t* (as in general one would expect that individual tastes do not change from situation to situation), but it may be generalised if desired *(Ortúzar and Willumsen 2001).*

This specification is identical to the MNL except for the fact that the coefficients $\boldsymbol{\theta}_n$ are not fixed but vary in the population. Now, the vector for coefficients $\boldsymbol{\theta}_n$ for each individual may be expressed as the sum of its population mean $\boldsymbol{\theta}^*$ and individual deviations $\boldsymbol{\eta}_n$ represent individual tastes in relation to the average tastes of the population:

$$U_{int} = \boldsymbol{\theta}^{*\boldsymbol{T}} \mathbf{X}_{\boldsymbol{int}} + \boldsymbol{\eta}^{\boldsymbol{T}}{}_{\boldsymbol{n}} \mathbf{X}_{\boldsymbol{int}} + \varepsilon_{int}. \tag{19}$$

The unobserved part of the utility ($\boldsymbol{\eta}^{\boldsymbol{T}}{}_n \mathbf{X}_{int} + \varepsilon_{int}$) is correlated over options and situations due to the influence of $\boldsymbol{\eta}_n$, and it is possible to achieve very general patterns of correlation, taste variations and heteroscedasticity with an appropriate specification of parameters and variables *(Ortúzar and Willumsen 2001).*

However, maximum likelihood estimation in a mixed logit model formulation is not trivial. Instead, and as in the case of the multinomial probit, the probability must be approximated by a simulation and where the simulated log-likelihood is maximised.

In most of the literature, Normal-distributed error components are used. This, however poses problems eg. with value of time (VOT) calculations because with so few observa-tions the ratio between two independently distributed standard normal variables is Cauchy distributed, not normal distributed, as will be stated in Section 3.8. One could assume that the cost-coefficient does not follow a distribution (only the time-coefficient) in which this problem is avoided. However, Danish experiences have shown that it is not possible to reject on statistical grounds the notion that no heterogeneity would exist in the cost coefficient *(Nielsen and Jovicic 2003)*.

A far better solution then is to assume that the coefficients are lognormal *(Ben-Akiva et al. 1993)*. In the thesis by *Sørensen (2003a)* this is actually shown to be the best fit for most empirical distributions of the coefficients. The positive aspect of a lognormal distribution is that it is nonnegative and multiplicative. This means that the ratio between two lognormal distributions is also lognormal, and that the VOT therefore also follows a lognormal distribution, i.e. if θ_{cost} is lognormal and θ_{time} is lognormal, then $\theta_{\text{time}}/\theta_{\text{cost}}$ is also lognormal. The parameters of the joint lognormal VOT distribution can easily been calculated, and the mean and variance follows directly. However, it should be noted that in model transfer the advantage gained when using lognormal distribution might be open to question, if it is used to correct the effect in a small sample size. That is, if the coefficients of time and cost are made negative (minus-lognormal assumption), it may cause bias in the other coefficients. The other important point to remember is that to avoid identification problems only the cost or time variable can usually be estimated using lognormal distribution. In fact *Sørensen (2003b)* has shown that the specification and estimation of a mixed logit is not a simple task. The number of alternative specifications for the shape of distributed terms grows rapidly with the number of coefficients that are expected to be distributed and with the number of considered shapes of distribution.

In this thesis an error component model is not used because it is a relatively new method (to date there are no software applications available). If error component models had been used, the user interface would have needed to have been reformulated as well. It is obvious that some results, in particular those relating to VOT, would have been different if a mixed logit model had been used. Nevertheless, some Monte Carlo simulations were performed to examine the effect where coefficients are assumed to be stochastic variables. It was found that the distribution of the VOT did not differ greatly from that based on the empirical data.

3.7 Transfer methods

3.7.1 General aspects

The model transfer approach is based on the idea that estimated model parameters from a previous study in a different context may provide useful information for estimating the parameters for the same model in a new context, even when the true values of the parameters are not expected to be equal. Depending on the quality and type of the travel survey (which has been carried out) the model transfer can be done in at least four different ways. These are:

- transfer scaling,
- Bayesian method,
- combined transfer estimation and,
- the joint context estimation.

Naive transfer is considered here as a special case of transfer scaling. The transfer methods differ from each other in emphasising the estimation and the application context data. Errors expressed in model transfer can be caused by sampling errors or errors that are caused by differences between the coefficients in the initial and the final stages.

In principle, the model transfer usually decrease the variance of model parameters resulting that the transferred models are more precise than the corresponding application context models that are estimated solely from small data samples. On the other hand, if the sample size of the application context data is small, the variance of the model parameters increases and this is reflected in transferred models, too. The problem in model transfer is that we do not know how much the difference between the coefficients estimated in estimation and application contexts is caused by the impreciseness of parameter estimates, and how much is caused by the difference in travel behaviour.

The transfer scaling approach uses the new data only to correct the transfer bias (the difference between the coefficients in estimation and the application context) and therefore does not explicitly consider the differences in sampling errors between the two data sets. In the Bayesian method the initial point is that the transfer bias is assumed to be zero; that is, the estimation and application contexts share the same underlying set of parameters. The combined Bayesian estimator extends the Bayesian procedure to take into account the presence of a transfer bias (*Ben-Akiva and Bolduc 1987*). When transferring with any of these three methods, only the application context data are required. When we are using joint context estimation, the new models are estimated using both the estimation and the application context data *(Badoe 1994, Badoe and Miller 1995b).* The basic assumption has normally been, that the two data sets are drawn from the same underlying probabilistic choice process. Based on this assumption the ratios of coefficients are expected to be the same and common coefficients are used and only the alternative-specific constants are estimated separately for these two contexts. However, in this study, we assume that particularly in spatial transfer there may be differences (transfer bias) between the estimation and application context parameters, and the best way to apply the joint context estimation is using both data-specific and common coefficients.

3.7.2 Transfer scaling

There are at least four different ways applying the transfer scaling approach:

Naive approach: The naive approach uses the existing model with its original coefficients in the new situation. This approach assumes that all the factors relevant to the choice process are embodied in the model, an assumption that can never be fully justified.

Adjustment of constant terms: If the validity of the model coefficients other than the constant terms is accepted, then aggregate data can be used to adjust the constant terms so that the model replicates existing aggregate data.

In this case the ***utility function*** of the application context model is:

$$V_i(x) = \gamma_i * c_i + \beta_1 * x_{l1} + \beta_2 * x_{l2} + ... + \beta_3 * x_{lk} + \beta_4 * x_{(l+1)k} + ... + \beta_k * x_{mk} \text{ , where}$$
c_i = the alternative specific constant in estimation context
$\beta_1 - \beta_k$ = the coefficients in estimation context
$x_{l1} ... x_{lk}$ = travel system variables in application context
$x_{(l+1)k} ... x_{mk}$ = other variables in application context (20)
γ_i = the scale factor to the alternative specific constant.

Re-estimating new constants and scale: In this case all or some of the parameters in the application context utility function are scaled and alternative-specific constants are estimated from an application context sample, assuming that the remaining utility function parameters are transferable from the estimation context. As is shown in *Ben-Akiva and Bolduc (1987)* and *Atherton and Ben-Akiva (1976*), much of the transfer bias can be eliminated by adjusting the model constants and scales. The transfer problem then becomes one of determining the application context alternative-specific constants and shifts in the transfer scaling factor relative to the estimation context. For example, given a set of estimation context parameters β_1, one can assume that the application context systematic utilities, $V_{in,2}$ take the form:

$$V_{in,2} = \boldsymbol{\gamma_{i,2}} \boldsymbol{\beta_1}^{\boldsymbol{T}} \boldsymbol{X_{in,2}} + \alpha_{i,2}, \tag{21}$$

where $\gamma_{i,2}$ is a vector of scale factors for each group of variables to be scaled, T is transpose, $X_{in,2}$ is a vector of explanatory variables for alternative i for individual n and $\alpha_{i,2}$ is the alternative-specific constant for alternative *i (Ben-Akiva and Bolduc 1987).*

The transfer scale factor can also be determined by using SP data. In this case e.g, the coefficients for travel time and travel costs are first estimated by using SP data. The estimates for the coefficients are substituted to the estimation context model, which is to be re-estimated by using the estimation context data and the fixed coefficients for travel time and costs. Then this new estimation context model is transferred using the application context RP data.

Re-estimating the estimation context model: This procedure uses a small sample of observations of individual choices and assumes that the sample is representative of behaviour in the study area. The small disaggregate sample can be used to re-estimate the coefficients from the original model specification. This is identical to transfer scaling in the case where all the coefficients are re-scaled. In this study models based on this approach are referred as “new sample models”.

The first three methods test the transferability of model parameters. The fourth method also takes into consideration the validity of the model specification. In fact, the first and the last methods represent the extremes in model transfer and can also be regarded as independent methods as well. The naive approach is the simplest approach to model transfer. The estimation of new models is not exactly the same as in the transfer method but rather the estimation corresponds to the situation in which all the coefficients are separately re-scaled.

3.7.3 Bayesian method

In the Bayesian method parameter estimates $\hat{\beta}_2$ from a small application context sample are combined with the estimation context parameter values $\hat{\beta}_1$ using a classical Bayesian analysis to yield an updated set of parameters. The basic assumption is that the behaviour of an individual is not different in estimation and application contexts. Thus, the new and the old data share the same underlying set of parameters.

Based on the normality assumption the variance $\sigma^2_{transferred}$ of the coefficient can be defined by using the variances of the original coefficients (σ_1^2 and σ_2^2) estimated in estimation and application context *(Atherton and Ben-Akiva 1976)*:

$$\sigma^2_{transferred} = [(1/\sigma_1^2) + (1/\sigma_2^2)]^{-1}. \qquad (22)$$

Thus in a one dimensional case the Bayesian estimator is stated as a weighted average of the direct estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.

$$\hat{\beta}_{transferred} = \frac{(\hat{\beta}_1/\sigma_1^2)+(\hat{\beta}_2/\sigma_2^2)}{(1/\sigma_1^2)+(1/\sigma_2^2)} = \frac{\sigma_2^2\hat{\beta}_1+\sigma_1^2\hat{\beta}_2}{\sigma_1^2+\sigma_2^2}. \tag{23}$$

In the multivariate case the coefficients are to be estimated according to Equation 24 - *(Badoe and Miller 1995 b):*

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{transferred}} = (\textstyle\sum_1^{-1} + \sum_2^{-1})^{-1}(\sum_1^{-1}\hat{\boldsymbol{\beta}}_1 + \sum_2^{-1}\hat{\boldsymbol{\beta}}_2)$$

and covariance matrix

$$\textstyle\sum_{\text{transferred}} = (\sum_1^{-1} + \sum_2^{-1})^{-1}, \text{where}$$

β = [K∗1] vector of model parameters where
K=M+N−1, where
M is number of explanatory variables and
N is number of alternatives

$\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2$ = estimated parameter vectors in the estimation context 1 and application context 2, respectively (24)

$\sum_t$ = covariance matrix of estimated parameters for context t (t=1,2).

3.7.4 Combined transfer estimation

We can assume that during a short time period an individual's travel behaviour will be rather stable. Thus the true values of the parameters can be expected to be equal $\beta_1=\beta_2=\beta$ (which is the situation the Bayesian method assumes). For longer period or in spatial transfer this assumption is not justified. Then in model transfer there is an error, which is caused by the differences between the coefficients in the estimation and application context. This difference, $\Delta = \beta_2 - \beta_1$, is called the transfer bias and it describes the change in individuals behaviour.

Ben-Akiva and Bolduc (1987) present a generalization of the Bayesian approach which accounts for a non-zero Δ, and which yields the minimum mean squared error estimate of $\hat{\beta}_{transferred}$ achievable from a linear combination of the estimation and application context parameter estimates. This minimum square error estimate is provided by:

$\hat{\beta}_{transferred} = ((\sum_1 + \Delta\Delta^T)^{-1} + \sum_2^{-1})^{-1}((\sum_1 + \Delta\Delta^T)^{-1}\hat{\beta}_1 + \sum_2^{-1}\hat{\beta}_2)$

and covariance matrix

$\sum_{transferred} = \begin{pmatrix} \sum_1^2 & 0 \\ 0 & \sum_2^2 \end{pmatrix}$, where

β = $[K*1]$ vector of model parameters where

$\hat{\beta}_1, \hat{\beta}_2$ = estimated parameter vectors in the estimation context and application context, respectively

$\sum_t$ = estimated parameter covariance matrix for context t ($t=1,2$) (25)

Δ = $\beta_2 - \beta_1$

Δ^T = transpose of Δ.

Comparison of equation [25] with equation [24] indicates that the omitted transfer estimator reduces to the Bayesian estimator in the case of $\Delta=0$. In practice, the unknown transfer bias Δ is approximated by the estimated bias d= $\hat{\beta}_2 - \hat{\beta}_1$. Ben-Akiva and Bolduc also demonstrate theoretically that the combined transfer estimator is superior to simply using the application context parameter estimates $\hat{\beta}_2$, providing the transfer bias, Δ, is small.

3.7.5 Joint context estimation

Joint context estimation is used to estimate a new joint application context model, using both the estimation context and the application context data sets. In econometrics the estimation of models with different data sources is called "mixed estimation". Often these data are divided into two sets: primary and secondary data. The primary data provide direct information about the main modelling parameters. The secondary data provide additional (indirect) information about the parameters. For example, in discrete choice modelling the primary data could be information coming from a survey at the disaggregate level, and the secondary data could be data coming from an aggregate survey. In our case RP data collected in an estimation context constitute the primary set, since these data capture the actual behaviour of the individuals, and application context data constitute the secondary set.

The principle of joint context estimation used here is quite similar to the combined estimation of RP and SP data sets, and is discussed for example in *Bradley and Daly (1991)* and in *Ben-Akiva and Morikawa (1990*). *Badoe and Miller (1995b)* have applied this method also in model transfer.

The following notation is used in developing the joint context estimation procedure:

t = superscript denoting context (=1 for first data set; =2 for second data set)
U_{in}^{t} = utility of alternative i in context t for person n
V_{in}^{t} = deterministic utility for alternative i in context t for person n
ε_{in}^{t} = random component of utility for alternative i in context t for person n
$\boldsymbol{\gamma}$ = vector of utility function parameters assumed to be constant across contexts
$\boldsymbol{s}_{in}^{t}$ = vector of explanatory variables for alternative i common to contexts 1 and 2 (i.e., associated with the constant parameter vector γ), but with values given for person n in context t
$\boldsymbol{\alpha}^{t}$ = vector of utility function parameters assumed to be specific to context t (at a minimum, this includes the alternative-specific constants for context t)
$\boldsymbol{r}_{in}^{t}$ = vector of context-specific explanatory variables for alternative i for individual n within context t
μ = utility function scale for alternative i in context 2 (the context superscript has been suppressed to simplify the notation; context 1 scales are assumed to be "embedded" within α and γ; given this, μ is actually the ratio of the context 2 scale to the constant scale 1 for alternative i, with the absolute values of neither of these scales being identifiable)

$$\boldsymbol{\beta} = [\boldsymbol{\alpha}^{1}\ \ \boldsymbol{\alpha}^{2}\ \ \boldsymbol{\gamma}]^{T} \tag{26}$$

$\boldsymbol{\beta}$ = combined vector of all parameters to be estimated within the joint context model, excluding the utility function scales

$\boldsymbol{x}_{in}^{t}$ = combined vector of all explanatory variables in the joint context model, for alternative i for person n in context t (T is transpose)

$$\begin{aligned} \boldsymbol{x}_{in}^{t} &= [\boldsymbol{r}_{in}^{1}\ 0\ \boldsymbol{s}_{in}^{1}]^{T}, \quad for\ t=1 \\ &\quad [0\ \boldsymbol{r}_{in}^{2}\ \boldsymbol{s}_{in}^{2}]^{T}, \quad for\ t=2 \end{aligned} \tag{27}$$

Given the definitions above, the systematic utility components for the two contexts are:

$$\begin{aligned} U_{in}^{1} &= \boldsymbol{\alpha}^{1T}\boldsymbol{r}_{in}^{1} + \boldsymbol{\gamma}^{T}\boldsymbol{s}_{in}^{1} + \epsilon_{in}^{1} &= \boldsymbol{\beta}^{T}\boldsymbol{x}_{in}^{1} + \epsilon_{in}^{1} \\ U_{in}^{2} &= \boldsymbol{\alpha}^{2T}\boldsymbol{r}_{in}^{2} + \boldsymbol{\gamma}^{T}\boldsymbol{s}_{in}^{2} + \epsilon_{in}^{2} &= \boldsymbol{\beta}^{T}\boldsymbol{x}_{in}^{2} + \epsilon_{in}^{2}. \end{aligned} \tag{28}$$

The vector of coefficients, γ, is common to both estimation and application context models while α^1 and α^2 are specific to each model. Normally, it is supposed that sharing γ, in both models, implies that trade-offs among attributes s are the same in both contexts. However, due to the possible transfer bias in model transfer, the situation is essentially different from the norm. Thus, it may be valuable to combine the common and data-specific variables in a different way than previously, and in such a way to optimally emphasize the estimation and application contexts by being certain to take into consideration the impreciseness of some coefficients. This issue is discussed in Chapter 6.

The multinomial logit form requires that the unobserved effects are independently and identically distributed (IID) across the alternatives in the choice set, according to the Gumbel distribution *(Hensher and Johnson 1991, Ben-Akiva and Lerman 1985).* The combined estimation method assumes that random terms have IID property within each type of data. However, since the effect of unobserved factors may well be different between the two data sets, there is no reason for assuming that ε_{in}^1 and ε_{in}^2 have an identical distribution, or more specifically, have the same variance. In order to make the variances of ε_{in}^1 and ε_{in}^2 equal, the utilities in context 2 are multiplied by an unknown utility function scale μ defined by:

$$\mu^2 = \frac{\mathrm{Var}(\varepsilon_{in}^1)}{\mathrm{Var}(\varepsilon_{in}^2)}. \tag{29}$$

If the random components in each of the data sets follow Gumbel distribution and are identically and independently distributed (IID) then the combined data set also has an IID Gumbel distributed random noise if the utility function in context 2 is multiplied by μ. The final utility function in context 2 is then

$$\mu U_{in}^2 = \mu(\boldsymbol{\alpha}^{2T}\boldsymbol{r}_{in}^2 + \boldsymbol{\gamma}^T\boldsymbol{s}_{in}^2) + \mu\varepsilon_{in}^2 = \mu\boldsymbol{\beta}^T\boldsymbol{x}_{in}^2 + \mu\varepsilon_{in}^2. \tag{30}$$

The multiplication of the utility function in context 2 with the unknown variance factor μ makes the utility function non-linear in parameters. So, ordinary logit estimation methods cannot be used directly. *Bradley and Daly (1991)* have given a method where the Maximum Likelihood estimation of the model coefficients is done with an artificial tree structure. The tree structure used in the present study is shown in Figure 6. In the estimation alternatives of context 1 are given as such, but the alternatives in context 2 are given structured below a dummy alternative. The only variable in the dummy alternative is the logsum of the lower level alternative in context 2. If the coefficients of the logsum variables of the dummy alternatives are forced to be equal then the θ values presented in Equation 31 correspond to the μ factors given in Equation 30.

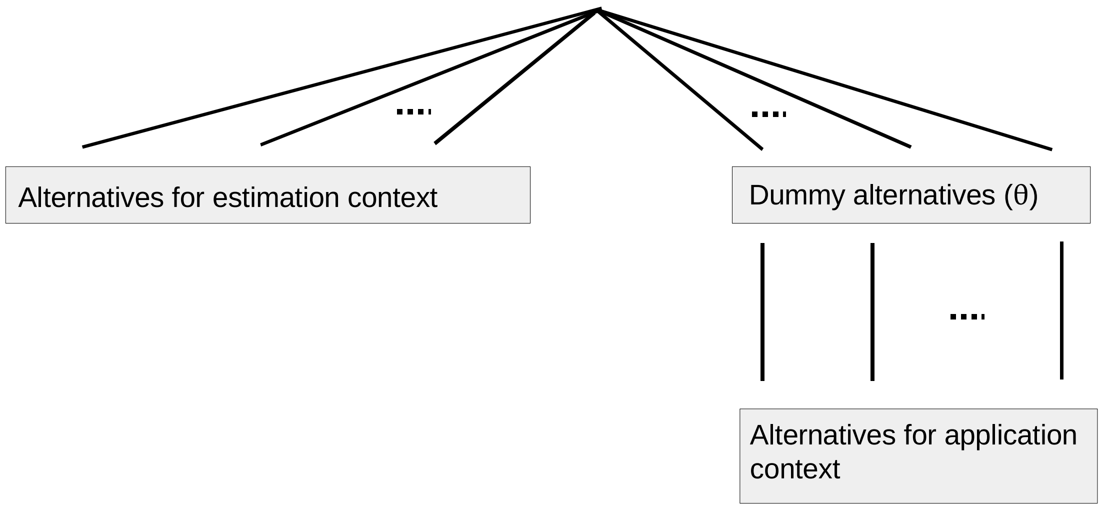


Figure 6: The artificial tree structure used in the combination of different data sets.

For the observations in context 2, the expected maximum utility of each of the dummy composite alternatives is computed as usual in a tree logit model:

$$V_{in}^{COMP} = \theta \ln \sum e^{V_{in}^2} = \theta V_{in}^2. \tag{31}$$

where the sum is taken over all of the alternatives in the nest corresponding to the composite alternative (each nest contains only one alternative in this specification) and

$$V_{in}^2 = U_{in}^2 - \varepsilon_{in}^2. \tag{32}$$

3.8 Goodness of fit measures

3.8.1 Introduction

When a logit model has been specified and the parameters are estimated so that the model fits the observed sample "as well as possible", we are naturally interested in how good that best possible fit is. In general, the validity of the model depends on how well the alternatives can be described (does the model include all the important variables or not). The model variables should be simple, easy to understand and predictable. In addition, studying the effects of changes in a transportation system also has to be possible. The same requirements also apply to model transferability. In addition, it is desirable that the coefficients in the initial and final stages are similar to each other and that the same transport system-related variables are included in both stages.

Several measures used to assess the model accuracy and the effectiveness of model transfer have been formulated in a number of papers (i.e. *Atherton and Ben-Akiva 1976, Koppelman and Wilmot 1982, Koppelman and Rose 1983, Abdelwahab 1991).* The following definitions are based largely on the paper by *Koppelman and Wilmot (1982*). First, some general tests used to measure the goodness of model parameters are presented. Then the most important traditional transferability tests that are used in this research are discussed.

3.8.2 General goodness-of-fit tests

The **likelihood ratio test** is a very general test that is used in nearly all contexts to measure how well the model fits the data. Stated more precisely, the statistic measures of how well the model, with its estimated parameters, performs compared to a model in which all the parameters are zero. This comparison is made on the basis of the log likelihood function being evaluated at both the estimated parameters and at zero for all parameters. The likelihood ratio index is defined as *(Ben-Akiva and Lerman 1985)*:

$$\rho^2(0) = 1 - \frac{L(\hat{\beta})}{L(0)}, \tag{33}$$

where $L(\hat{\beta})$ is a value of log likelihood function at its maximum (when all parameters have been estimated)
$L(0)$ is a value of log likelihood function when all parameters are zero.

The likelihood ratio index ranges from zero, when the estimated parameters are no better than zero parameters, to one, when the estimated parameters allow for perfect prediction of the choices of the sampled decision makers.

There are no general guidelines explaining when a $\rho^2(0)$ is sufficiently high. However, conventionally the $\rho^2(0)$ should be at least 0.15. The value 0.25 is quite good and the value of 0.40 excellent *(Talvitie 1983).* The value should be interpreted carefully, because both the number of alternatives and observations affect to $\rho^2(0)$ value. The value of $\rho^2(0)$ also depends on the shares of alternatives. The greater the proportion of the more preferred alternative is, the greater the values of $\rho^2(0)$ tend to be. It might be better to use $L(c)$ in which the log likelihood function is compared to the model explained only with alternative specific constants (it means the so-called market share model in which all the choice probabilities correspond to the observed distribution), because this value is more comparable across different samples *(Ortúzar and Willumsen 1994).*

Another useful goodness-of-fit measure is derived in Equation (34) remembering the fact that $\rho^2(0)$ will allways increase or at least stay the same when new variables are added to the utility functions. To cope with this problem, an **adjusted likelihood ratio index** is defined as follows *(Lerman 1984):*

$$\rho^2_{adj} = 1 - \frac{L(\hat{\beta}) - K}{L(0)}, \tag{34}$$

where K is the number of variables (including alternative-specific constants).

The evaluation of model parameter preciseness

In model transfer, the differences in coefficients and predictions are caused by differences in initial data creation, by random variation, and by real differences in behaviour between the estimation and application context. Hence, it is important to study the effect of the sample size. However, because the variance of model parameters can not be determined analytically according to the sample size *(Ben-Akiva and Lerman 1985, Rao 1973)* the effect of sample size is studied here by taking random samples from the entire set of data.

If the estimators of model parameters are asymptotically normally distributed, the asymptotic t-test can be used and the confidence interval with significance level 95 % can be stated as *(Ortúzar and Willumsen 1994)*:

$$\hat{\beta}_K \pm 1{,}96\sqrt{\mathrm{Var}(\hat{\beta}_K)}. \tag{35}$$

However, in this study, the confidence intervals are used minimally, because the test measures for different methods are not generally normally distributed. As a result, we have primarily based our analysis on the Bootstrap variation of estimates rather than confidence intervals.

Different transfer methods are generally compared using statistical tests. However, before comparing the effective model, the different combinations used in common and data-specific coefficients in joint context estimation are roughly investigated by calculating the **Mean Deviation-**values (MD) for each model coefficient. The test measure is defined as the percentage Mean Deviation from the coefficients estimated using the full data set.

$$MD = 100 * \frac{\sum_{i=1}^{N} |\hat{\beta}_{i_{transferred}} - \hat{\beta}_{full\ data}| / N}{\hat{\beta}_{full\ data}}, \text{where} \quad (36)$$

$\hat{\beta}_{i_{\mathrm{transferred}}}$ =the coefficient estimated for transferred model i

$\hat{\beta}_{\mathrm{full\ data}}$ =the coefficient estimated for the full data set

N =the number of models.

Based on our own theories, the smaller the MD-value, the more precise is the model coefficient and the more appropriate for it to be used as a data-specific coefficient when using joint context estimation.

The model parameters preciseness is also evaluated through the examination of coefficient ratios. A special case of the relative assessment of different attributes is the ratio R between the assessment of time and cost attributes:

$$R = \beta_{travel\ time} / \beta_{cost} \ . \quad (37)$$

This ratio can also be interpreted as a being value of time (VOT). In this thesis, the value of time is chosen as an indicator because:

- it is a generally understood policy measure and
- it is also reasonable to study the ratio of some coefficients so as to remove the effect of scale parameter μ.

However, it should be noted that with few observations the ratio between two independently distributed standard normal variables can be seen to follow a Cauchy probability density function, which then, with a number of (perhaps even finite) observations, approach Normal distribution. That is, VOT is undefined or unstable in the cases where there are few observations. As a consequence of this the mean of the time-coefficients $\hat{\beta}_{travel\ time}$ over respondents divided by the mean of the cost-coefficients $\hat{\beta}_{cost}$ over respondents is not the same as the mean of their VOT, but the latter is an underestimation of the true VOT *(Armstrong 2001).*

Thus, the ratio of time and cost coefficients[1] is only used to study the model parameters' preciseness. That is, the results should not be interpreted as results related to VOT because many of the advanced theories and estimation problems have not necessary been taken into account for the purpose of this work.

The test measure for this ratio, known as the **percentual VOT error** is calculated by comparing the VOT based on a transferred model with the VOT based on a model estimated from the full application data set. The maximum error we define as being "acceptable" is 25 percent.

$$\mathrm{R}_{\mathrm{error}} = 100 * \frac{R_{transferred} - R_{full\ data}}{R_{full\ data}} = VOT_{error}. \tag{38}$$

3.8.3 Methods to evaluate model transferability

Disaggregate transferability measures

To evaluate if the application and estimation contexts can be represented by a single model **Nested Likelihood Ratio Test (LR)** can be used. The statistic is defined as *(Horowitz 1982)*:

$$\mathrm{LR} = -2[L_{ij}(\hat{\beta}_{ij}) - (L_i(\hat{\beta}_i) + L_j(\hat{\beta}_j))], \tag{39}$$

where

$L\ (\hat{\beta}_i)$ = log-likelihood value in context i using model developed from context i
$L_j\ (\hat{\beta}_j)$ = log-likelihood value in context j using model developed from context j
$L_{ij}\ (\hat{\beta}_{ij})$ = log-likelihood value in context i using model developed from the pooled data from contexts i and j.

The statistic is chi-squared distributed with degrees of freedom equal to the number of model parameters in each model. In the event the null hypothesis is rejected, an asymptotic t-test can be performed to identify which parameter(s) were responsible for the rejection. The null hypothesis is in this case $\beta_i = \beta_j$. If the test statistic is greater than a

1 Instead of pure cost variable, the variable cost/income is used. The denominator incomes are used to manage different time points in which data has been collected (that is: the HMA data is collected in 1995 and the Turku data in 1997). The cost and time are OD-pair dependent and income zone dependent. However, when calculating VOT, the average income over the whole area is used.

critical value, the null hypothesis is rejected and it is reasonable to let the parameters have different values in the model. This statistic is given by (*Ben-Akiva and Lerman 1985)*:

$$t = \frac{\hat{\beta}_i - \hat{\beta}_j}{\sqrt{\mathrm{Var}(\hat{\beta}_i) + \mathrm{Var}(\hat{\beta}_j)}} \ , \tag{40}$$

where $\hat{\beta}_i$ is the estimate of parameter in context *i* and $\mathrm{Var}(\hat{\beta}_i)$ is the variance of the estimate. *Galbraith and Hensher (1982)* recommend the application of this test only to parameters with low standard error (which often imply high t-ratio); otherwise, the t-statistic may reject the alternative hypothesis (i.e the parameters are different) even if they exhibit substantial differences.

A natural measure of transferability of a model estimated in context *i* for application in context *j* is the difference in the log-likelihood between this model and the corresponding one estimated in context *j*. **The Transferability Test Statistic (TTS)** is defined as the absolute value of twice this difference *(Koppelman and Wilmot 1982, Abdelwahab 1991).* It tests the hypothesis that the underlying parameter values in context *i* are equal to the estimated values in context *j*.

$$\mathrm{TTS} = -2[L_j(\hat{\beta}_i) - L_j(\hat{\beta}_j)], \tag{41}$$

where $L_j(\hat{\beta}_i)$ is the log-likelihood of the transferred model and $L_j(\hat{\beta}_j)$ is the log-likelihood of the application context model based on the full data set. It is assumed that when applying the test the parameter vector of the transferred model is fixed, and the distribution of this test statistic is chi-squared with degrees of freedom equal to the number of model parameters. Because the chi-squared distribution is asymmetric, the results will depend on the direction in which the models are transferred.

Transfer Index statistic **(TI)** describes the degree to which the log likelihood of the transferred model exceeds the log-likelihood of a reference model (we use the market shares model) relative to the improvement provided by a model developed in the application context. TI has an upper bound of one which is attained when the transferred model performs as well as an estimated model on the application data. It has no lower bound. Negative values imply that the transferred model is worse than the local base (market shares) model. TI is expressed as *(Koppelman and Wilmot 1982, Koppelman and Rose 1983, Abdelwahab 1991)*:

$$\mathrm{TI} = \frac{L_j(\hat{\beta}_i) - L_j(c_j)}{L_j(\hat{\beta}_j) - L_j(c_j)}, \tag{42}$$

where $L_j(c_j)$ is the log-likelihood of the local model evaluated at the market share values and all other variables are as defined before.

Transfer Goodness-of-Fit Measure describes the degree to which the log likelihood of the transferred model exceeds a reference model (such as the market shares model), relative to the improvement in log-likelihood achieved with a perfect estimation context model, over this reference model. This measure is similar to the goodness-of-fit measure $\rho^2(c)$ which is used for discriminating between different specifications on the same data set. This is expressed mathematically as *(Koppelman and Wilmot 1982, Koppelman and Rose 1983, Abdelwahab 1991)*:

$$\begin{aligned} \rho_j^2(\hat{\beta}_j) &= \frac{(L_j(\hat{\beta}_i) - L_j(c_j))}{(L_j^{*} - L_j(c_j))} \\ &= 1 - \frac{L_j(\hat{\beta}_i)}{L_j(c_j)}, \end{aligned} \tag{43}$$

where L_j* denotes the log-likelihood value obtained when the choices are predicted perfectly, and this has a value of zero. This measure has an upper bound value equivalent to the local rho-square for the application context. Negative values indicate a transfer model performance worse than that of a market share model estimated on the application context data.

The last three measures defined above are interrelated by their dependence on the difference in log-likelihood between transferred and application context models. However, they offer different perspectives on model transferability: TI provides a relative measure to new estimation context model, TTS a statistical test measure and finally the transfer rho-square, an absolute measure based on ideal situation ($L*$).

Aggregate transferability measures

One of the most popular approaches to treat model performance is **sample enumeration**, by which the choice probabilities of each individual in a sample are summed, or averaged, over individuals *(Ben-Akiva and Lerman 1985, Hallipelto 1993)*.

In its simplest form, sample enumeration uses a random sample of the population as ”representative” of the entire population. The predicted share of the sample choosing alternative *i* is used as an estimate for *W(i)*, which is the fraction of population T choosing alternative *i*:

$$\hat{W}(i) = \frac{1}{N_s}\sum_{n=1}^{N_s} P(i|x_n), \text{where} \tag{44}$$

N_s is the number of individuals in the sample.
x_n is defined as all the attributes affecting the choice that appear in the model, regardless of which utility function they appear in.

To forecast the changes in aggregate shares under some policy, one simply changes the values of the appropriate variables for each affected individual in the sample.

In this thesis we use a sample enumeration test and other aggregated elasticity tests to decribe how the variation of model parameters affects the model’s ability to predict changes in travel behaviour when the transferred coefficients are applied to the full application context data set. The test measure, known as the **percentual Relative Sample Enumeration Error (RSEE)** is calculated by comparing the predicted change based on a transferred model with the predicted change based on a model estimated from the full application data set. The maximum error we define as “acceptable” is rather high, 25 percent. The motivation for using rather high error limit is to show that the sample size required to estimate precise models is larger than was previously thought to be the case. In addition, we have tried to find a limit that would react to the sample size in a reasonable way (that is, when using small sample sizes the error limit is quite often exceeded but when using larger sample sizes, the limit is rarely exceeded). The sensitivity analysis of different error limits is presented in Chapter 5.

$$\text{RSEE} = 100 * \frac{\textit{predicted change}_{\textit{transferred}} - \textit{predicted change}_{\textit{full data}}}{|\textit{predicted change}_{\textit{full data}}|} \tag{45}$$

Other elasticity test is undertaken by running the whole four step model system. The test measure is known as RSEEF. Note, in this thesis, RSEEF is referred to as an elasticity test and not as a sample enumeration test because it is not applied to individuals, but to the zonal values determined when running the full forecast process.

Koppelman and Wilmot (1982) have presented various aggregate prediction test statistics, all of which compare in various ways the aggregate number of predicted trips by mode *m* for a given aggregate group *g*, with the observed number of trips by this mode for this group. In this context, only one application of these test statistics is discussed, the **Mean Absolute Error for Forecast** (MAEF) defined as:

$$\mathrm{MAEF} = \frac{\sum_o \sum_d |\hat{N}_{od} - N_{od}|}{\sum_o \sum_d N_{od}} \,, \tag{46}$$

where

$\hat{N}_{od}$ = predicted trips made from zone o to zone d with mode

N_{od} = observed trips made from zone o to zone d with mode.

In this thesis this test is used when running the model system, in its aggregate implementation, to derive base year matrices in division of four aggregated zones. The predicted and observed trips are compared to each other, and the total error over the aggregated groups is calculated.

4 DATABASE, MODEL SYSTEM AND STUDY PROCESS

4.1 Overview of the study structure

The purpose of the research is to compare alternative methods of spatial transfer as a function of sample size and identify the factors affecting the models' quality and the impreciseness of the model's parameters.

The study is mainly based on the travel surveys carried out in the Helsinki Metropolitan area (HMA) in 1995 and in the Turku region in 1997. The HMA travel database is used for the estimation of the models which are to be transferred. The Turku region database represents the application context to which the HMA models are transferred to, for evaluation of transfer effectiveness. The model transfer is studied using two different trip groups, namely home-based work trips (HBW) and other home-based trips (OHB). This makes it possible to study different kinds of combinations related to the model parameters' preciseness and the transfer bias. The studied modes are walking and bicycle, car (driver and passenger), and public transport (bus, train and tram).

The model transfer is studied using all four methods described in Section 3.7. These methods are:

- transfer scaling,
- Bayesian method,
- combined transfer estimation, and
- joint context estimation.

In addition, models are estimated using different sizes of samples from the 1997 Turku data. The models estimated from these samples are referred to as new sample models (Figure 7). The new sample models are used to compare the impact which the additional information used in the model transfer has on the predictive performance of the models in the application context. The performance of the four transfer methods and the new sample models are evaluated for each sample size-model specification combination in terms of how well they replicate the full sample of the 1997 Turku study set of observed trips.

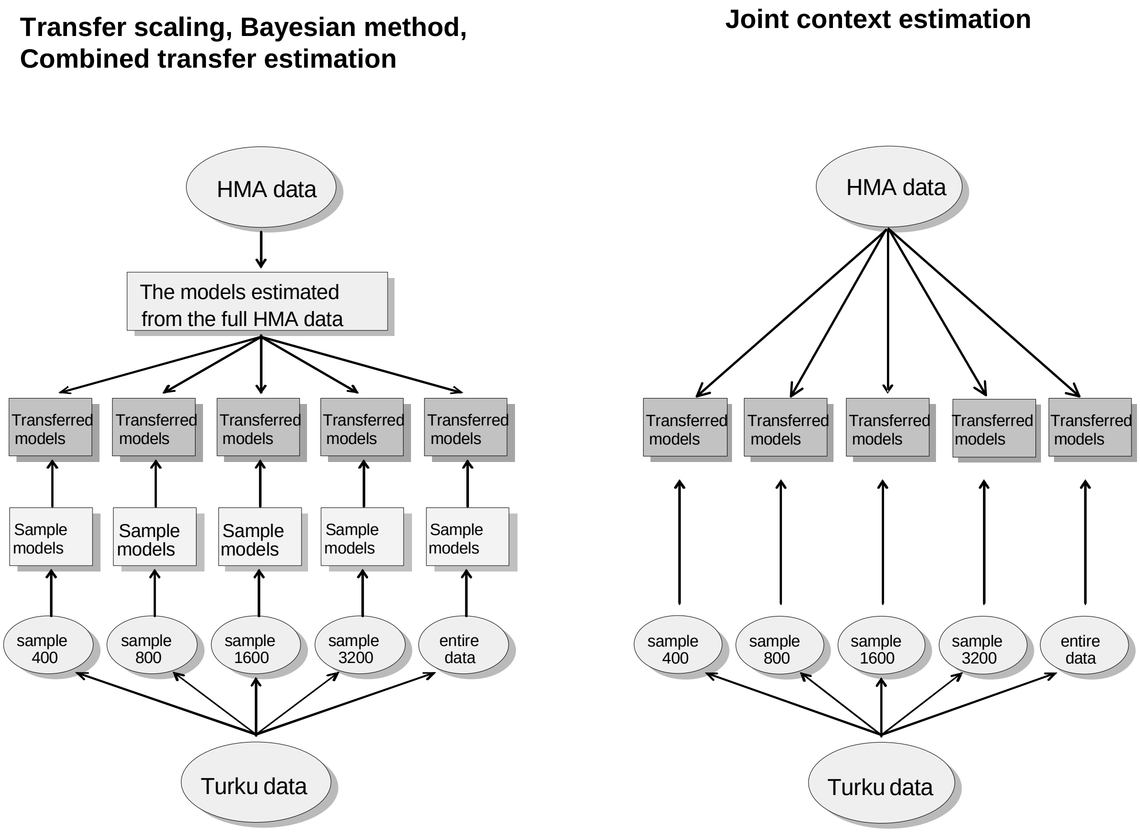


Figure 7: The principle of model transferability analysis.

Model transferability is considered from the viewpoint of the equality of model parameters in estimation and the application context, as well considering the factors, that may cause differences between the estimation and application context. As a large part of the differences in coefficients and predictions is due to random variation, and only part due to real differences in the estimation and application context, the main emphasis of this study is in investigating the relationship between the transfer bias and the impreciseness caused by the sample size of the application context. Different transfer methods are compared to each other, as well as the amounts of data needed to estimate mode and destination choice models.

The study is divided into two main parts. The first part of the thesis (Chapter 5) considers the factors, which decrease model quality and affect the differences between the estimation and application context models. For example the importance of data gathering methods and the sample size are studied. The second part gives (Chapter 6) a general overview of the spatial transferability of the HMA models.

4.2 Study areas

Table 1 presents some statistical background information about the HMA (estimation context), in 1995, and the Turku region (application context) in 1997.

Table 1: The characteristics in the areas studied (Tilastokeskus 1996, Tilastokeskus 1998).

	HMA in 1995	**Turku region in 1997**
population	891,056	240,481
number of jobs	445,800	100,495
land area	742 km^2	698 km^2

The HMA consists of four cities: Helsinki (525,031 inhabitants), Espoo (191,247 inhabitants), Vantaa (166,480 inhabitants), and Kauniainen (8,298 inhabitants). The city centre of Helsinki is located on a peninsula in the Gulf of Finland, and the metropolitan area forms a half circle around it with a radius of 25 to 30 km. In 1995 there were about 91,000 jobs and 63,000 inhabitants in Helsinki's central area. The number of jobs in the whole metropolitan area is about 445,800.

In 1995, the car density in the area was about 320 cars per 1,000 inhabitants. The public transport system in the area consists of bus and tram, commuter and ordinary trains, and one subway line east of the city centre.

The Turku region is one of the five biggest cities in Finland and it is located 150 km west of the HMA. The whole study area consists of the four cities: Turku (168,772 inhabitants), Kaarina (19,309 inhabitants), Raisio (22,854 inhabitants), Naantali (12,769 inhabitants) and the two communities: Lieto (13.138 inhabitants), and Piikkiö (6,367 inhabitants) surrounding it.

In 1997, car density in the area was about 360 cars per 1,000 inhabitants, and 60 percent of all households had at least one car. The public transport system in the area consists of bus, commuter and ordinary trains. Of the 0.8 million internal daily trips made by the inhabitants of the area, 51 percent are made by car, 11 percent by public transport, and 38 percent by bicycle or on foot.

4.3 The database

4.3.1 Travel surveys in the HMA in 1995

The basic travel surveys for the HMA were made during 1987 and 1988. The main field studies were an origin-destination (OD) survey of automobile traffic *(YTV 1990a),* an OD survey of public transport *(YTV 1990b)*, and a mobility survey *(YTV and liikenneministeriö 1990b, YTV and liikenneministeriö 1991).* The mobility survey was repeated in 1995 *(Kaartokallio 1997).* The mobility survey carried out in 1995 was done simply for updating purposes. Hence origin-destination (OD) surveys were not conducted in 1995.

The mobility survey was an individual based study. The objective of the survey was to gather data on daily travel behaviour pattern plus sosio-economic and other background information. The sample size was 8,065 people (0.9 percent of the corresponding population) in 1995. Only people of 7 years of age or older were included in the original random sample.

The data were gathered using an informed telephone interview. This means that the questionnaire plus travel diary for a survey date were sent in advance to people who owned telephone. After the survey date these people were contacted by telephone and the data were directly recorded into a computer system. The number of acceptable telephone interviews was 2,758. In addition, 324 acceptable answers were collected by mail *(Kaartokallio 1997).*

The sample data were corrected for age and sex to represent the total population. The weighted data were used for forecasting and the unweighted data were used in modelling.

In all, the data for modelling consisted of 3,082 respondents. The average trip generation rate calculated from the unweighted model data for internal trips was 2.76 trips/person/day. The total trip generation rate, including external trips, was 2.89 trips/person/day.

Table 2 presents the share of trip groups. The HMA data was divided into four different trip groups, from which, the home-based work trips and other home-based trips are considered in this research. The trips for respondents with an incomplete trip diary (lacking the mode, origin or destination, trip purpose or travel time) are not included in the data. However, respondents with no trips are included in the data.

Table 2: The share of trip groups pertinent to modelling in the HMA in 1995.

Trip group	**Internal trips in the HMA in 1995**	
	no. of trip observations	%
Home-based work trips	1,993	23.5
Home-based school trips	873	10.3
Other home-based trips	4,149	48.8
Non home-based trips	1,479	17.4
Total	8,494	100.0

Table 3 presents the mode shares, by trip groups, for the model data. Approximately 39.2 percent of the trips were made by car, 30.5 percent by public transport, and 30.3 percent of the trips were walk and bicycle trips in 1995.

Table 3: The mode shares by trip groups in the HMA in 1995 (The shaded trip groups are considered in this research).

Mode	**HMA 1995**			
	Home-based work trips	Other home-based trips	Home-based school trips	Non home-based trips
Walk and bicycle	17.7	31.2	65.5	23.5
Public transport	40.8	28.1	26.3	26.0
Car	41.5	40.7	8.2	50.5
Total	100.0	100.0	100.0	100.0

Figure 8 presents the trip length distribution of all trips, home-based work trips and other home-based trips. The average trip length of all trips (including all four trip groups) was 6.5 km. The average trip length of home-based work trips was 9.9 km, and for other home-based trips the figure was 5.7 km.

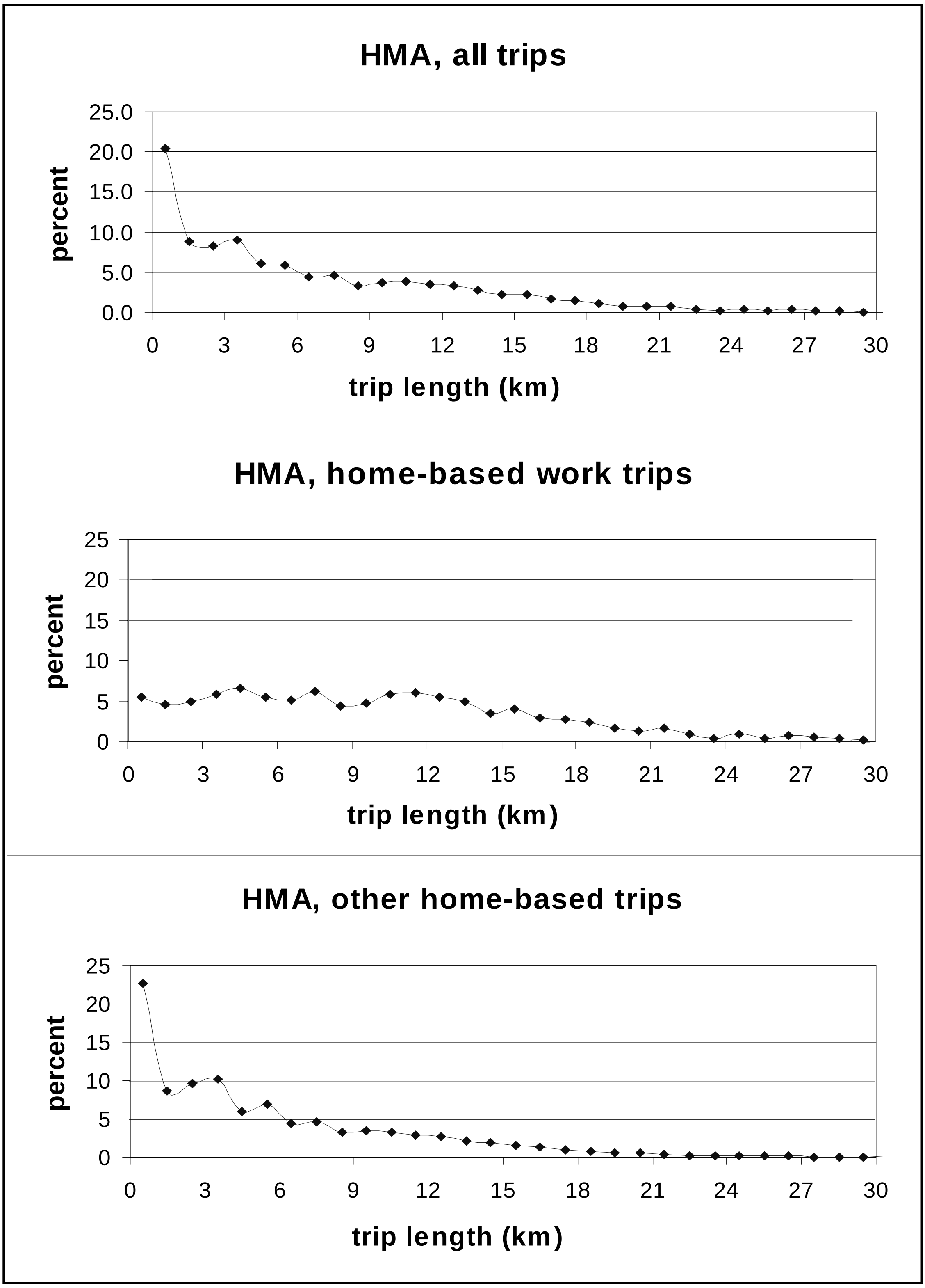


Figure 8: Trip length distribution of the HMA trips in 1995.

4.3.2 Transportation network description in the HMA

The travel times, trip lengths, waiting times and number of transfers used in the modelling are zonal values based on the traffic assignment. We used zonal values because the HMA data was not geocoded and thus had no individual information on travel times and distances.

The car and walk networks were divided into 117 zones (Appendix A). The public transport network was originally based on 596 zones. The models and forecasts were made in a division with 117 zones. That is, when the model and forecasts have been made, the impedances based on the 596-division were converted to a 117-division by combining some zones and calculating the weighted average values of impedances, to represent the corresponding larger area. The number of network links is approximately 2,500 (depending on the scenario) and the number of regular nodes 1,414. A more precise network description is given in *Karasmaa et al. (1997*).

The traffic assignments were made in the Helsinki Metropolitan Area Council using a standard multipath equilibrium model (Emme/2). The trip lengths and travel times were based on a traffic assignment using the 1995 car network and trip matrices, which were based on the forecasts made for the year 1995, using the models based on the 1988 mobility survey. The impedance matrices were produced separately for morning and evening peak periods. A transpose of the morning peak matrix was taken to produce the evening peak hour matrix. The peak periods were defined to be between 6.30-8.30 and 15.30-17.30.

The public transport impedances are based on the Emme/2 multi-path transit assignment. The 1995 public transport network and lines were used. The available modes in the public transport assignment were bus, train and tram.

The intra-zonal travel times were formulated using a regression model based on the travel times of the observed OD-trips under five kilometres. The regression line is determined separately for car and public transport intra-zonal travel times.

The intra-zonal distances were calculated by using a regression model made for the Turku region (see Section 4.3.4). The regression model is based on the distances calculated from the geocoded data in the Turku region. We have used the Turku data because geocoded data was not available in the HMA. Travel times were formulated simultaneously for all modes because consideration of only one mode does not produce enough data to be accurate for modelling purposes.

4.3.3 Travel surveys in the Turku region in 1997

The survey in the Turku region was the first study in Finland in which the origin-destination surveys for internal trips were conducted by postal questionnaires and telephone interviews. Roadside interviews were carried out only in order to collect data on external trips. The aim of the study was to collect an adequate amount of data for travel system planning and to study model transferring *(Turun kaupunki et al. 1999a)*. In addition, an important objective was to systematically compare different data gathering methods *(Kurri and Karasmaa 1999, Turun kaupunki et al. 1999b, Turun kaupunki et al. 1999c)*.

The data were gathered between late October and early December 1997 (October 20-23 the postal questionnaires, October 20-23 telephone interview, and the reminders November 3-6 and 17-20, and December 1-4). The travel diaries were collected between Monday and Thursday. The sample size was exceptionally large, 21,000 persons, which means that every 10th person living in the area (excluding children under seven years) was included in the study. In order to compare different data gathering methods the sample was divided randomly into four sub-samples, one of which acted as a reference. The people in the reference group were sent a postal questionnaire with a two-day trip diary. One reminder-to-respond postcard and one reminder with a new questionnaire with new survey dates were sent to each person. The people in the second group were sent a postal questionnaire with a two-day diary and three reminders. In the third group a one-day trip diary and three reminders were used. The people in the fourth sub-sample were interviewed by telephone (Table 4).

Table 4: The structure of the mobility survey carried out in the Turku region in 1997.

Sample size	Response rate (%)	Method	Number of reminders with a new questionnaire	Number of days
12,000 (reference sample)	49	postal questionnaire	1 reminder	two-day trip diary
3,000	55	postal questionnaire	3 reminders	two-day trip diary
3,000	60	postal questionnaire	3 reminders	one-day trip diary
3,000	55	telephone interview	-	two-day trip diary

Different data collection methods are compared in Section 5.2 and in *Kurri and Karasmaa (1999)*. The data described in the following section is based only on the reference group, which is also used in studying model transferability.

On the whole, the basis of the data for modelling in Turku consisted of 4,675 respondents who had reported 28,315 internal trips (all these respondents were asked to fill in two-day diaries).

The average trip generation rate calculated from the unweighted reference data for internal trips was 3.49 trips/person/day. The total trip generation rate, including external trips, was 3.70 trips/person/day.

Table 5 presents the share of trip groups for the entire set of reference data. The trips for the respondents with an incomplete trip diary (lacking the mode, origin or destination information, trip group or travel time) are not included in the data. However, people with no trip on either of the two days (661 persons) are included in the data.

Table 5: The share of trip groups pertinent to modelling in the Turku region in 1997.

Trip group	**Internal trips of the reference sample**	
	no. of trip observations	%
Home-based work trips	4,442	15.7
Home-based school trips	2,689	9.5
Other home-based trips	13,989	49.4
Non home-based trips	7,195	25.4
Total	28,315	100.0

Table 6 presents the mode shares, by trip groups, for the data used in model estimation. Approximately 51 percent of the trips were made by car and 11 percent by public transport.

Table 6: The mode shares by trip groups in the Turku region in 1997 (The shaded trip groups are considered in this research).

Mode	**Turku region**			
	Home-based work trips	Other home-based trips	Home-based school trips	Non-home-based trips
Walk and bicycle	31.7	42.4	63.4	41.7
Public transport	13.9	11.8	24.1	8.6
Car	54.4	45.8	12.5	49.7
Total	100.0	100.0	100.0	100.0

Figure 9 presents the trip length distribution of all trips, home-based trips and other home-based trips in the Turku region. The average trip length of all trips (including all four trip groups) was 4.8 km. The average trip length of home-based work trips was 6.1 km and, other home-based trips 4.9 km.

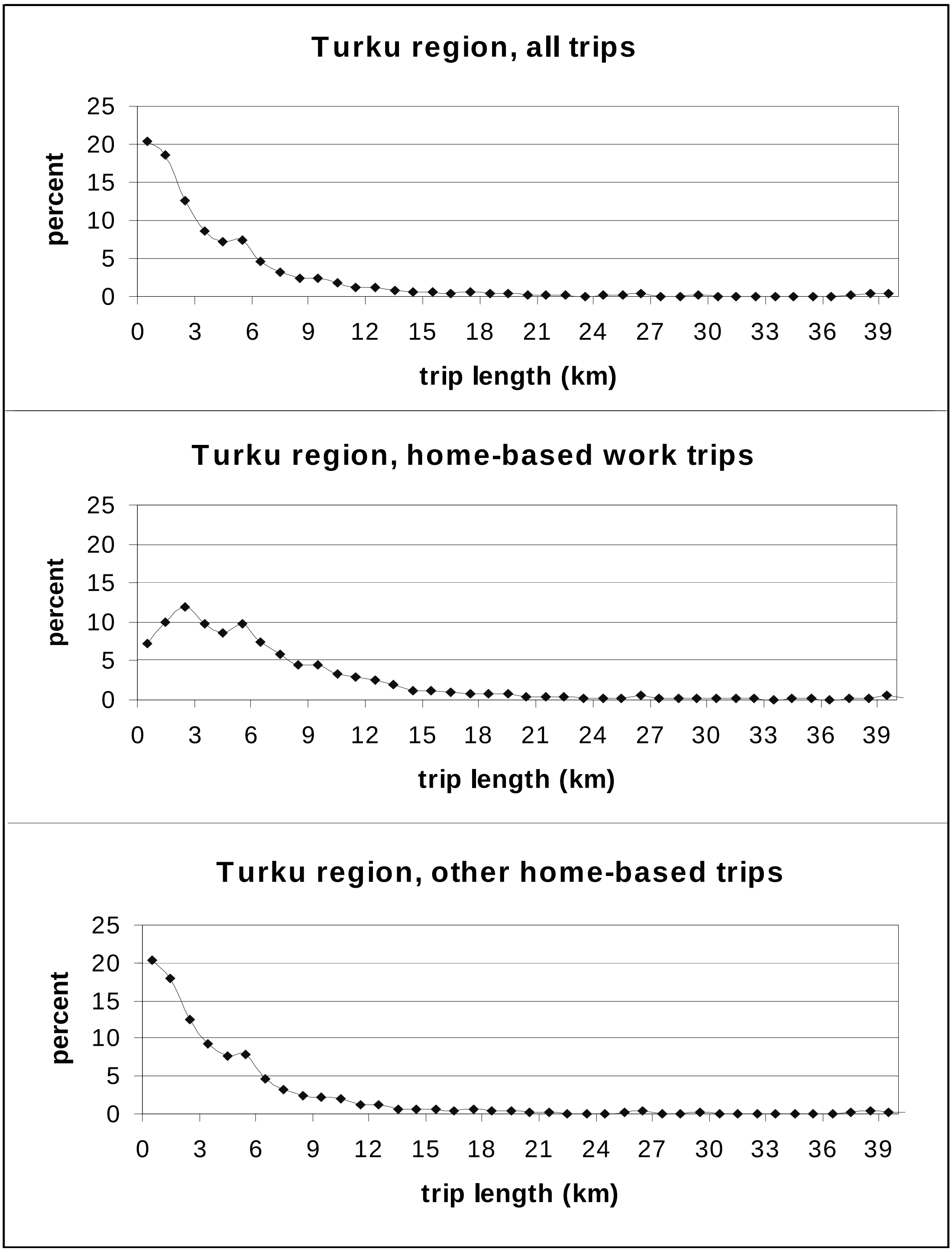


Figure 9: Trip length distribution of all the trips in the Turku region in 1997.

4.3.4 Transportation network description in the Turku region

The travel times, trip lengths, waiting times and number of transfers used in the modelling are zonal values based on the traffic assignment. The car, walk, and public transport networks were made in division of 118 zones (Appendix B). The number of network links is 5,908 and the number of regular nodes is 2,362. The trip lengths and travel times are based on the travel assignment made by the Emme/2 program, using the 1997 travel demand matrices, which are based on the OD-surveys (external traffic) and the mobility survey (internal traffic) carried out in the Turku region in 1997. The network assignment was made by the outside research institution in co-operation of HUT. The role of HUT was to ensure that the network assignment principles are as similar as possible to those used in the HMA. The detailed description of assignment principles can be found in the technical report *(Turun kaupunki et al. 1999d).* In addition, different assignment parameters were studied to observe how sensitive the models are to different definitions. It was found that assignment parameters do not greatly affect the model coefficients of home-based work trips, but taking into account other home-based trips, the coefficient for travel time varied greatly when the “number of transfers” was included in the mode choice model.

The intra-zonal travel times were calculated using a regression model based on the travel times of the observed OD-trips under five kilometres. That is, the travel distances based on the traffic assignment for OD-trips under five kilometers are used as an explanatory variable and the corresponding travel times have been used as a dependent variable. The regression line is determined separately for car and public transport intra-zonal travel times. When estimating the regression model for public transport travel times, the average value of initial waiting times is indicated as the intersection point of the y-axis. Hence, it is assumed that for the minimum length of a trip, at least one waiting time is included.

The intra-zonal distances were calculated by using regression model, which is based on the distances calculated from the geocoded data in the Turku region. The squared land area is used as an explanatory variable.

4.4 Model system

4.4.1 Models

The present travel demand model system in the HMA and in the Turku region is basically a traditional four-step model (trip generation, distribution, modal split and assignment) with feedback between the last three stages *(Karasmaa et al. 1997)*. The trip generation ”model” is currently based on simple cross-tabulation of the data. The most important categorization is the division according to the person’s accessibility to a car. Mode and destination choice are mostly modelled using nested logit models and sequential estimation *(Ben-Akiva and Lerman 1985).* Table 7 represents the model types used. For network assignment a standard multipath equilibrium model (Emme/2) is used. Trips are divided into four categories according to their purpose: home-based work trips, home-based school trips, other home-based trips, and non-home-based trips. This study concentrates on home-based work trips and other home-based trips. In mode choice modelling there are three alternative modes: walk and bicycle, car, and public transport. The access trips to rail are made on foot or by bus. Less than 2 percent of rail passengers use park-and-ride. These trips are coded based on the main mode choice (usually rail) respondents indicated they used.

Table 7: The model types used to forecast internal trips in the HMA and in the Turku region (only the shaded trip groups are considered in this thesis).

Trip group	**Trip generation rate**	**Mode choice**	**Trip distribution**
Home-based work trips	trips/working person/day	logit model	nested logit model
Home-based school trips	trips/school-aged person/day	distance matrix	logit model
Other home-based trips	trips/person, age > 6	logit model	nested logit model
Non home-based trips	trips/person, age > 6	logit model	nested logit model

The trip generation rates are calculated from the sample data by dividing the number of trips in each trip group and population group by the relevant population. There are two population groups in home-based work trips: working persons age 18-64 years and other population, and three in other home-based trips: age 7-17, EHAP-persons age 18- and HAP-persons (HAP-person is a person that practically always has access to a car for personal trips. Others are EHAP-persons).

The mode and destination choice models are mainly logit models. The same mode and destination choice models are used for different time periods of the day. However, different impedances were applied to different time periods. The time periods used were:

- morning (6.30 - 8.30)
- evening (15.30 - 17.30)
- day (any other time).

The models and forecasts for the HMA are based on 117 zones. In addition, variable cars per household was constructed by using 19 zones. In the Turku region a division with 118 zones is mainly used. A division of 23 zones is used for constructing the variables cars per household and income.

The model structure used for mode- and destination choice modelling is presented in Figure 10. The estimation of the destination choice models is based on a (simple) random sample of 25 sub areas, including the chosen destination. The mode and destination choice models are basically estimated sequentially. The differences between the sequential and simultaneous estimation are considered in Appendix C. Due to the logsum coefficients over one when estimating destination choice models for other home-based trips, the inverse model structure was also tested. However, the prediction performance of these models was not as good. The "artificial" tree structure used in joint context estimation has already been presented in Section 3.*7*.

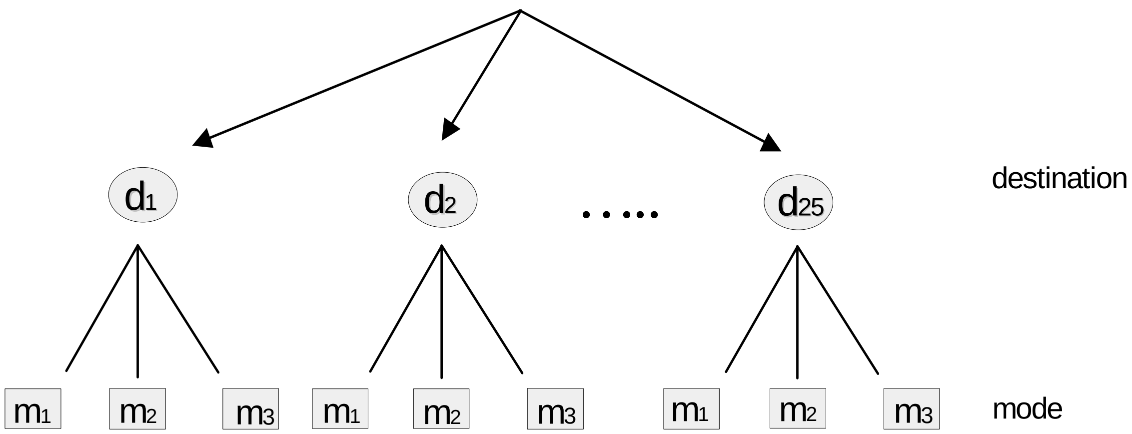


Figure 10. Structure of combined mode and destination choice models.

4.4.2 Forecasting process

The whole four-step forecast process is presented in Figure 11. The regional travel forecast interface was originally developed for the modelling purposes required for the Turku region and it applied the same principles as were used in the HMA in 1988 (*YTV 1990a*) and 1995 (*Karasmaa et al. 1997*). In this research, the current forecasts (forecast based on the data of the time of the survey) are made to evaluate the transferred model's prediction accuracy and its ability to predict changes in a transportation system. The model system is applied separately for each trip group. Appendixes D and E present the forecasting process according to the trip group.

The dark grey boxes in Figure 11 mark the repeated elements of the process when testing different transfer methods and sample sizes. The light grey boxes mark the constant elements such as trip generation rates and the HAP/EHAP model. That is, the trip generation rates are calculated only once using the entire set of data and the same trip generation rates are used for every transfer method and sample size. So, the error caused by the impreciseness of trip generation rates is not taken into account when testing the current forecast. However, this is considered separately in Section 5.3.2. The HAP/EHAP model is also only produced once for the entire set of data. Hence, these two error sources are ignored in this review and only the errors caused by mode-and destination choice level are considered.

The main steps of the forecast are

1. Estimation of the car-availability model for other home-based trips by using the model presented in Appendix E. The definition HAP has been used for a person who has a driving license and whose family has a car and who has, during an interview, stated that, he/she always or nearly always has a car available. Other people belong to the EHAP-group. (The car ownership model has not been used for home-based work trips).
2. Cross-tabulation of trip production rates. The data are extended to respond to the whole population, that is those over 7 years of age, and then the mode shares and trip generation rates are cross-tabulated from this weighted data. The weighting is done according to sex, age and the location of the residence.
3. Multiplying the trip production rates by the corresponding population groups to get the absolute trip production for each zone.
4. Estimation of mode and destination choice models.
5. Splitting the travel demand of each trip groups for different zones and modes (according to the mode- and destination choice models).
6. Multiplying the destination probabability matrix (formulated in the previous step) by the absolute trip production vector calculated in step 4 => the total travel demand matrix for each trip group.
7. Splitting the total travel demand matrices for different time periods.
8. Multiplying the total travel demand matrixes represented for each time group by mode choice probabilities.

Normally, the forecast process is made iteratively, to include the network assignment for the modelled travel demand matrix and then the whole process is repeated until the predefined accuracy of the travel demand matrix is achieved. However, these steps, which are indicated by the white boxes, are not included in this study. The forecast run does, however, include the iterative process (Equation 47) so as to correct the alternative-specific constants, so that the modelled modal shares represent the modal shares based on the weighted mobility survey *(Talvitie 1981)*.

$$D_{new} = D_{old} - \frac{P_j - s_j}{s_j(1-s_j)}, \tag{47}$$

where D_{new} is the new dummy-coefficient, D_{old} the original dummy, P_j is the proportion of mode j based on the estimated model and s_j the proportion based on the weighted mobility survey data.

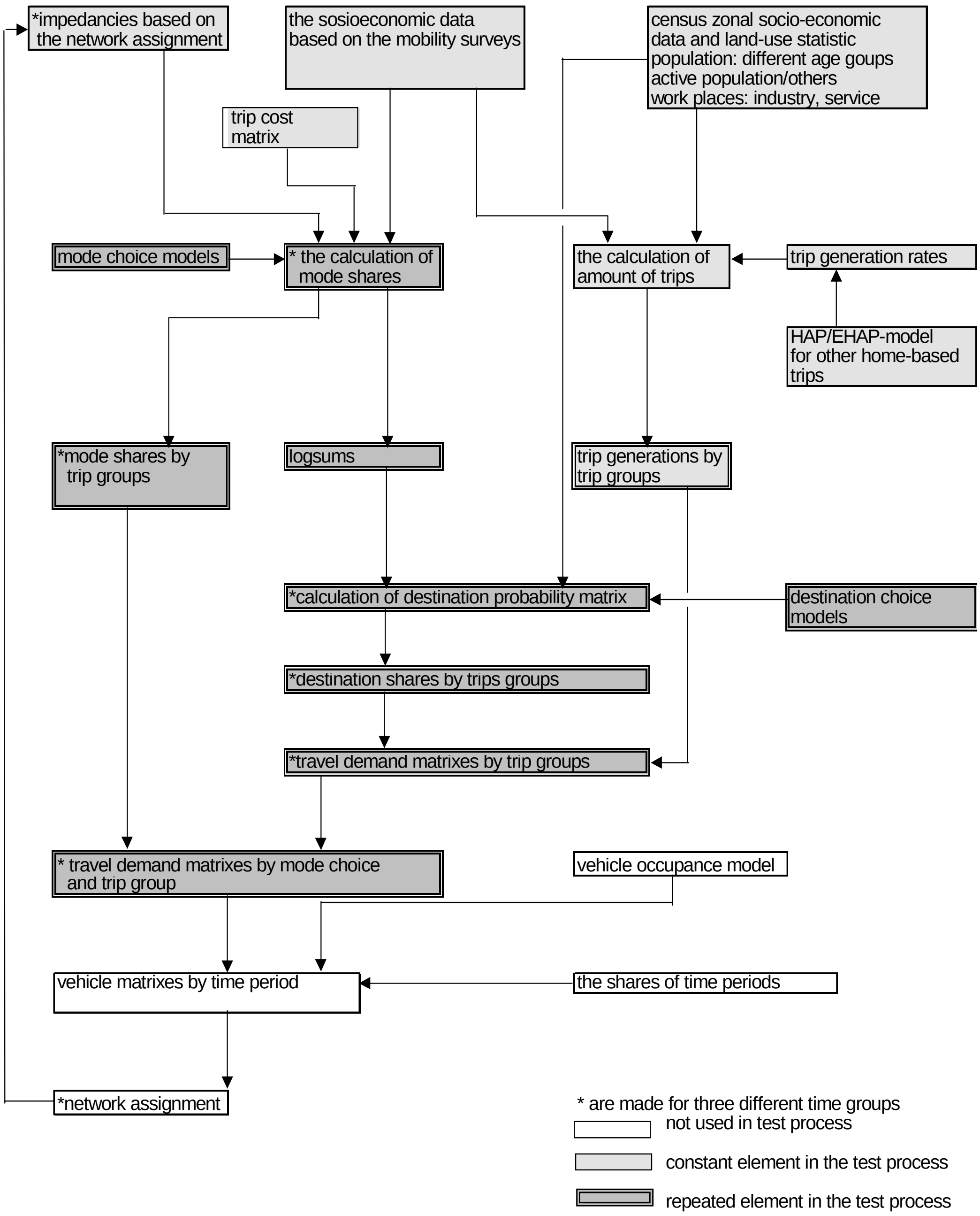


Figure 11: The calculation of the forecasts (user interface of the forecasting system) (YTV 1990a, Karasmaa et al. 1997).

4.5 The principles of the study

The model transferability and preciseness are studied by using the mode and destination choice models presented in Chapter 6. The HMA data and the “reference” data collected in the Turku region are used to study model transferability. The HMA travel database is used for estimation of models which are to be transferred. The database of Turku region represents the travel context to which the HMA models are transferred to, for evaluation of transfer effectiveness. Different factors affecting impreciseness of model coefficients are studied by using the data collected in the Turku region. The data quality issues are considered in Chapter 5. Next, a short description of the model transfer study is given.

The model transferability is examined in particular from the perspective of mode and destination choice models. The transferability of trip generation rates has not been studied. That is because in Finland, trip generation rates have traditionally based on the cross-tabulation instead of logit formulation (see Section 5.3.2.1). However, we have studied the sample size required to calculate trip generation rates based on the small samples. The trip generation rates are considered in Chapter 5 and the results of model transfer in Chapter 6.

The transferability of mode and destination choice models is studied by all four methods described in Section 3.7. The studied methods are:

- transfer scaling,
- Bayesian method,
- combined transfer estimation, and
- joint context estimation.

Scale factor is used for distance variables and for variables “travel time” and “number of transfers” in transfer scaling. In addition, the joint context estimation is studied by using different kind of combinations of common and data-specific coefficients.

The model specifications investigated in this work are presented in Table 8 and the definitions of the variables in Table 9. The distance variables for walking trips are given in kilometres for a one-way trip between the origin and destination. Many different distance functions were examined based on the log-likelihood values before the choice of a piecewise linear function with a point of inflexion of 5 km was made. The travel times, travel costs, and numbers of transfers are for a round trip. Travel times are in minutes, and costs are in euros (1 euro=5.946 FIM). Parking costs (for one-way trips) have been cross-tabulated from the mobility survey based on the costs car drivers had paid, and are included in the car costs. Travel costs for public transport are zonal values, which have been cross-tabulated from the mobility survey data based on the passenger ticket type. Travel costs for cars are based on 0.10 euro/km in the HMA in 1995 and 0.12 euro/km in the Turku region in 1997. The formulation of cost variables is considered more precisely in Appendix F and in *Karasmaa et al. (1997)* and *Turun kaupunki et al. (1999d).* In estimation, generic coefficients for time and cost variables are used, because the mode-specific coefficients did not differ statistically from each other. The only exception was when estimating mode choice models for home-based work trips. In this case, the cost coefficient for the car differed slightly from that estimated for public transport. However, if the cost coefficient was assumed to be generic (as is usually the case), the travel time coefficient for the car was equal to the coefficient estimated for public transport. The use of generic coefficients is further justified by the multicollinearity problems, which are discussed in more detailed in Appendix C.

Table 8: Model specifications for 1995 HMA and 1997 Turku area databases.

Variable	Home-based work trips	Other home-based trips
Distance 0-5 km (Walk)	X	X
Distance > 5 km (Walk)	X	X
Total travel time (Car, Ptr)	X	X
Number of transfers (Ptr)		X
Trip cost/income (Car, Ptr)	X	X
Cars/household (Car)	X	X
Walk dummy (Walk)	X	X
Car dummy (Car)	X	X
Log sum (Dest.)	X	X
Ln(jobs) (Dest.)	X	
Size variable		
- Population		X
- Service employment		X
- Retail employment		X

Walk=Walk and bicycle
Ptr =Public transport
Car =Car

The size variable is the weighted sum of the inhabitants and jobs in the area. Rather simple models, particularly for home-based trips, are used so as to be able to forecast the values of chosen variables in the future. That is, simple models are assumed to be more convenient when doing forecasts and complex models, if the behaviour is studied. The coefficient of size variable is fixed to one (based on the theoretical consideration of the collective choice of destination points). The attraction variables included in the size variable are normally logarithmic, so that the model is independent of the zoning system. Thus, the logarithm expression is usually referred to as a size variable and the coefficients of its components are known as weight-factors. The coefficients for the components of size variable are, in principle, estimated in the same way as the other model coefficients. Nevertheless, in this case, note that t-values from the estimation results do not test the deviation from zero as the coefficient has been estimated inside the exponent function.

Table 9: Definition of variables specified in mode and destination choice models.

Variable	Definition
Distance 0-5 km (Walk)	One-way distance for walk and bicycle trips between 0 and 5 km. If the distance is > 5 km, the value of variable is 5.
Distance > 5 km (Walk)	= 0 for walk and bicycle trips between 0 to 5 km. = distance-5 for walk and bicycle trips over 5 km.
Total travel time (Car)	Round trip total travel time (including walk access times, min)
Total travel time (Ptr)	Round trip total travel time (including walk access times, in-vehicle time, waiting times, min)
Number of transfers (Ptr)	Round trip number of transfers
Trip cost/income (Car)	Round trip out of pocket travel cost (including parking cost) divided by monthly household income/1000 (zonal average)
Trip cost/income (Ptr)	Round trip out of pocket travel cost divided by monthly household income/1000)
Cars/household (Car)	Number of cars per household (zonal average)
Walk dummy (Walk)	= 1 for walk and bicycle mode; =0 otherwise
Car dummy (Car)	= 1 for auto mode; =0 otherwise
Log sum (Dest.)	Natural log of the denominator of the mode choice model
Ln(jobs) (Dest.)	Natural log of number of jobs in the destination area (zone)
Size variable	Natural log of weighted numbers of population and jobs. (For OHB different groups of jobs are used)
- Population	The amount of inhabitants in the area
- Service employment	The number of jobs in service industries
- Retail employment	The number of jobs in retail industries

Studying the effect of sample size

To explore the impact of sample size on transferring perfomance, model transferability is tested using three to four different sample sizes. Consequently, all transferability tests have been carried out by using 100 samples (resampled from the entire set of 1997 Turku data) for each trip group, transfer method and sample size category. The sample sizes used by trip groups are presented in Table 10. The respondents who did not make any trips are not included in the data used in mode and destination choice modelling. The number of respondents is fixed in each sample category, and the number of trips variates around the average value presented in the table.

The resampling is performed by using bootstrap (see appendix G for more specific information). That is, bootstrap samples are created by randomly resampling observations from our original sample of n observations with replacement *(Efron and Tibshirani 1993)*. In each drawing, each observation has the same probability of being drawn, 1/n. This resampling procedure implies that some observations will be drawn several times and some not at all. The advantage of using bootstrap is that the variance in the coefficients can also be studied for the entire set of data. However, the main reason for using bootstrap, other than sampling without replacements, is that this procedure represents the situation in which samples would have been drawn from the whole population.

Table 10: The sample sizes and average numbers of trips in the study.

Turku region in 1997 (entire set of data: 4,675 respondents and 28,315 trips)*			
Home-based work trips (4,442 trips)		**Other home-based trips (13,989 trips)**	
respondents	average no. of trips	respondents	average no. of trips
425	400	140	400
850	800	275	800
1,700	1,600	550	1,600
3,400	3,200	1,100	3,200
4,675	4,450	2,300	6,400
-	-	4,675	14,000

* two-day diaries used

Test of transferability

The testing process is divided into two phases. First the best possible way to transfer models by using joint context estimation is defined by comparing the model coefficients, based on small samples, to the coefficients estimated from the entire set of application context data. Nine different combinations for using common and data-specific variables are tested. The combinations (models A-I) studied in mode choice level are presented in Table 11. The model definitions used in model transfer are presented in Table 8.

Table 11: The combinations of data-specific variables for mode choice models tested in the joint context estimation. The data-specific variables are marked by "X" and letters A to I refer to different models.

	Data-specific coefficient								
Variable	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Distance 0-5 km (Walk)	-	X	X	X	X	X	X	X	-
Distance > 5 km (Walk)	-	-	X	X	X	X	X	X	X
Total travel time (Car, Ptr)	-	-	-	-	-	X	-	X	-
Number of transfers (Ptr)	-	-	-	-	-	-	-	-	-
Trip cost (Car, Ptr)	-	-	-	X	X	-	-	X	X
Cars/household (Car)	-	-	-	-	X	-	X	-	X

The initial premise was, that almost all possible combinations would be tested. However, during the pre-testing some very unsatisfactory combinations were dropped. For example the number of transfers is always estimated as being common because the coefficient of this variable is so imprecise. This means that the estimation context data have to be emphasized (the use of common variables emphasize the estimation context more than the use of data-specific variables).

In addition, two different combinations of common and data-specific variables are used for the destination choice level, for other home-based trips. First, all variables are kept as common. Then the components (service employment, retail employment) of the size variable (see Table 8) are estimated as data-specific. The destination choice model, for home-based work trips, only includes one unconstrained coefficient (ln(jobs)), thus there are no different combinations to be tested for this trip group.

The best possible way to transfer models by using joint context estimation is evaluated by comparing the model coefficients, based on small samples, to the coefficients estimated from the entire set of application context data. The test measure is the percentual Mean

Deviation (MD) from the coefficients estimated using the entire set of data. The model, which produces the minimum error calculated as a percentual Mean Deviation from the coefficients estimated from the entire set of data, is chosen to represent the joint context estimation.

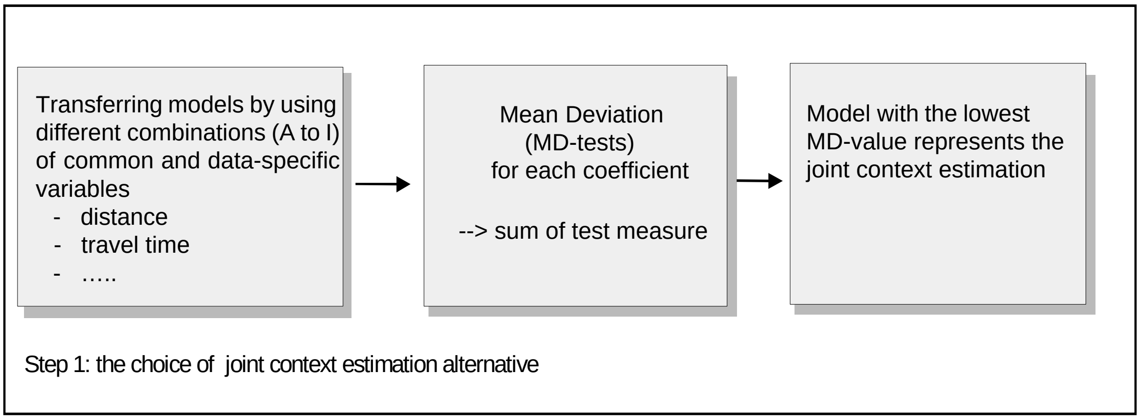


Figure 12: The choice of the best way to use common and data-specific parameters in a joint context estimation.

Secondly, after choosing the best possible way to use joint context estimation, the different methods are compared to each other. Two approaches are undertaken to determine model transferability: One compares model parameters between demand models, and the other compares the forecasting results with the observed values. The steps in the testing process are presented in Figure 13.

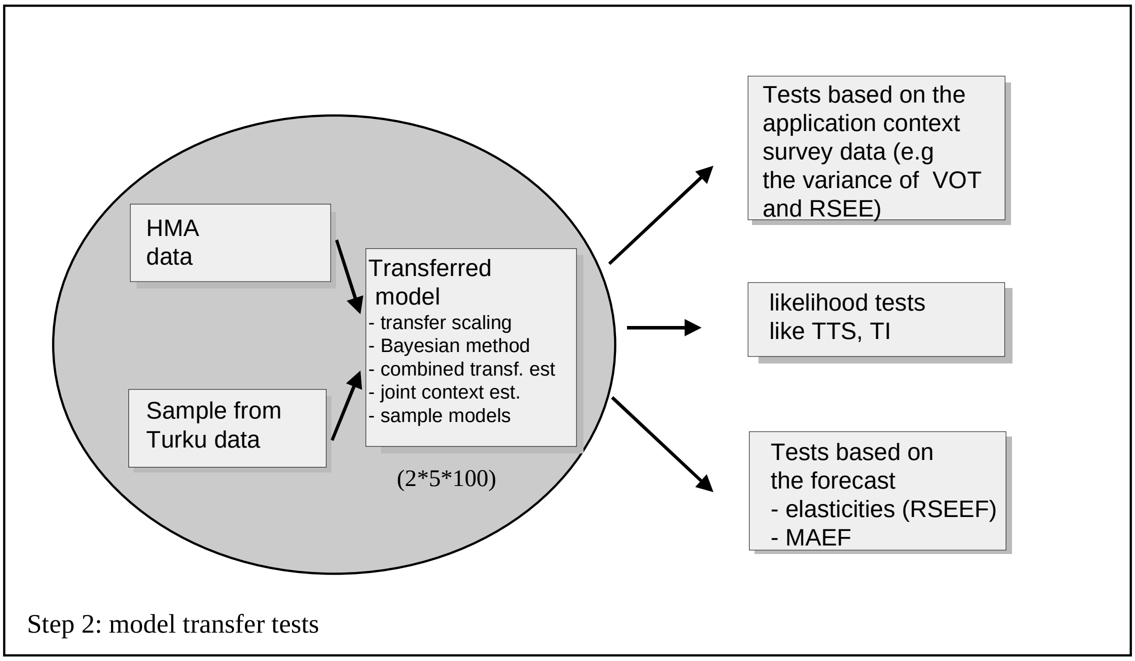


Figure 13: The comparison process of different transfer methods.

Model validation is carried out at three levels. Firstly, validation tests of the disaggregate models are conducted through examination of the coefficient ratios, e.g. **values of time**, and production of elasticities with respect to a cost and time variables. Two **elasticity tests** are conducted. Firstly, in order to test the cost elasticity of the mode choice models, a test is carried out to examine the effect of increasing car costs by 10 %. Secondly, the effect of a 30 percent increase in public transport travel times is studied. Transfer effectiveness is measured by calculating the proportion of models in which the error of VOT, or **sample enumeration error** (RSEE), is less than 25 percent. Note that the error limits defined may, in some cases, affect the conclusions. This is especially so if models based on one given method slightly exceed the limit and the same method is slightly under the limit when a different test measure is used. This is, why the average relative sample enumeration errors are also calculated and presented in the appendixes.

Secondly, the disaggregate measures such as TTS and TI are used to test the model parameters' equality in two contexts. **The Transferability Test Statistic (TTS)** is used to test the difference between the transferred model and the model estimated from the full application context data set. **The Transfer index (TI)** is used to describe the disaggregate transfer prediction accuracy or goodness-of-fit relative to the corresponding goodness-of-fit of a similarly specified model estimated in the application context. The test statistic is described more specifically in Section 3.8.

Finally, validation tests are undertaken by running the model system in its aggregate implementation to derive base year matrixes and to extract the elasticities with respect to cost and time variables at this level. The predicted and observed trips are compared to each other, and the total error, known as **MAEF**, is calculated over four aggregated groups. The effects of a 30 percent increase in the public transport travel time on the public transport share, as well as the effects on car share of a 10 percent increase in travel costs by car, are studied also in this aggregate level. The relative error for these tests is labelled **RSEEF.**

Table 12: Summary table of transfer evaluation measures used in the case study.

Quantity	**The description of the field of application**
Mean Deviation (MD) (Equation 36)	This measure describes the mean deviation of the transferred coeffients from the set of coefficients based on the full application context data set.
The percentage of observations with the error of VOT under 25 percent (Equation 38)	This measure describes the accuracy and preciseness of VOT.
Relative Sample Enumeration Error (RSEE, RSEEF) (Equation 45)	This measure describes the transferred models' ability to predict the effects of changes in travel system. RSEE means that the test has been applied by using the full application context data set. RSEEF means that elasticities are based on the running of the whole model system.
Transfer Test Statistic (TTS) (Equation 41)	This tests the hypothesis, if the parameter values in context i are equal to the parameter values in context j.
Transfer Index (TI) (Equation 42)	This compares the predictive performance of the transferred model to the performance of a similarly specified model, estimated on the application context sample only.
MAEF (Equation 46)	This is the mean of the absolute values of the forecast error based on the entire traffic forecast data.

5 DATA-COLLECTION METHOD AND SAMPLE SIZE AS SOURCES OF VARIATION IN MODEL COEFFICIENTS

5.1 Introduction

This chapter presents some special issues related to the modelling work. According to *Horowitz (1981)*, the errors concerning disaggregate, random-utility demand models can be divided into two categories: A distinction is made between errors that cause correct models to yield incorrect forecasts (such as errors in aggregating disaggregate models or in forecasting explanatory variables), and those that cause models to yield incorrect forecasts, even if the models are used correctly (for example, sampling, specification, and data errors). In model transfer, the above mentioned errors may cause apparent differences between the estimation and application context parameters making it difficult to conclude whether the observed transfer bias is real or not. Therefore, in this chapter the uncertainty caused by the latter class of errors is examined due to its importance in model transfer. First, the issues related to data gathering methods are considered. After that, the sample size requirements in modelling are investigated.

5.2 The influence of data gathering methods

5.2.1 Data and methods

The aim of this section is to analyse how the data gathering methods affect the quality of the data, indicated by the number of trips reported by the respondents, and other key variables. A comparison of data gathering methods is presented here, to see if different methods used in the HMA and in the Turku region affect the results in relation to model transferability. The HMA data used in the model transfer, was collected by using telephone interviews and one-day diaries. In the Turku region, the data used in the model transfer, was based on postal questionnaire and two-day diaries.

The effect of data gathering methods is studied using the special data collected in the Turku region for this purpose. That is, before the data collection, the random sample was divided into four different sub-samples according to the data gathering method employed. One of these samples, namely the sample collected by postal questionnaires with one reminder is used in the model transfer. Other sub-samples are used in studying the effect of data gathering methods. All these samples have already presented in Table 4 in Section 4.3.3. The data gathering method is studied both at the trip generation level and at the mode choice level. The main issues to be compared are:

- postal questionnaire vs. telephone interview,
- two-day vs. one-day diary,
- the effect of three reminders vs. the effect of one reminder only.

The data quality is studied by comparing the basic statistical parameters as response rates and the number of trips and mode shares as well as by using statistical tests. A t-test has been used in mode choice level to test if the coefficients estimated in two contexts differ significantly from each other. A likelihood ratio test (Equation 39) has been used to see if the two contexts can be represented by a single model. The model's ability to predict the effects of change in traffic is examined by sample enumeration tests. The effects of a 30 percent increase in travel time on the public transport share, as well as the effects of a 10 percent increase in car costs are studied. We employed an "unrealistically" high percentage to study the increase in travel time in public transport in order to examine the differences between transfer methods, and not to study the effects of travel time in practice.

5.2.2 Results

Comparison of trip-related data

The distributions of trip groups and modal split according to the data gathering method in the Turku region are presented in Figures 14 and 15. It is obvious that more bicycle and walk trips were reported in the telephone interview than in the postal questionnaires. However, the differences were smaller than in many similar studies. The distribution of trip purpose groups was quite similar in different data gathering methods. The most significant difference was in the telephone interviews, which included more other home-based trips (53 %) and less non home-based trips (21 %) than the postal questionnaires. It can probably be explained by the fact that a greater share of bicycle and walk trips were reported in telephone interviews. Most short trips are other home-based trips. It must also be noted that in practice the data will be weighted for age and sex to represent the total population. In such cases the mode and trip distributions may change. The results presented here are based on unweighted data, which are also used in estimating mode- and destination choice models. The questions concerning data weighting are discussed in Section 5.3.2.

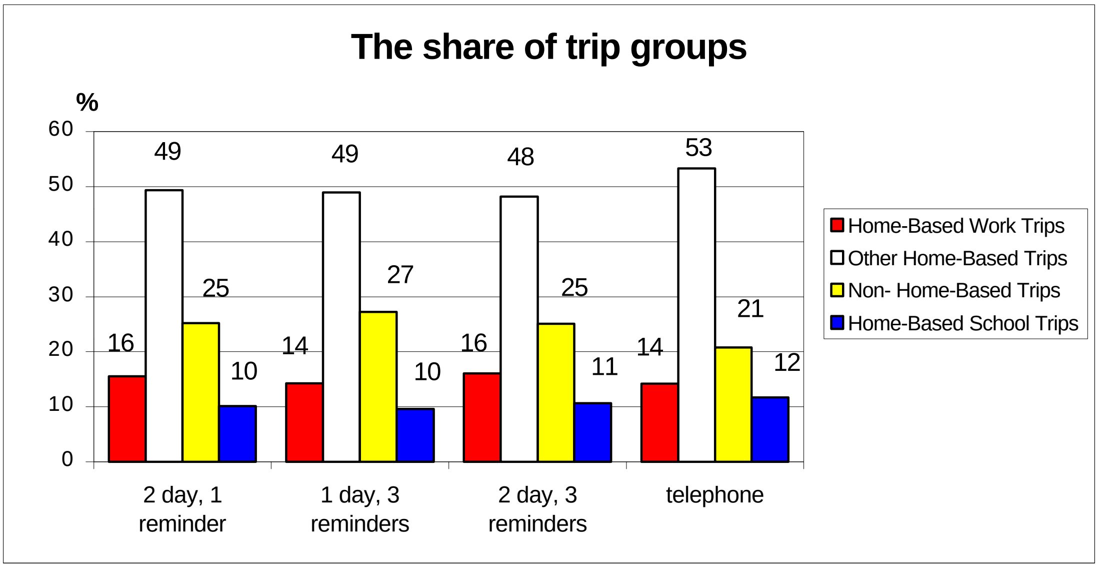


Figure 14: The shares of trip purpose groups according to the data gathering methods in the Turku region.

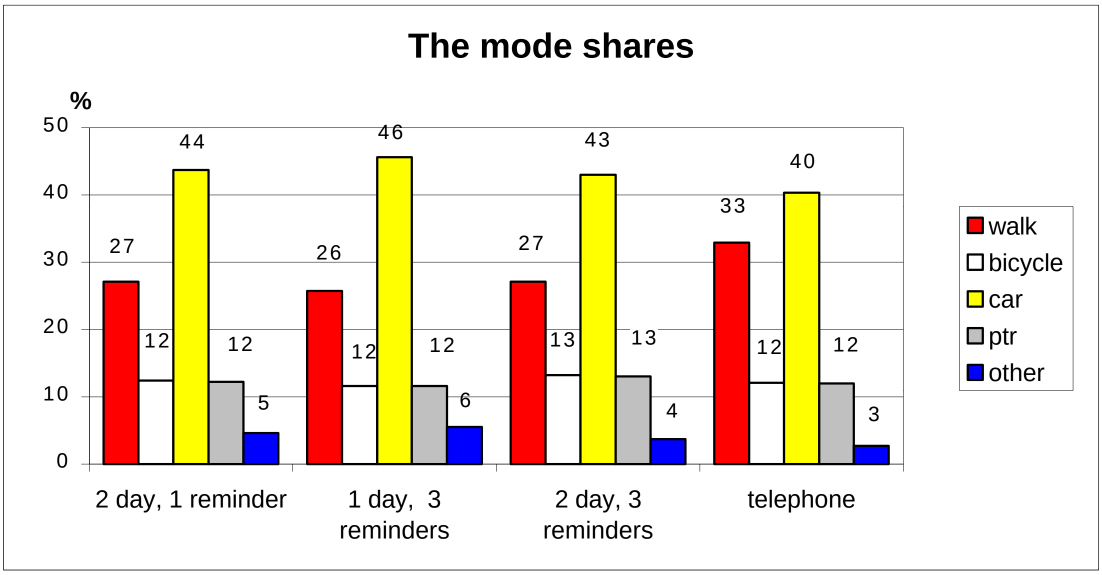


Figure 15: The mode shares according to the data gathering methods in the Turku region.

The trip generation rates

Table 13 presents the weighted average trip generation rates defined by the trip groups used in the Turku region. (There were two population groups for home-based work trips: working persons age 18-64 years and others, and three for other home-based trips: age 7-17, EHAP-persons age 18- and HAP-persons). In the first part of the table the trip generation rates calculated from the first day observations are presented and in the middle part, those representing the second day. In the third part of the table the trip generation rates are calculated based on the assumption that the personal trip generation rate is the average of these two days, and the first and second day observations are not independent of each other. This method is also used in the model transfer study. The average trip generation rate for these two days is not the average of the first and second day because the number of respondents on the first day differ from the number of respondents on the second day; that is, the respondent burden stops some people from informing us about the trips made on the second day.

On the whole, the trip generation rates based on first day observations were somewhat higher than those based on second day observations. This is mainly due to the respondent burden on the second day. The use of postal questionnaires created greater differences for other home-based trips than the use of telephone interviews. That was probably due to the ability of the interviewer in the phone interview to motivate and help the respondent disclose all the trips made. The trip generation rates calculated as an average of the two days were quite similar to the rates calculated from the first day observations being, however, slightly smaller than the first-day trip generation rates, but this difference was not of great importance.

According to the results the total trip generation rate did not vary very much, but the share of people with “no trip” (Table 14) was higher (13.5 percent) in the telephone interviews than in the postal questionnaires (8.5 percent). This is because, in the mail survey, especially in the early stage of the response, it is likely that those with “no trip” probably thought that it was not worthwhile returning the questionnaire. On the other hand, walk and bicycle trips were better reported during telephone interviews (Figure 15). Thus, the non-response error and, the better representativeness of the walk and bicycle trips by using telephone interviews maybe “compensated” each other and the total trip generation rate did not vary excessively between the methods.

Table 13: The trip generation rates (trips/person/day) by the data gathering method (Kurri and Karasmaa 1999).

The first day	**2 days, 1 reminder**	**1 day, 3 reminders**	**2 days, 3 reminders**	**Telephone**
Total trip generation rate	3.58	3.43	3.51	3.56
Home-based work trips age 18-64 (working population)	1.25	1.13	1.29	1.26
Home-based work trips others aged >6	0.07	0.07	0.07	0.05
Other home-based trips age 7-17	1.38	1.50	1.27	1.79
Other home-based trips, HAP*	1.87	1.82	1.76	1.92
Other home-based trips, EHAP*	1.80	1.57	1.79	2.06
The second day				
Total trip generation rate	3.02	-	2.95	3.32
Home-based work trips age 18-64 (working population)	1.11	-	1.09	1.13
Home-based work trips others aged >6	0.06	-	0.07	0.06
Other home-based trips age 7-17	1.13	-	1.09	1.67
Other home-based trips, HAP*	1.62	-	1.52	1.70
Other home-based trips, EHAP*	1.42	-	1.41	1.64
The two day average	**2 days, 1 reminder**	**1 day, 3 reminders**	**2 days, 3 reminders**	**Telephone**
Total trip generation rate	3.50	-	3.46	3.54
Home-based work trips age 18-64 (working population)	1.25	-	1.29	1.23
Home-based work trips others aged >6	0.07	-	0.07	0.05
Other home-based trips age 7-17	1.34	-	1.27	1.77
Other home-based trips, HAP*	1.86	-	1.72	1.87
Other home-based trips, EHAP*	1.73	-	1.73	1.91

* HAP is a person who mainly use car for travelling

* EHAP is a person who usually cannot use a car to travel

Table 14: The amount of persons with "no trip" and of the people who made more than four trips in the different data gathering methods studied (Kurri and Karasmaa 1999).

	Trips	Initial question-naire (1st call)		1st reminder (2nd call)		2nd reminder (3rd call)		3rd reminder (4-9 call)	
2 day, 1 reminder	0 >4 total	630 2 383 7 512	8.4 % 31.7 %	202 316 1 368	14.8 % 23.1 %	-	-	-	-
1 day, 3 reminders	0 >4 total	139 565 1 803	7.7 % 31.3 %	45 92 409	11.0 % 22.5 %	18 28 136	13.2 % 20.6 %	5 16 84	6.0 % 19.0 %
2 day, 3 reminders	0 >4 total	100 373 1 080	9.3 % 34.5 %	27 47 213	12.7 % 22.1 %	12 20 96	12.5 % 20.8 %	7 8 56	12.5 % 14.3 %
2 day, 1 reminder, telephone	0 >4 total	145 234 966	15.0 % 24.2 %	58 106 458	12.7 % 23.1 %	23 45 192	12.0 % 23.4 %	10 45 126	7.9 % 35.7 %

Table 14 shows that the share of people with "no trip" does not appear to decrease noticeably after the first reminder in postal questionnaires. The share of people who have made more than four trips decreases in postal questionnaires when more reminders are sent out; however, in telephone interviews the number of calls does not appear to play so important a role. This is the opposite of the preconception that reaching people who make many trips requires more attempts than reaching less mobile people.

Results based on the mode choice models

Tables 15 and 16 contain the mode choice models for home-based work trips and other home-based trips defined by the data gathering methods. The model specifications have already been explained in Section 4.5.

The models cannot be compared, as such, because a great deal of the variation between the coefficients is due to random variation (due, for example, to the small sample) and only a part is caused by the differences between the data gathering methods. However, all the models are presented in order to provide a better overview of the situation, and to enable us to evaluate the differences between the methods as well as can be done. In addition, the first and second-day models for the data based on the two-day diaries with one reminder are presented, so we can better evaluate the effect of two-day diaries versus one-day diaries. In addition to the number of observations, the mode shares are also presented because the modal share of the least represented mode definitely affects model accuracy. The mode shares may also give some additional information about the possible reasons why model parameters differ from each other.

Table 15: Estimation results of home-based work trips using different data gathering methods; estimated from data collected in the Turku region in 1997.

	Postal Survey					**2 day, Telephone Interview**
	2 day, 1 reminder			**2 day, 3 reminders**	**1 day, 3 reminders**	
	total	**1st day**	**2nd day**			
	A	**B**	**C**	**D**	**E**	**F**
Distance 0-5 (Walk)	-0.878 (-24.2)	-0.848 (-17.0)	-0.914 (-17.0)	-1.034 (-13.5)	-0.896 (-9.2)	-0.778 (-8.4)
Distance > 5 (Walk)	-0.339 (-11.0)	-0.371 (-8.6)	-0.298 (-6.7)	-0.375 (-6.3)	-0.441 (-4.4)	-0.540 (-6.7)
Total travel time (Ptr, Car)	-0.022 (-8.2)	-0.022 (-5.8)	-0.022 (-5.7)	-0.029 (-5.1)	-0.023 (-3.3)	-0.041 (-5.0)
Walk-dummy (Walk)	2.137 (11.4)	2.067 (7.8)	2.242 (8.3)	2.014 (5.0)	1.949 (4.0)	0.166 (0.3)
Car-dummy (Car)	-0.506 (-2.6)	-0.481 (1.8)	-0.576 (-2.0)	-1.515 (-3.8)	-0.720 (-1.4)	-1.760 (-3.3)
Cars/household (Car)	0.658 (3.8)	0.633 (2.7)	0.707 (2.8)	1.189 (3.6)	0.798 (1.8)	1.094 (2.8)
Trip cost/income (Ptr, Car)	-0.884 (-7.2)	-0.994 (-5.8)	-0.723 (-4.1)	-1.157 (-4.8)	-0.875 (-2.9)	-1.698 (-5.3)
Number of trip obs.	4 442	2 347	2 095	1 223	665	758
mode share distribution						
* walk and bicycle	1 407	760	647	369	175	209
* car	2 417	1 272	1 145	651	388	443
* public transport	621	315	303	203	102	106
$\rho^2(c)$	0.2011	0.1985	0.2044	0.2208	0.2190	0.1598
VOT (Car, Ptr)	21.4	19.0	26.1	21.5	22.6	20.7
Sample enumeration (RSEE)						
Ptr travel time + 30 %	-4.8	-4.7	-4.9	-7.0	-5.6	-7.5
Car cost + 10 %	-0.8	-1.2	-0.9	-1.5	-1.1	-2.1

Walk=walk and bicycle
Ptr=public transport (bus)
Car =driver or passenger

Table 16: Estimation results of other home-based trips using different data gathering methods; estimated from data collected in the Turku region in 1997.

	Postal Survey					**2 day, Telephone Interview**
	2 day, 1 reminder			**2 day, 3 reminders**	**1 day, 3 reminders**	
	total	**1'st day**	**2'nd day**			
	A	**B**	**C**	**D**	**E**	**F**
Distance 0-5 (Walk)	-1.038 (-49.0)	-1.050 (-36.4)	-1.023 (-32.7)	-0.884 (-22.0)	-0.937 (-17.6)	-0.948 (-20.7)
Distance > 5 (Walk)	-0.247 (-13.5)	-0.260 (-10.4)	-0.232 (-8.6)	-0.343 (-7.5)	-0.425 (-6.6)	-0.312 (-7.4)
Total travel time (Car, Ptr)	-0.014 (-7.3)	-0.014 (-5.6)	-0.013 (-4.7)	-0.017 (-4.2)	-0.021 (-3.8)	-0.018 (-3.9)
Number of transfers (Ptr)	-0.437 (-10.5)	-0.495 (-8.4)	-0.378 (-6.4)	-0.301 (-3.7)	-0.543 (-4.4)	-0.435 (-4.6)
Walk-dummy (Walk)	2.157 (20.4)	2.148 (14.8)	2.173 (13.9)	2.033 (9.2)	1.516 (5.1)	2.116 (8.2)
Car-dummy (Car)	-0.893 (-7.7)	-0.864 (-5.5)	-0.922 (-5.4)	-0.798 (-3.5)	-0.857 (-2.8)	-0.901 (-3.5)
Cars/household (Car)	1.173 (13.3)	1.106 (9.3)	1.251 (9.6)	1.110 (6.0)	0.949 (4.2)	0.969 (4.9)
Trip cost/income (Car, Ptr)	-1.739 (-20.4)	-1.809 (-15.6)	-1.655 (-13.1)	-1.147 (-7.1)	-1.716 (-7.7)	-1.734 (-8.5)
Number of trip- obs.	13 989	7 609	6 380	3 587	2 164	3 042
mode share distribution						
* walk and bicycle	5 928	3 279	2 652	1 494	835	1 493
* car	6 407	3 459	2 942	1 659	1 086	1 226
* public transport	1 651	871	780	434	243	323
$\rho^2(c)$	0.2265	0.2323	0.2198	0.2126	0.2199	0.2216
VOT (Car, Ptr)	6.9	6.6	6.7	12.7	10.5	8.9
Sample enumeration (RSEE)						
Ptr travel time +30 %	-2.3	-2.3	-2.3	-2.8	-3.1	-2.6
Car cost + 10 %	-1.5	-1.5	-1.4	-1.0	-1.4	-1.5

Walk=walk and bicycle

Ptr=public transport (bus)

Car =driver or passenger

The differences between the coefficients were compared with t-tests. In addition, the differences between the model parameters were tested by comparing the two models estimated in context i and j to see if they are equal to the model estimate with the pooled data from context i and j with the likelihood ratio tests (Equation 39). The results of the statistical tests are presented in Table 17. However, note that due to repeated measurements (see Section 5.3.3.3 and Appendix H), t-values are higher than they would be for independent observations and the null hypothesis of the equality of the coefficients to be rejected too easily. When comparing the one-day and two-day interviews there are also different number of repeated observations between these two data sets.

Table 17: The test statistic of the difference between the mode choice model coefficients by different data gathering methods.

The Data Gathering Method	**t-test**	**$\chi 2$ -value***
Home-Based Work Trips		
The first day versus the second day	-	6.7
Telephone interview versus postal questionnaire (2 day, 3 reminders)	distance 0-5 km (t-value 2.1)	**22.3**
One day (E) versus two day (D) postal questionnaire	-	8.4
One reminder (A) versus three reminders (D)		10.4
Other Home-Based Trips		
The first day versus the second day	-	2.2
Telephone interview versus postal questionnaire (2 day, 3 reminders)	trip cost/income (t-value 2.3)	**51.8**
One day (E) versus two day (D) postal questionnaire	trip cost/income (t-value 2.1)	**16.7**
One reminder (A) versus three reminders (D)	trip cost/income (t-value 3.1) distance 0-5 km (t-value 2.6)	**30.5**

* The critical risk level value of 0.05; 14.1 in home-based work trips and 15.5 in other home-based trips.

According to the results, based only on the home-based work trips, the models estimated via the telephone interview and the postal questionnaire differed from each other due to the differences in the distance variable. It also seems to be the case that the level of coefficients for time and costs is higher when using telephone interviews than when using postal questionnaires, inspite of the fact that the differences are not statistically significant.

When we consider the other home-based trips, the results based on the one-day diary also differed from those based on the two-day diary and the number of reminders significantly affected the model parameters as well. The table shows that the "cost per income" coefficient had, in all cases, the biggest deviations between the models. The difference between the coefficients of cost per income was highly problematic. When the two-day, three times reminded data were used, without the data collected by the second and third reminder, the difference still existed (this examination was made to ascertain the effect of reminders particularly for this "two-day, 3 reminders" sample data). Thus, it can be concluded that the reminders, at least, did not cause all the differences. When the model was estimated using the individually observed values instead of the zonal values or using only the cost variable, without the denominator income, these differences did not exist. That is, the differences between the coefficients are probably derived from the use of zonal mean values for income variable.

5.2.3 Conclusions

The research suggested the general opinion that the telephone interview is usually a more reliable method of collecting information than the postal questionnaire. The walk and bicycle trips, in particular, were better reported in telephone interviews, and also "no trip" responses were better captured in the telephone survey.

The differences between the data collection methods were greater when crosstabulating mode shares and trip generation rates than when considering the mode and destination choice models. The differences between the trip generation rates or mode shares are not very important in our case, because the main focus is studying the transferability of mode and destination choice models. Thus, the forecasts, which are used when testing the whole model system, are made by using constant trip generation rates which are calculated using the entire set of application context data. In addition, the forecast process includes the iteration of the alternative-specific constants, so that the mode shares represent the weighted mode shares in the area. However, generally when estimating totally new models or transferring models the correct trip generation rates and mode share distribution have a necessarily greater importance. Therefore, in this case, the telephone interview should be preferred.

The differences in quality between one- and two-day-questionnaires seem to be quite small. Nevertheless, the lower response rate and the respondent burden appear to play a greater role when two-day-questionnaires are used. The advantage of a two-day diary is the large trip sample size with reasonable costs. This is especially useful in the kind of surveys done in the Turku region, where additional origin-destination field surveys were not carried out for internal trips, but the travel demand matrix was formulated using mobility surveys. It must also be noted that the two-day diary does not increase the number of respondents. Consequently, the larger number of trips based on two-day diaries does not increase the statistical accuracy to the same extent as the number of respondents increases does, because the trips made during the first and the second day are not totally independent of each other. About 43 percent of the home-based trips and 11 percent of the other home-based trips made in the first day were also made in the second day. (The trip was regarded as being the same if the type of origin or destination was the same on both the first and second day and if the main mode-choice was the same on both days. In addition, the departure time was required to vary by no more than two hours on both days).

With regard to model transfer, it can be said that the survey method appears to affect the results at many different levels. However, the differences e.g. between the postal and telephone interviews are not so important when studying model transferability, but they do have more importance in real transfer situations. There was also evidence, that the data gathering method can affect the mode choice level, thereby suggesting that some of the differences between the coefficients in the estimation and the application context discussed in Chapter 6 may be caused by the data gathering method and not just by real differences in local conditions in the HMA and in the Turku region.

The results were quite similar to other international studies, although the differences between the methods were smaller than in most previous studies. Maybe, due to the time resources of the outside survey organization the telephone interview was not planned as carefully as it should have been (the telephone interview and the postal questionnaire had the same questionnaire form). In spite of that, the results relating to the response rate, the share of people with no trip and trip generation rates can be considered quite reliable; however, more detailed analyses concerning the mode choice models can not be made because the sample size achieved by the telephone interview is too small.

5.3 The importance of sample size

5.3.1 Introduction

Choosing the sample size is another important aspect of the planning and design of the survey. Samples may be chosen on the basis of the maximum acceptable levels of error in certain information to be collected, or on the basis of the sample size required for model building in travel demand modelling.

Smith (1979) and *Stopher (1982)* have proposed methods for estimating the required sample size for given levels of accuracy required from a cross-classification model of trip generation. It is relatively simple to specify the sample-size requirements for a given accuracy of trip rate. However, sample sizes tend to be more difficult to determine for calibration of trip distribution models, and there have been some significant differences of opinion on the sample-size requirements for calibrating logit models of mode choice *(Stopher 1991*). No method has been adopted by the profession for determining the required sample size for a household travel survey. Instead, sample sizes are determined sometimes on the basis of statistical assessment (such as minimum sample sizes for certain jurisdictions within the survey region), and sometimes by simply determining how many observations can be obtained for the available budget.

In this chapter we consider the sample size required for mode and destination choice modelling, as well as for the trip generation level. The aim is to consider the sample size in the first three steps of the traditional four-step model (excluding trip assignment) and to evaluate which part of the model system is in a critical path in the model transfer. However, our purpose has not been to give any precise sample size recommendations. The purpose has rather been to examine different criteria and evaluate the accuracy which can be reached using these criteria.

The study is based on the reference data collected in the Turku region in 1997. The importance of sample size is studied with respects to:

- how much data is required to cross-classify trip generation rates (Section 5.3.2),
- how much data is required to determine accurately mode shares (Section 5.3.2),
- what is the sample size required in mode and destination choice modelling and how (the absolute) sample size affects the variation of mode and destination choice model parameters (Section 5.3.3), and
- how the use of a two-day diary instead of a one-day diary affects the sample size that is necessary for estimating mode and destination choice models.

All these factors affect the final outcome of model estimation and transfer. The first three items are directly connected to the issues concerning the data requirements in model transfer. By comparing the two-day diaries to the one-day diaries we also try to evaluate the possible effects of different kinds of data in an estimation and application context. In our case, a one-day diary was used in the HMA and a two-day diary in the Turku region.

5.3.2 The sample size required in a trip generation level

5.3.2.1 Problem description and methodology

Next, the sample size required to accurately define the trip generation rates, as well as for finding the correct trip shares, is considered. The trip generation models have been calculated using simple cross-tabulations, which were made separately for each sample examined. Under the circumstances, the trip generation rates have not been actually transferred, but have been calculated in every trip group and sample size studied.

The trip generation rates are calculated from the sample data by dividing the number of trips in each trip group and population group by the relevant population. There are two population groups in home-based work trips: working persons age 18-64 years and other population, and three in other home-based trips: age 7-17, EHAP-persons age 18- and HAP-persons (HAP-person is a person that practically always has access to a car for personal trips. Others are EHAP-persons). The preciseness of trip generation rates undergoes two phases of testing. First, the trip- and mode shares of unweighted data are calculated. Second, the data are extended to respond to the whole population and then the mode shares and trip generation rates are calculated again from this weighted data. The weighting was done according to five age groups. To test the required sample size the trip generation rates calculated from the samples are compared with the trip generation rates calculated based on the full data set. We allow a maximum deviation of 5 percent from the real trip generation rate (a trip generation rate based on the average of 100 bootstrap samples using the whole dataset). A one percent maximum deviation is allowed for mode shares (thus, if the correct mode share is 25 percent, the values between 24 to 26 are allowed). The sample size is always given as a number of respondents unless otherwise stated.

5.3.2.2 Results

Figure 16 presents the trip generation rates and 95 percent confidence intervals (based on the normal distribution assumption) for home-based work trips by trip group and class for weighted data. Figures 17 to 18 present the proportion of the 100 samples with the error below 5 percent. The sample size requirement is defined with respect to the result which can be reached with the entire set of data. Thus; first, the number of "full" samples, based on the bootstrap samples, not exceeding a 5 percent deviation, is defined for the whole set of data. Then, a number, 5 percent less, is defined for the samples, e.g. if the number of "full" samples which do not exceed a 5 percent deviation is 99 percent, the number required for the samples not exceeding the 5 percent deviation is 94. The "relative" criteria with respect to the full sample is used because it is difficult to give any exact values which should be gained, and in many cases even by using the entire set of data results can be regarded as being too imprecise. Hence, we try to find an acceptable sample size after which results show only a small level of improvement.

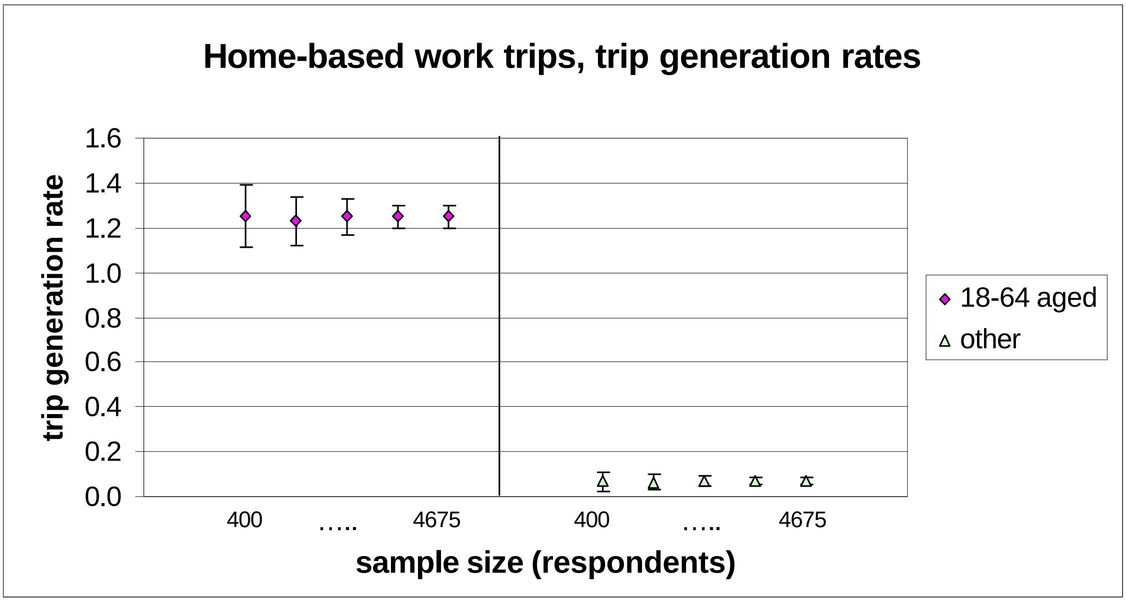


Figure 16: *The variation of trip generation rates and 95 % confidence intervals for home-based work trips as a function of sample size based on 100 bootstrap samples and weighted data.*

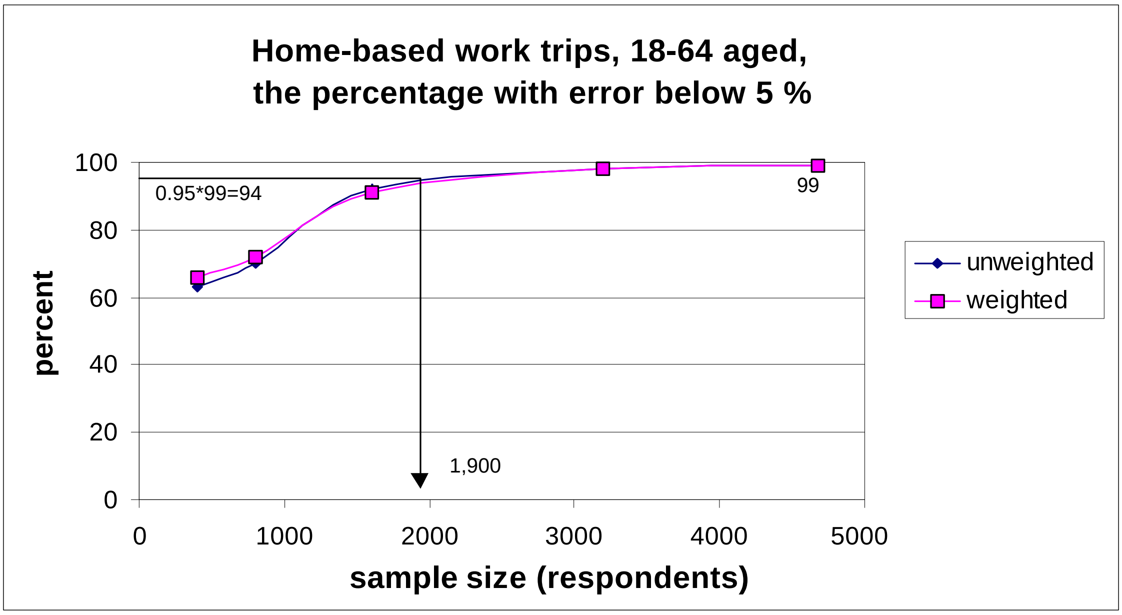


Figure 17: *The percentage of trip generation rates with the error below 5 percent as a function of sample size; home-based work trips, 18-64 aged persons.*

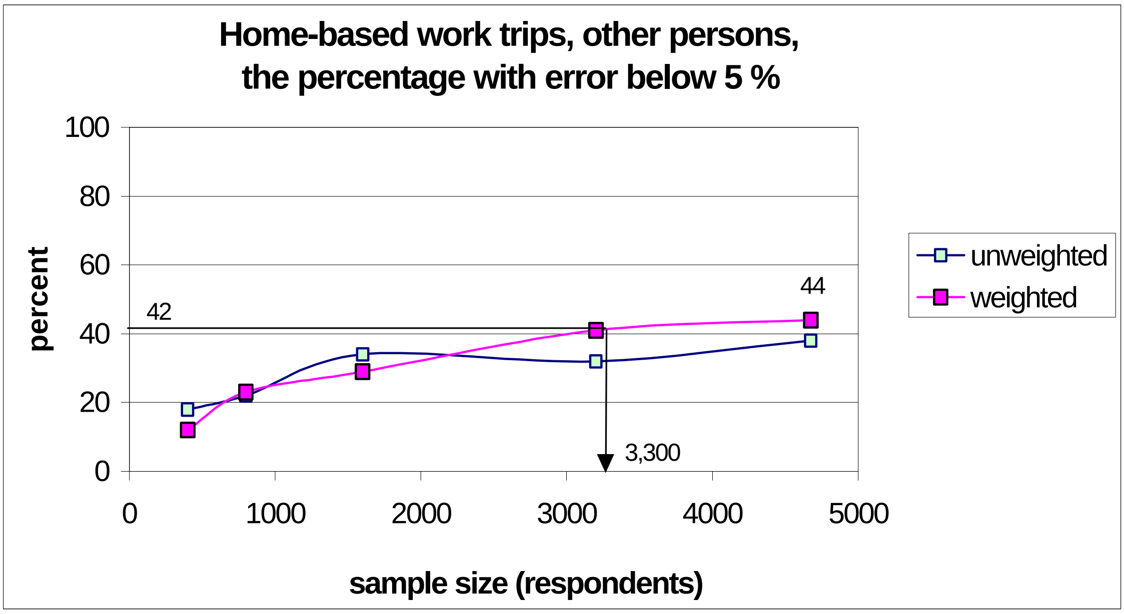


Figure 18: The percentage of trip generation rates with the error below 5 percent as a function of sample size; home-based work trips, other persons.

The figures show that the sample size required to accurately cross-tabulate trip generation rates for home-based work trips is 3,300 respondents with this cross-classification strategy. However, because the trip generation rate of "other" is so small, the absolute effect of this class is marginal (e.g. the weighted amount of trips for 18-64 aged employed is 124,594, whereas it is 7,425 for others). Decreasing the sample size does not result in a significantly worse total generation, even when the sample size of 1,900 respondents is used.

The weighted and unweighted values do not differ from each other remarkably. The weight factor does not actually always improve the model. That is, sometimes the trip generation rate based on the unweighted data is closer to the "right" trip generation rate (based on the weighted data) than the small sample trip generation rate based on the weighted data. This is because errors in the estimations of the two additional parameters are not always comparable with the problem caused by the differences.

Figures 19 to 22 illustrate the variation of trip generation rates for other home-based trips.

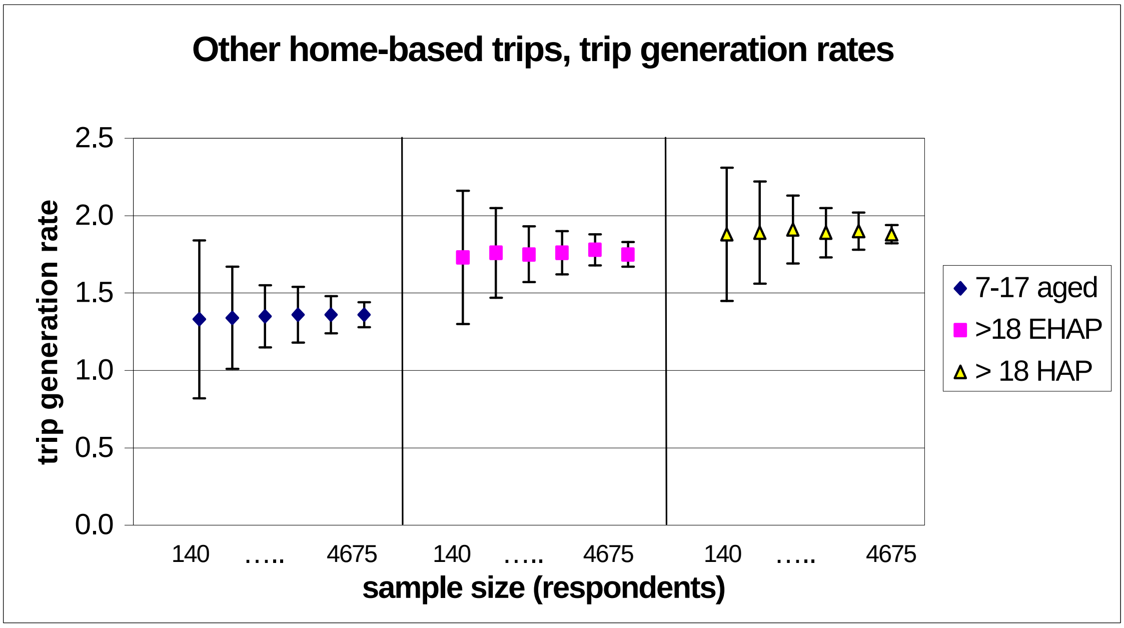


Figure 19: *The variation of trip generation rates and 95 % confidence intervals for other home based trips as a function of sample size based on 100 bootstrap samples and weighted data.*

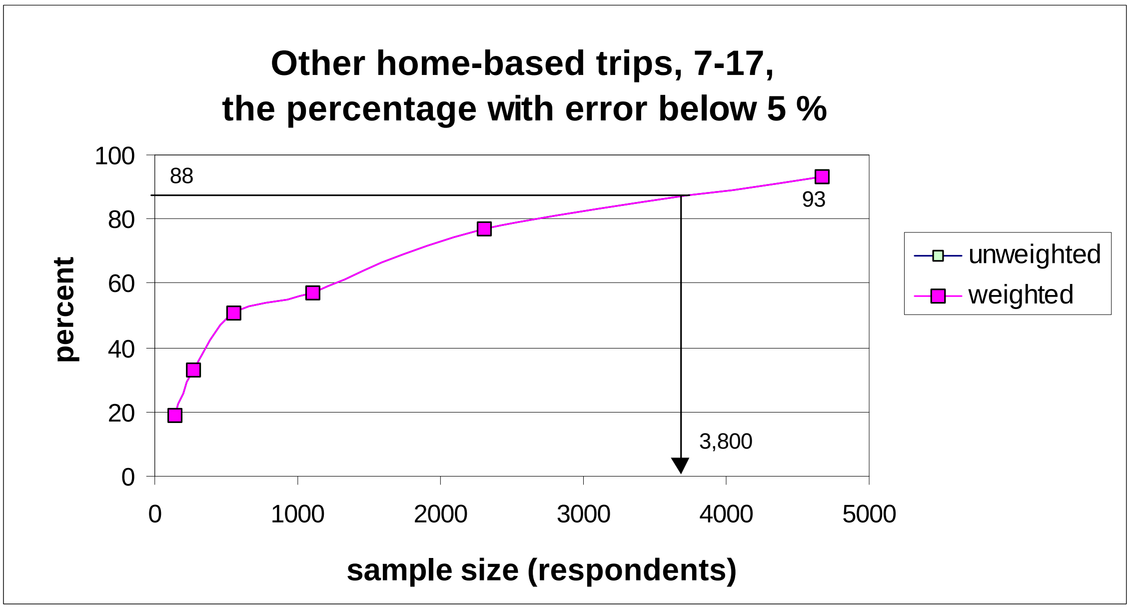


Figure 20: *The percentage of trip generation rates with the error below 5 percent as a function of sample size; other home-based trips, 7-17 aged persons (the unweighted curve is behind the weighted curve).*

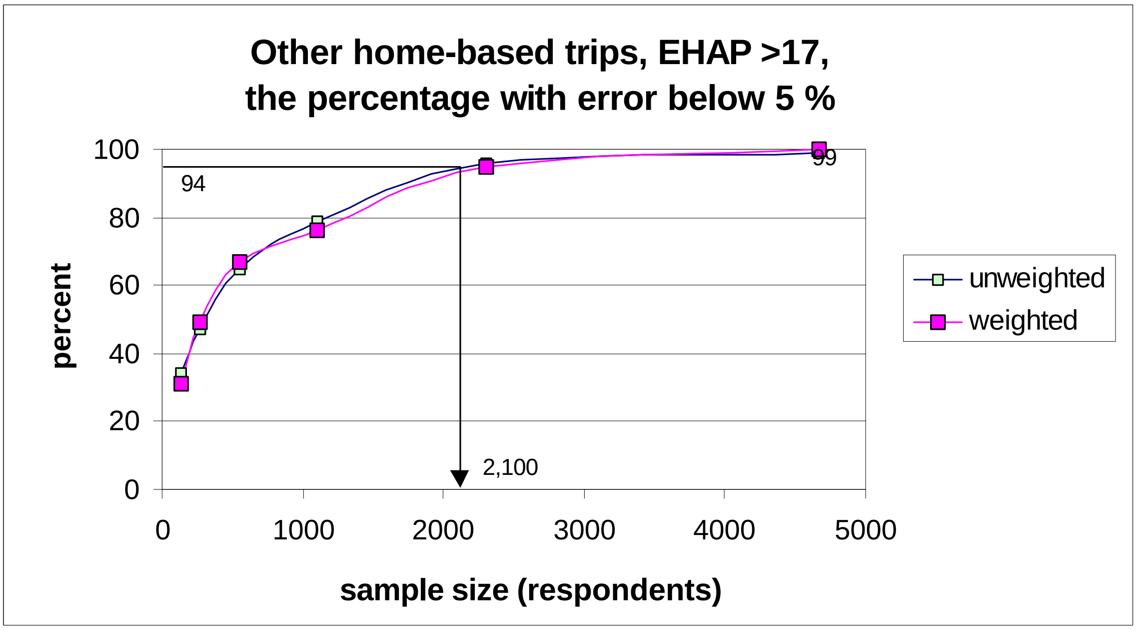


Figure 21: *The percentage of trip generation rates with error below 5 percent as a function of sample size; other home-based trips, EHAP-persons over 17 years of age.*

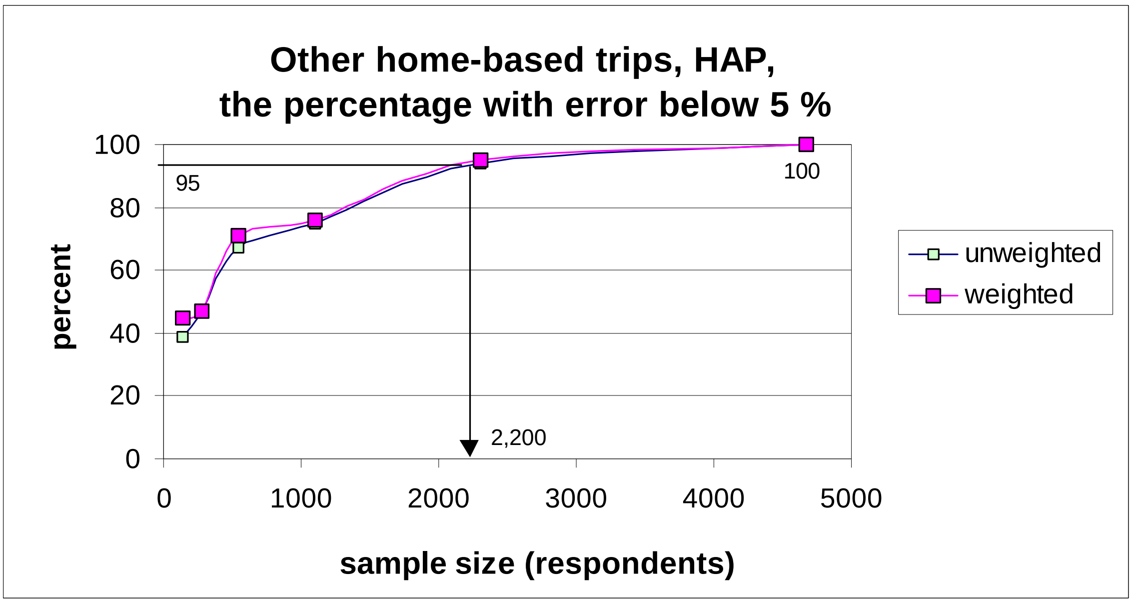


Figure 22: *The percentage of trip generation rates with error below 5 percent as a function of sample size; other home-based trips, HAP-persons.*

As for home-based work trips, the trip generation rates are quite precise even though the variation for the smallest sample sizes is quite large. Also, in this case the weighted and unweighted values do not differ from each other remarkably. The most restrictive group is “persons 7-17" requiring the sample size of 3,800 respondents to achieve an error below the given criteria; however smaller sample sizes still produce reasonable results.

Mode shares

Next the sample size, which is required to accurately get the mode shares by cross-tabulating from the data, is examined. Figures 23 to 26 present the variations in mode shares according to the sample size, and the sample size required for each (walk and bicycle, ptr, car) mode share. The sample size requirement is defined by the criteria that the maximum allowable deviation from the correct mode share is 1 percent; thus, e.g. if the correct mode share is 25 percent, the values between 24 and 26 are allowed. In other words; first, the number of “full” samples, based on the bootstrap samples, not exceeding a 1 percent deviation, is defined for the whole set of data. Then, a number, 5 percent less, is defined for the samples, e.g. if the number of “full” samples which does not exceed a 1 percent deviation is 80 percent, the number required for the samples not exceeding the 1 percent deviation is 76 (0,95*80).

Figures 23 to 26 show that when considering the smallest sample size, the mode shares are very imprecise, and even for the largest sample size (representing the whole data set), only 58 to 76 percent of observations for home-based work trips and 78 to 95 percent for other home-based trips go below the 1 percent error limit. This 1 percent absolute error limit, is actually, stricter than the 5 percent relative limit used in relation to trip generation rates. However, in real life the 1 percent deviation is regarded to be as the maximum error to be allowed in a major study.

In Turku, the results for public transport shares appear to be more precise than those for car and walk and bicycle. This is because the absolute deviation favours the less represented mode share by allowing a larger relative error for this mode. The relative errors are quire similar for all modes.

When allowing for 5 percent poorer result than obtained by using the maximum sample size, the sample size required for home-based work trips is 4,400 respondents and for other home-based trips 3,900 respondents. This means that 4,400 respondents, i.e. all the respondents, are required to get the preciseness required in the Turku region. The result also means that more observations are required to define the mode shares than is the case for the trip generation rates. The best possible result is also worse on a relative basis when defining mode shares than when defining trip generation rates.

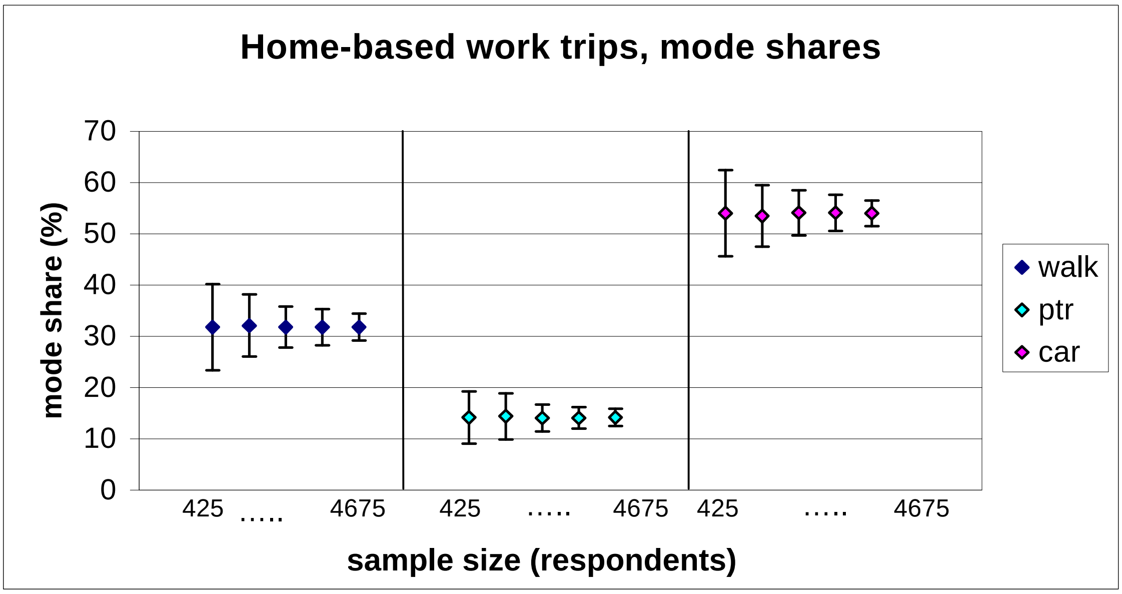


Figure 23: *The variation of mode shares and 95 % confidence intervals for home-based work trips as a function of sample size based on 100 bootstrap samples and weighted data.*

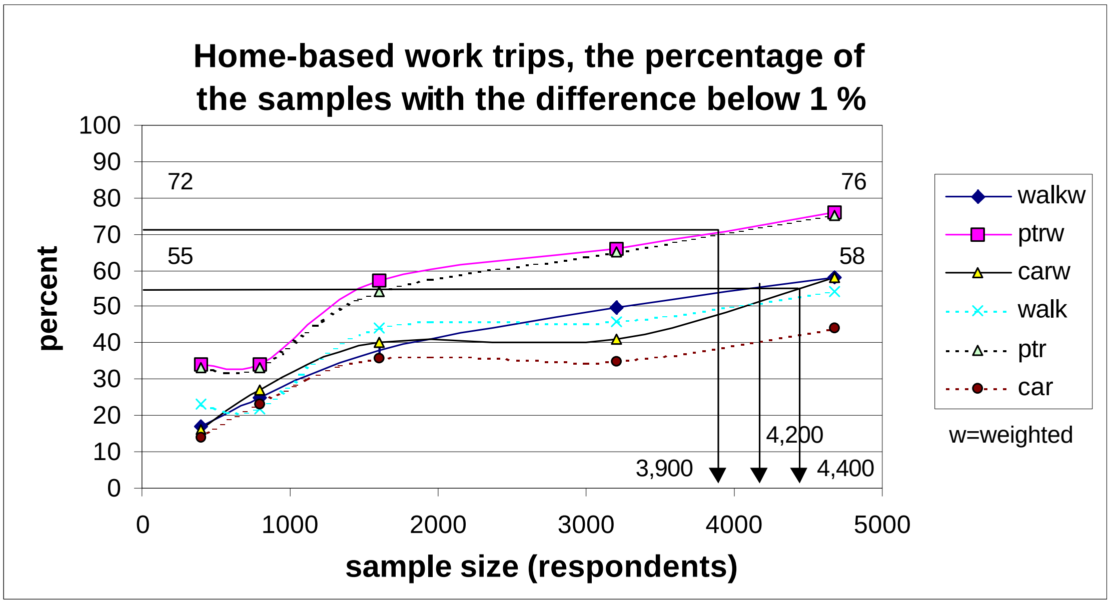


Figure 24: *The percentage of samples with mode share difference below 1 percent as a function of sample size; home-based work-trips.*

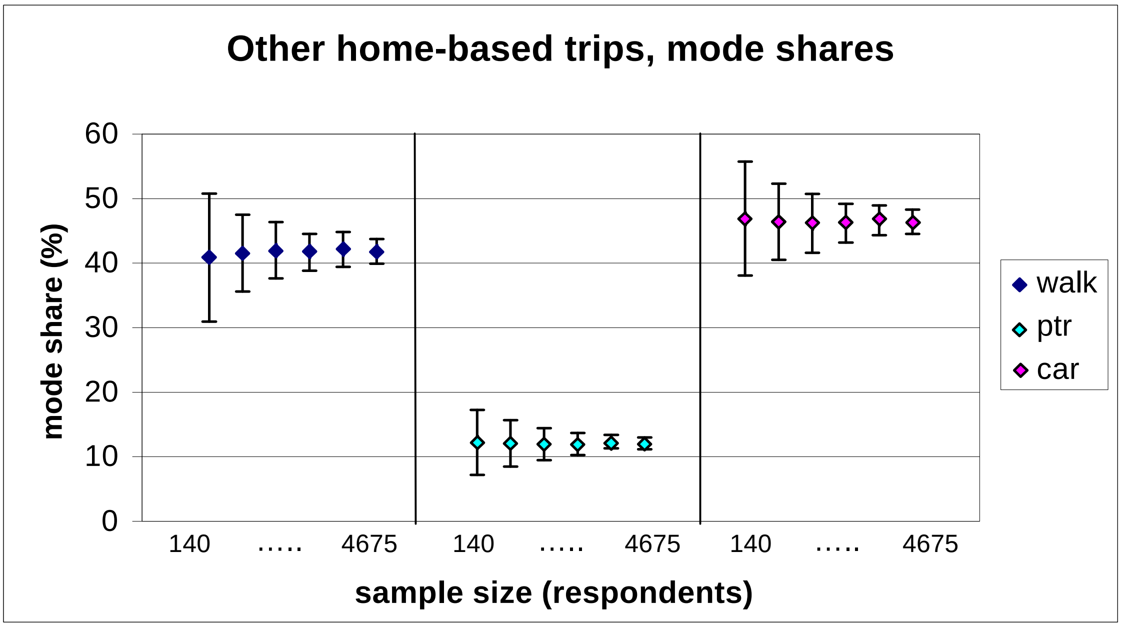


Figure 25: *The variation of mode shares and 95 % confidence intervals for other home-based trips as a function of sample size based on 100 bootstrap samples and weighted data.*

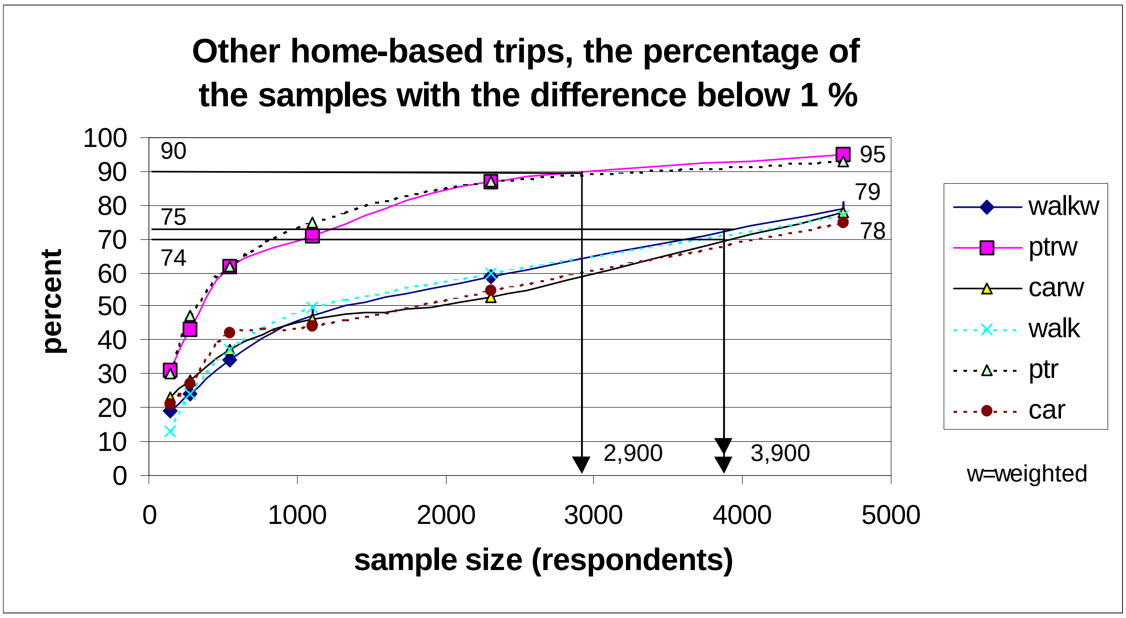


Figure 26: *The percentage of models with mode share difference below 1 percent as a function of sample size; other home-based trips.*

5.3.3 The sample size required in mode- and destination choice model estimation

5.3.3.1 Introduction

In this section the sample size required to estimate mode- and destination choice models is considered. It is well known that in principle sampling errors are roughly inversely proportional to the square root of the size of the estimation data set. Thus, for example, the data-set size must be quadrupled in order to halve the sampling errors *(Horowitz 1981)*. Based on this premise, *Koppelman and Chu (1983)* have formulated analytic relationships for sample size and the precision of parameter estimates. Their methodology estimates the sample size N_s required to obtain parameter estimates with a prescribed percent confidence (1-α) within a given percent z of true values, by the mathematical relationship:

$$N_s = (S * \frac{t_{\alpha/2}}{z})^2, \qquad (48)$$

where S is the standard error of the parameter estimate, and z is expressed as the accepted mean deviation of the true parameter value. This equation performs well in ideal situations in which only one observation per person is used. However, in the Finnish travel surveys, questions are asked for all the trips made for one or two days. This means that there will be correlation between the answers provided by the same person. This is known as the repeated measurement problem. Repeated measurements violate the assumption that the error terms are independent and means that we can no longer rely on the variance estimates obtained in this way *(Cirillo et al. 2000).* Thus, the sample size requirements have to be studied in another way.

This chapter considers the sample size requirements for mode- and destination choice modelling. In addition the repeated measurement problem (described above) is discussed. The samples are drawn by using bootstrap and the sample size is always presented in persons if it is not indicated otherwise. The need for observations is considered from two different aspects which are also used in the model transfer experiment in Chapter 6. The test measures of these two aspects are VOT (value of time), RSEEF (relative sample enumeration error for forecast).

The investigation is divided into the three parts:

- In the first part of the subsection the model parameters' preciseness is considered at quite a general level. That is, the model coefficients based on small samples are compared to the coefficients from the entire set of data. The test measure is the percentual Mean Deviation (Equation 36) from the coefficients estimated using the entire set of data.

- Next the variation of coefficients is studied using both estimated and observed standard deviations (Equations 49 to 50 presented in Section 5.3.3.3). The repeated measurement issue, in particular, is discussed in this subsection.

- In the third part of the study, different criteria for defining the sample size requirements are considered. Two different test measures (VOT, RSEEF) used in evaluating the model transfer effectiveness are discussed in this chapter (Equations 38 and 45). TTS is not considered due to its tendency allways reject the hypothesis of the model parameters equality.

5.3.3.2 Results based on the percentual MD and confidence intervals

Table 18 contains mode choice models for home-based work trips and other home-based trips estimated from the entire set of data in the Turku region. The value of time estimated for the home-based work trips is 3.60 euro/h (21.4 FIM/h) and for the other home-based trips 1.15 euro/h (6.7 FIM/h).

Table 18: Mode and destination choice models for home-based work trips (HBW) and other home-based trips (OHB) using the "reference data" (two-day diaries, 1 reminder) collected in the Turku region in 1997.

Variable	HBW			OHB		
	Coefficient	Std.	t-value	Coefficient	Std.	t-value
Distance 0-5 km (Walk)	-0.8775	0.0363	-24.2	-1.0380	0.0212	-49.0
Distance > 5 km (Kv)	-0.3388	0.0308	-11.0	-0.2470	0.0183	-13.5
Total travel time (Car, Ptr)	-0.0222	0.0027	-8.2	-0.0135	0.0019	-7.3
Number of transfers (Ptr)	-	-	-	-0.4372	0.0415	-10.5
Trip cost/income (Car, Ptr)	-0.8839	0.1220	-7.2	-1.7390	0.0853	-20.4
Cars/household (Car)	0.6578	0.1720	3.8	1.1730	0.0881	13.3
Walk dummy (Walk)	2.1370	0.1880	11.4	2.1570	0.1060	20.4
Car dummy (Car)	-0.5061	0.1950	-2.6	-0.8932	0.1150	-7.7
Log sum (Dest.)	0.7335	0.0170	43.2	1.2020	0.0089	134.8
Scale factor (Dest.)	1.0000	0.0000	-	1.0000	0.0000	-
Number of trip observations	4,442			13,989		
$\rho^2(c)$	0.2011 (mode), 0.1431 (dest)			0.2265 (mode), 0.2806 (dest)		
Walk = Walk and bicycle Ptr = Public transport Car = Car	Scale factor: - ln (jobs) 1.00			Scale factor: - inhabitants 1.00 - retail employment 14.89 - service employment 2.52		

Figures 27 and 28 present the variation of travel time coefficients by sample size for the two-day data. The variation of 100 model coefficients per each sample size is compared with the 95 percent confidence interval based on the average value of observed standard deviations of coefficients through bootstrap. The 95 percent confidence interval is marked by lines in the figures. From the figures we can see that the coefficients estimated by using small samples are distributed around the coefficient, which have been estimated by using the entire set of data. However, the variation of model parameters increases greatly when the sample size decreases and in some cases the signs of coefficients are actually wrong. The results also show that although the sample size is an important factor in model preciseness, there is a limit beyond which the results show only a small level of improvement.

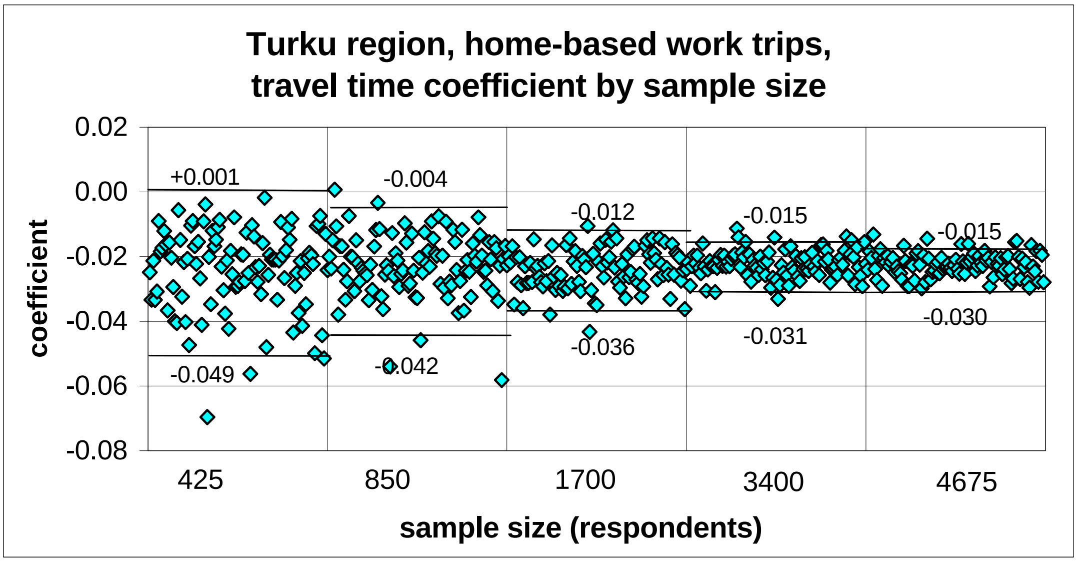


Figure 27: *The variation of travel time coefficients and 95 % confidence intervals of home-based work trips estimated from the two-day data in the Turku region.*

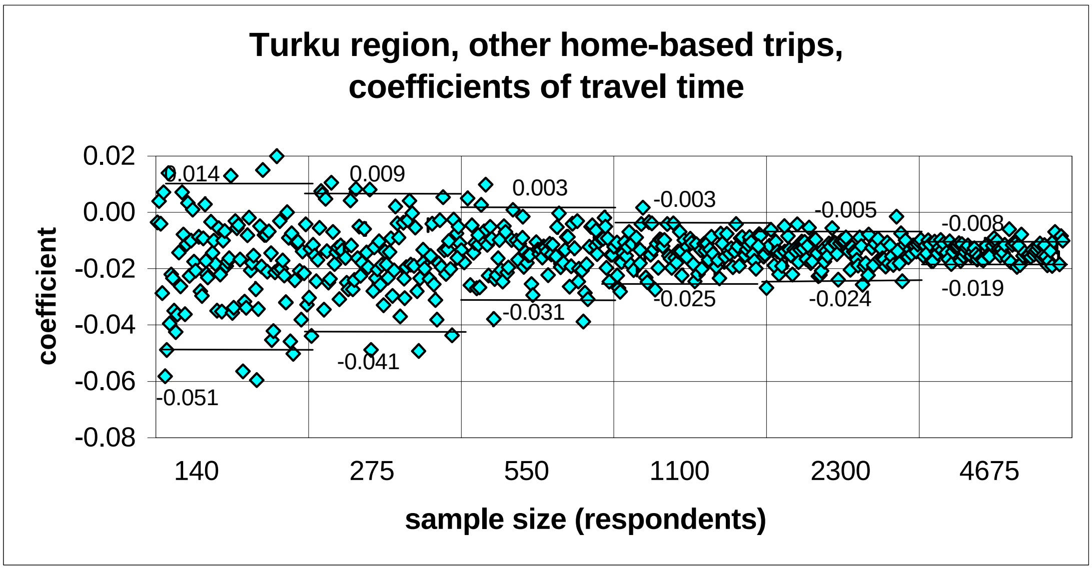


Figure 28: *The variation of travel time coefficients and 95 % confidence intervals of other home-based trips estimated from the two-day data in the Turku region.*

Table 19 presents the percentual mean deviation values for all coefficients. However, note that the results relating to the home-based work trips and other home-based trips are not to be compared directly according to the sample size due to the different number of trips per person in these trip groups. That is, the same sample size of persons represents a different number of trips depending on the trip purpose. In addition, because two-day diaries were used, about 43 percent of the home-based trips but only 11 percent of the other home-based trips made in the first day were also made in the second day. The more observations in common (in the two days), the less precise the coefficients are.

Table 19: The percentual mean deviations of the mode and destination choice model coeffients based on small samples versus the coefficients based on the entire set of original two-day data.

The percentual mean deviation of the coefficient (MD)									
Turku region									
trip group	sample size (persons)	sample size (trips)	travel time	cars/hh	dist >5	cost/inc	dist 0-5	no. of transfers	logsum
HBW	425	400	45.1	129.4	44.3	59.7	19.5	-	23.2
	850	800	32.3	101.5	30.7	36.5	12.6	-	18.2
	1,700	1,600	23.5	54.3	19.0	26.9	9.3	-	11.7
	3,400	3,200	14.6	38.5	13.1	17.5	5.5	-	9.0
	4,675	4,675	13.0	38.1	9.7	14.8	5.1	-	8.0
OHB	140	400	100.5	76.8	127.5	32.9	15.5	80.3	17.7
	275	800	73.0	42.0	53.2	24.7	11.1	54.6	15.9
	550	1,600	48.9	34.3	28.7	18.5	7.5	35.0	11.6
	1,100	3,200	32.9	25.2	27.4	13.2	6.0	24.5	8.4
	2,300	6,400	27.4	16.0	16.2	9.3	4.1	15.3	3.8
	4,675	13,900	17.3	11.8	10.4	8.8	2.2	12.0	3.7

HBW = home-based work trips
OHB = other home-based trips

The most precise mode choice coefficient is “distance 0-5" and the most imprecise coefficients are coefficients “cars per household”, “travel time” and “number of transfers”. On the whole, the use of “number of transfers” variable decreases the accuracy of the models. In particular, the variance of travel time coefficient increases. This is mainly related to the assignment package and also to the network coding. The variable “cars per household” is problematic because it is not based on the individual values but the zonal values by using a division of 23 zones. The same problem was detected using the ”cost per income” variable due to the fact that the incomes were calculated by using zonal values based on 23 zones. The coefficient “distance >5" of other home-based trips was also quite imprecise because of the small number of trips which in this category.

Amemiya (1978) has shown that when the sequential estimation is applied using standard multinomial logit estimation programs, the variance-covariance matrices of the estimates printed for the marginal probabilities of higher-level models (in that case the destination choice level) are incorrect and too small. That was also the case in this study. The coefficients for destination choice models (not presented here) are very precise. However, the sequential estimation, which was used and the lack of alternative specific constants in destination choice models may cause the absolute values of upper level coefficients to be overestimated.

5.3.3.3 Repeated measurement issue

Next the model parameters’ variation and repeated measurement issue are considered with respect to two different kinds of standard deviations. The std_t is the mean value of estimated standard errors based on the results derived from the Alogit estimation program *(Hague Consulting Group 1992)* and std_p is the observed standard deviation calculated from the distribution of 100 model coefficients through the bootstrap procedure.

$$\mathrm{std}_t = \frac{\sum_{i=1}^{N} \mathrm{std}_i}{N}, \text{ where} \tag{49}$$

std_i = the estimated std-value of the coefficient based on the Alogit

$$\mathrm{std}_p = \sqrt{\frac{\sum_{i=1}^{N} (\hat{\beta}_i - \bar{\beta})^2}{N-1}}, \text{ where} \tag{50}$$

N = the number of sample models

$\bar{\beta}$ = mean value of coefficients estimated from sample models

Theoretically the std_t should be close to std_p, if any bias does not exist. However, in cases repeated observations are taken from an individual, this is not the situation.

Table 20 shows that in our case there is a large difference between the estimated and observed standard deviation. The ratio of the std_p and std_t is approximately 1.5 regardless of the sample size. It is likely that large part of the difference is due to the repeated measurements. The repeated measurement error is greater in other home-based trips than in home-based work trips. This is due to the larger number of trips per person in the other-home-based trips, although there are fewer observations in common between the first and second day when considering the other home-based trips.

Table 20: The comparison of estimated (std_t) and observed (std_p) standard deviations based on the two-day data collected in the Turku region.

std_t versus std_p														
Turku region														
trip group	sample size	travel time		cars/hh		dist >5		cost/inc		dist 0-5		no. of trans-fers	
		std_t	std_p	std_t	std_p	std_t	std_p	std_t	std_p	std_t	std_p	std_t	std_p
HBW	425	0.010	0.013	0.61	1.08	0.13	0.18	0.43	0.65	0.13	0.21	-	-
	850	0.007	0.010	0.42	0.86	0.07	0.14	0.30	0.42	0.09	0.11	-	-
	1,700	0.005	0.006	0.29	0.47	0.05	0.08	0.21	0.29	0.06	0.10	-	-
	3,400	0.003	0.004	0.20	0.31	0.04	0.06	0.15	0.18	0.04	0.06	-	-
	4,675	0.003	0.004	0.17	0.33	0.03	0.05	0.12	0.17	0.04	0.06	-	-
OHB	140	0.012	0.017	0.54	1.16	0.18	0.35	0.55	0.78	0.13	0.20	0.27	0.54
	275	0.008	0.013	0.37	0.64	0.09	0.17	0.37	0.51	0.09	0.14	0.18	0.32
	550	0.006	0.009	0.26	0.49	0.06	0.10	0.26	0.40	0.06	0.10	0.12	0.19
	1,100	0.004	0.006	0.18	0.36	0.04	0.09	0.18	0.28	0.04	0.08	0.09	0.11
	2,300	0.003	0.003	0.13	0.17	0.03	0.04	0.12	0.13	0.03	0.04	0.06	0.06
	4,675	0.002	0.003	0.09	0.17	0.02	0.03	0.09	0.13	0.02	0.04	0.04	0.06

HBW = home-based work trips
OHB = other home-based trips

The common observations on the first and the second day as well as the other repeated answers provided by the same individual, do not affect the observed variation but decrease the estimated std_t-values. This means, that:

- estimated standard deviations are underestimated and t-values based on these standard deviations are incorrect because they are too high.

- in model transfer, the coefficients based on the Bayesian approach, or combined transfer estimation, which use the std_t in emphasizing the estimation and application context parameter values, may be biased. This issue is discussed further in Appendix H.

Although the estimated standard deviations are underestimated, the coefficients for new sample models are consistent and asymptotically efficient. However, *Brundell-Freij (1995)* has stated in her study, that for small sample sizes the absolute values of the coefficients are systematically overestimated, and the estimated standard deviations are systematically underestimated.

It must also be noted that the standard deviations of coefficients can never be exactly derived from sample size *(Rao 1973)* because they depend on the samples from which they have been estimated. That was also the reason why the importance of sample size was studied by drawing random bootstrap samples.

5.3.3.4 The sample size requirement based on the ratio of time and cost coefficients (VOT)

This chapter considers the sample size required for mode- and destination choice modelling based on the VOT-tests. Figures 29 and 30 present the value of time according to the sample size. The distribution of VOT is presented in the upper part of the illustration and three different sample size criteria, namely 10, 25 and 50 percent error limits, are described in the lower part of the illustration. From those, the middlemost, 25 percent error limits for the VOT estimates are marked by lines in the upper part of the illustration.

The sample size is considered to be large enough if the result is no more than 5 percent poorer than the result obtained by using the entire set of data. For example, 84 percent of the models yield an error below 25 percent when the entire set of data is used. Thus, the required share of observations, which go below the 25 percent error, is 80 percent (0,95*84); 5 percent less than if the largest sample size is used.

We have chosen this relative error criterion because it is difficult to define any acceptable absolute error limits, which would be relevant in all situations. The acceptable error is greatly dependent on the purpose of the model. In addition, in many cases, the absolute results which can be reached are rather poor even by using the entire set of data. The three error criteria presented are not definitive recommendations, but rather illustrate how the sample size requirements can be evaluated.

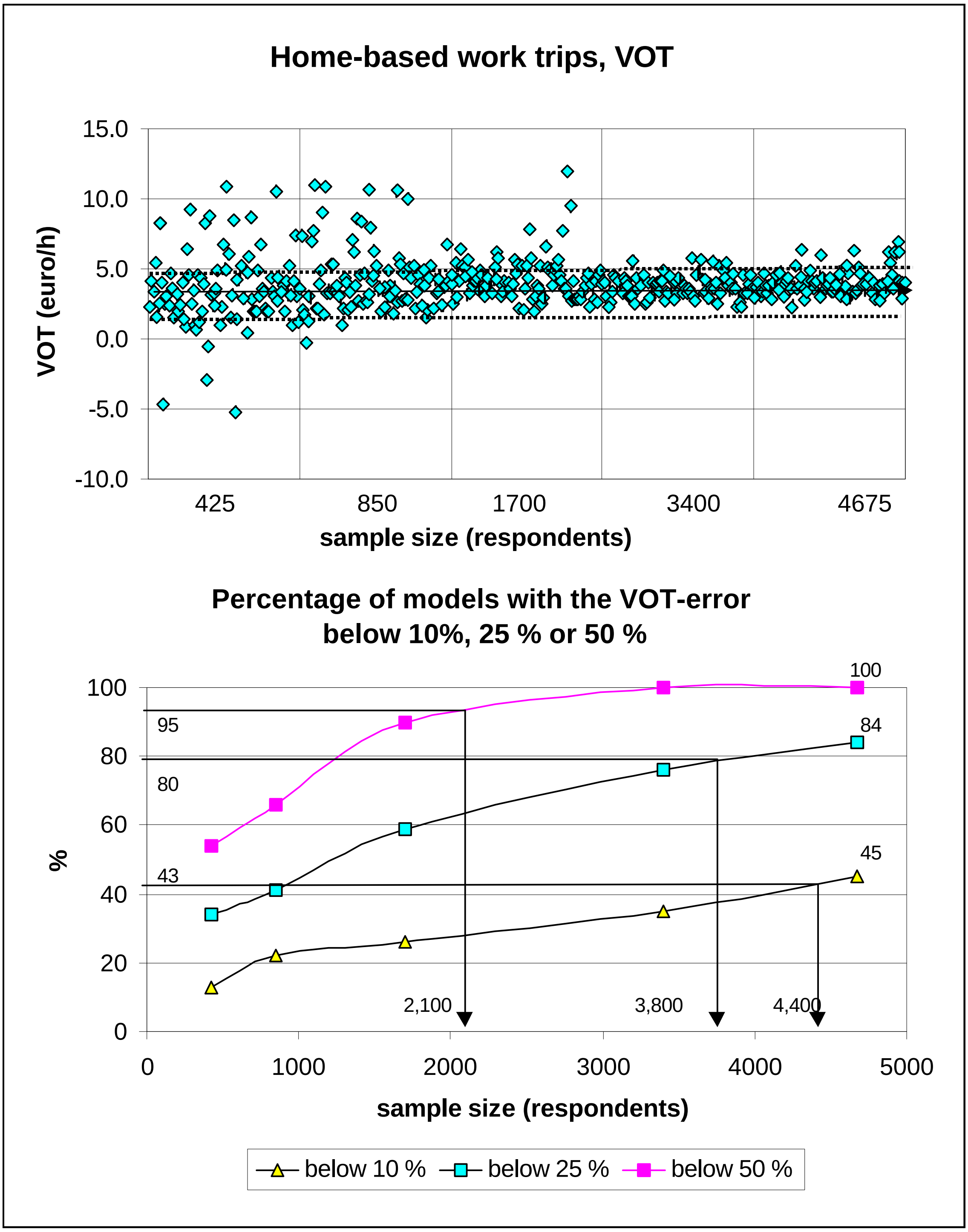


Figure 29: *The sample size required for home-based work trips based on the ratio of the time and cost/income coefficients (VOT criteria, 95 % of the best possible result).*

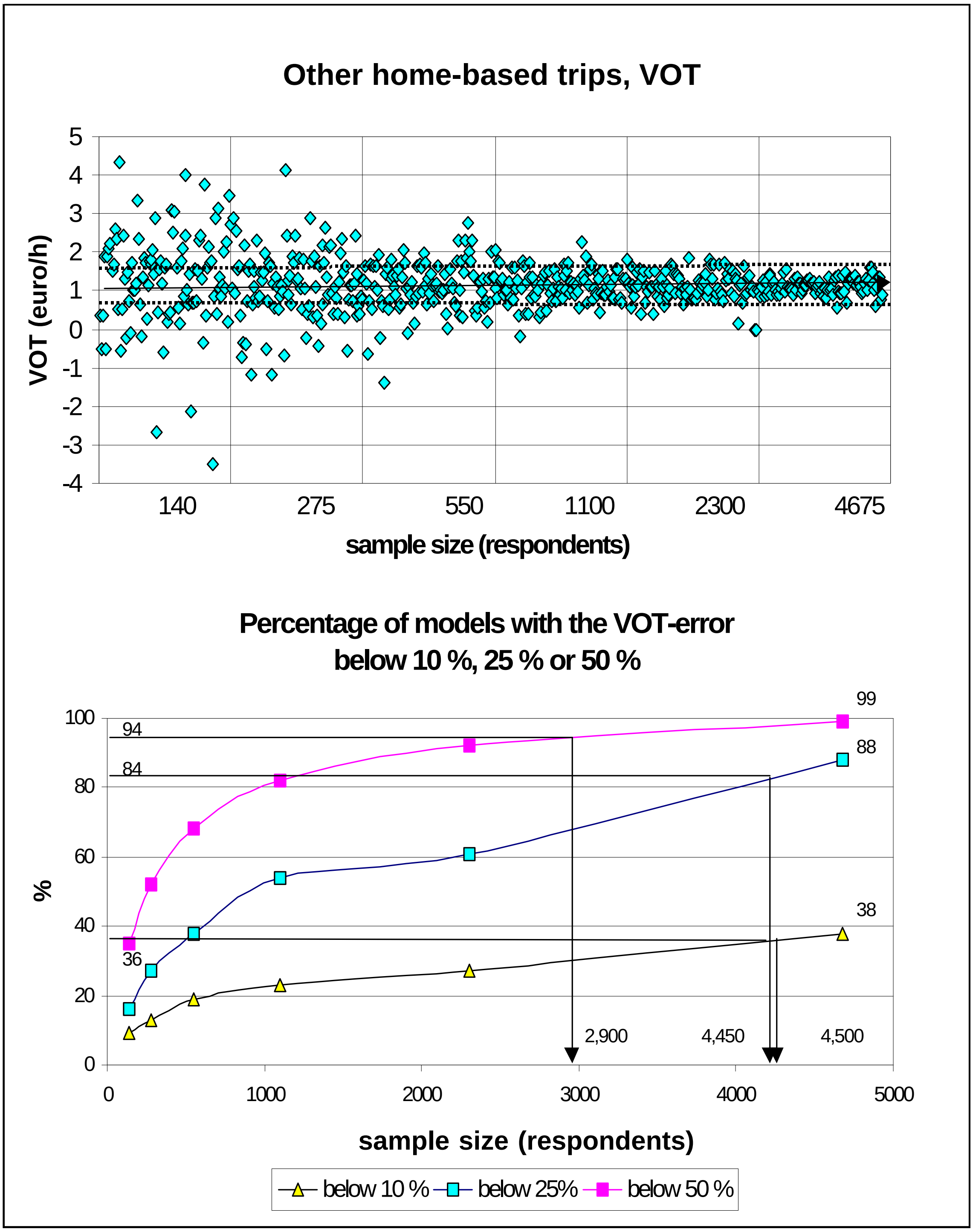


Figure 30: The sample size required for other home-based trips based on the ratio of the time and cost/income coefficients (VOT criteria ,95 % of the best possible result).

Figures 29 and 30 show, that when using 10 percent error criteria, the sample size required for modelling home-based work trips is 4,400 respondents and for modelling other home-based trips 4,500 respondents. The corresponding sample size requirements according to the 25 and 50 percent error-limits are, 3,800 and 2,100 for home-based trips and 4,450 and 2,900 for other home-based trips. Note that these recommendations provide results which are only 5 percent worse than "perfect" model yields. In many cases smaller preciseness may be considered as acceptable. On the other hand, the results, where only 45 percent (HBW) or 38 percent (OHB) of models yield error below 10 percent (although the entire set of data has been used) cannot be considered as acceptable.

5.3.3.5 The sample size requirement based on the elasticity tests

Figures 31 to 34 present how the variation of model parameters affects the model's ability to predict changes in public transport shares or car shares, if the public transport travel time increases 30 percent or car costs increase 10 percent. These tests are carried out by running the whole current forecast process (the whole forecast process based on the current land use and impedances) and including the iteration process of the alternative-specific constants, so that the mode shares represent the weighted mode shares in the area. The sample size requirement has been examined for three different error- criteria, that is the percentage of models with the RSEEF error under 10, 25 or 50 percent has been defined. The sample size is considered to be large enough if the result is no more than 5 percent poorer than the result obtained by using the largest sample size. From those, the middlemost, 25 percent error limits are marked by black dash lines in the upper part of the figure.

Figures 31 to 34 show that the sample size required according to the elasticity tests varied from 1,300 to 4,300 respondents in home-based work trips and from 300 to 4,200 respondents in other home-based trips. The 10 percent criteria was mostly in a dominating position. However, in some cases, the strict error-criteria may lead to a smaller sample size recommendation than when using broad criteria. This is because the comparison is not based on the absolute values, but is produced with respect to the best possible result for each criterion. Thus, if the curve is flat the sample size requirement may be higher for 50 percent criteria than for 25 percent criteria. However, normally the strict error-limit is in the critical path because it is more sensitive to the decrease in the sample size.

The results were mainly rather good. However, the elasticity test for home-based work trips shows that the models ability to predict changes in car shares affected by a 10 percent increase in car costs was quite poor even when using the entire set of data. On the other hand, note that the absolute errors are smaller than the relative errors, which are quite high due to the small real effect in this case. The models for other home-based trips are more precise than the models for home-based work trips.

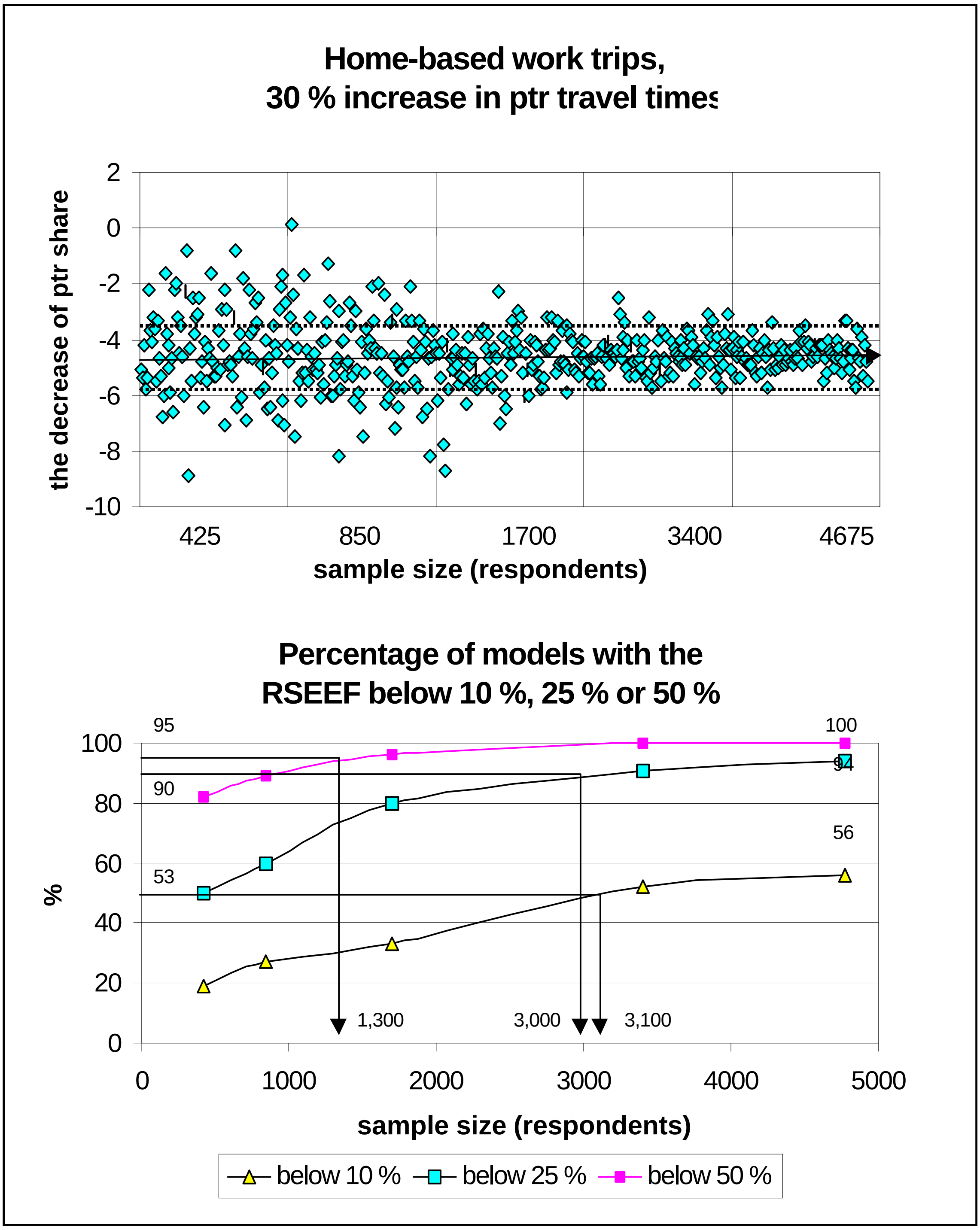


Figure 31: *The variation of change in public transport (ptr) shares, when public transport travel time is increased by 30 percent in home-based work trips.*

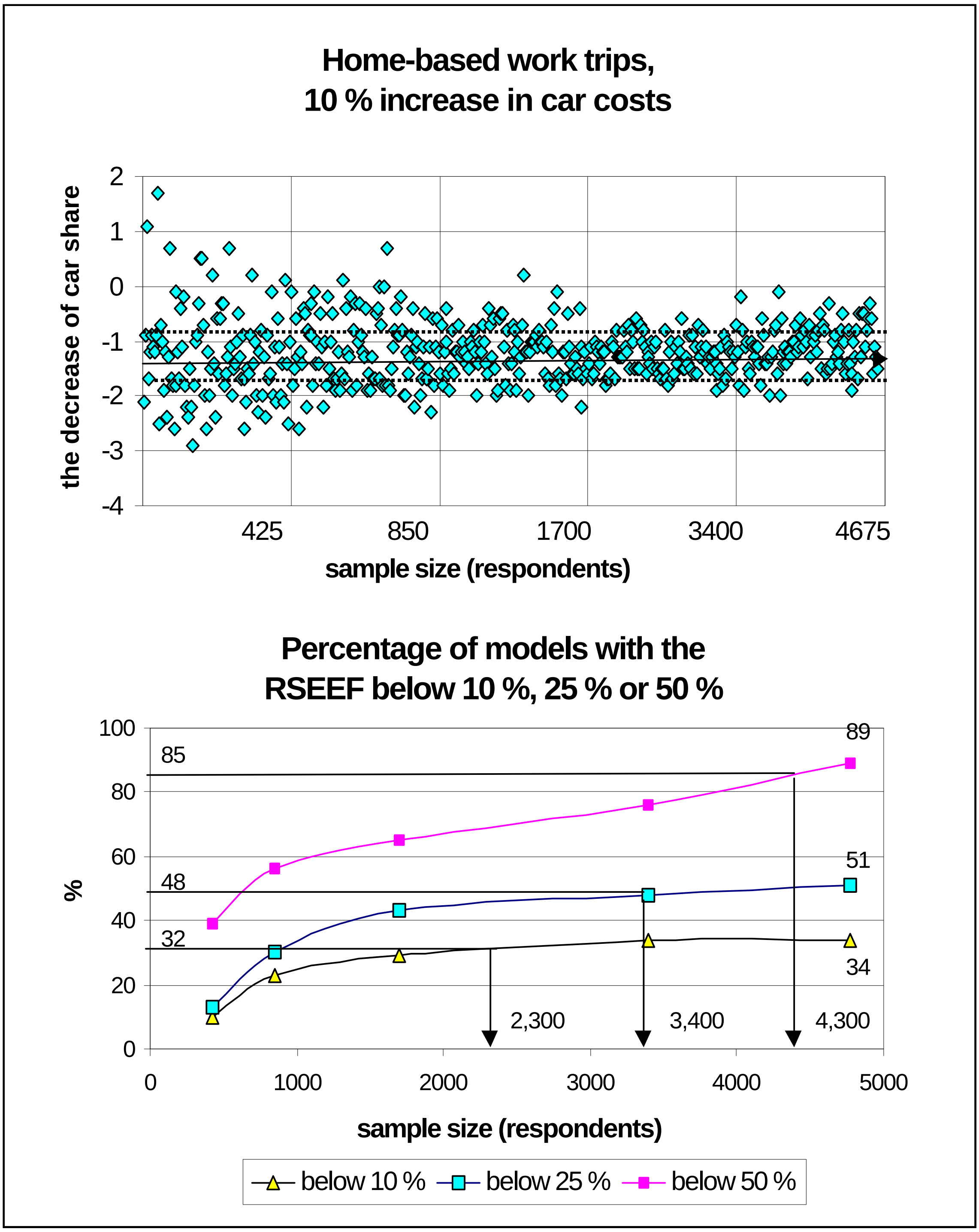


Figure 32: The variation of change in car shares based on the new sample models, when car costs are increased by 10 percent in home-based work trips.

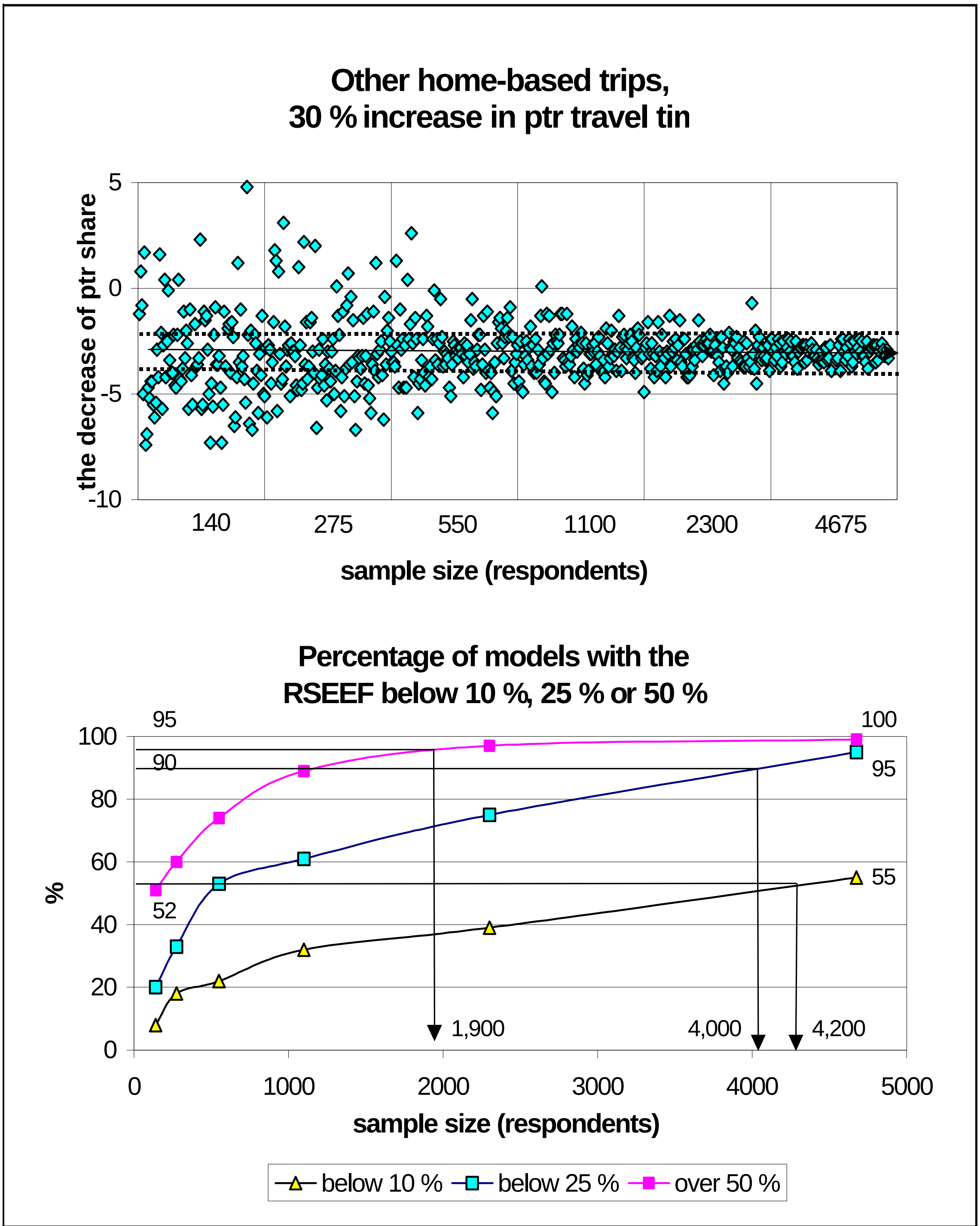


Figure 33: *The variation of change in public transport (ptr) shares based on the new sample models, when public transport travel time is increased by 30 percent in other home-based trips.*

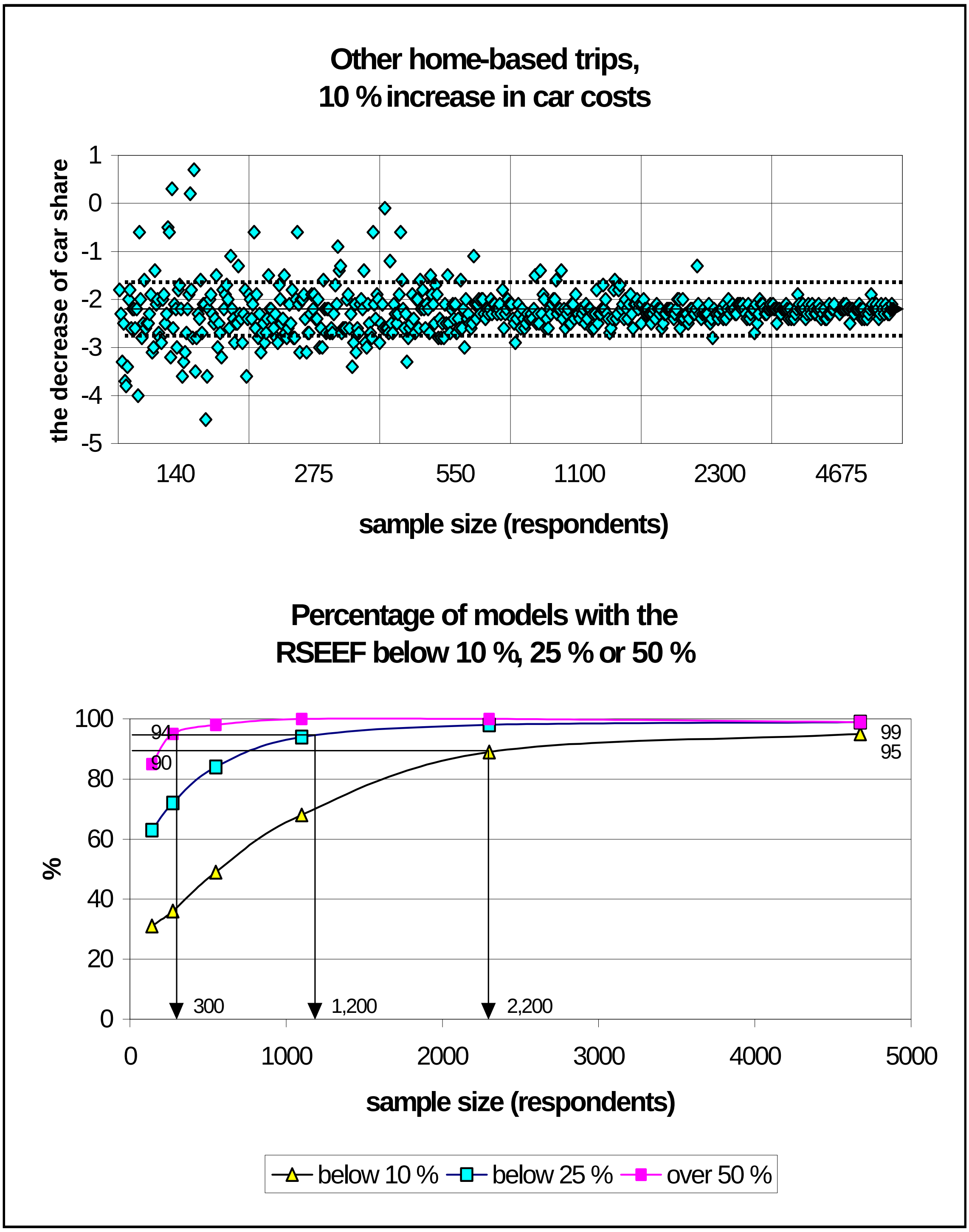


Figure 34: *The variation of change in car shares, when car costs are increased by 10 percent in other home-based trips.*

5.3.3.6 Summary of sample size requirements

The sample size requirements, based on the criteria studied, are presented in Table 21. The most restrictive (critical) values are indicated in the table in bold type. Table 19 and Figures 31 to 34 show that the travel time coefficient is more precise than the cost coefficient, when based on the treatment of the home-based work trips. Further, the cost coefficient is more precise than the travel time coefficient, when based on the treatment of the other home-based trips (even when the sample size requirement is, in both cases, larger for the elasticities based on the travel time effect).

Table 21: The sample size required to estimate new models based on the four different criteria.

		The sample size required					
criteria	trip group	below 10 %		below 25 %		below 50 %	
		respon-dents	corresp. test value	respon-dents	corresp. test value	respon-dents	corresp. test value
VOT	HBW	**4,400**	**45**	3,800	84	2,100	100
	OHB	**4,500**	**36**	4,450	84	2,900	94
RSEEF ptr time	HBW	**3,100**	**53**	3,000	90	1,300	95
	OHB	**4,200**	**52**	4,000	90	1,900	95
RSEEF car costs	HBW	2,300	32	3,400	48	**4,300**	**85**
	OHB	**2,200**	**90**	1,200	94	300	94

HBW= home-based work trips
OHB= other home-based trips

The minimum adequate sample size obtained from the VOT criteria (which is the most restrictive criterium) for home-based work trips is 4,400 respondents. The sample size is 4,500 respondents for other home-based trips. The corresponding values, based on the RSEEF criteria, are 4,200 and 4,300, respectively. The adequate sample sizes, based on a 25 percent error limit, which are later used in the transferability study (Chapter 6) are 3,800 for home-based work trips, and 4,450 respondents for other home-based trips.

The sample size requirements, based on the first day of the two-day travel diary, were also examined. The sample size required for modelling, would be 4,600 respondents when only the first day diaries, or correspondingly the one-day diary were used. It means that the one-day data responds in reducing the sample size equally with the two-day data. However, if the absolute level of errors is to be considered, the two-day data yielded clearly better results than the first day diaries alone.

For those coefficients where the elasticity tests were not carried out (e.g. cars/hh), the sensitivity to elasticity error can be roughly calculated from Table 19, based on the known relationship between the RSEEF and the MD for the travel time and costs. For example, when comparing the average RSEEF (presented in appendixes I and J) to the MD-values presented in Table 19, it can be noted that the 25 percent mean deviation error of time or cost per income coefficient affects the average RSEEF by less than 20 percent.

5.3.3.7 The CV-values

Finally, we have tried to find a unit of measurement by which the different model parameters' preciseness can be evaluated and compared. This unit of the measurement is the CV-value which is defined as a ratio of observed standard deviation of coefficient (std_p) divided by the coefficient estimated from the entire set of data (CV(B)), and it is used to systematically evaluate the preciseness of model parameters. Actually, the CV-value is inverse of t-value and it describes the stability of the model parameters only implicitely, as it is a measure of the risk of having an incorrect coefficient. The advantage of using CV-values is that they can be easily used in model transfer in order to evaluate in a variable-oriented way how the coefficients of estimation and application context should be emphasized. However, in model transfer, estimated standard deviations for application context parameters (std_t) must be used as the numerator and the coefficients estimated from the entire set of original estimation context data as the denominator (CV(A)) (see Chapter 6).

The CV-values calculated for coefficients estimated from the data collected in the Turku region are presented in Figure 35. The aim of the analysis is to compare the preciseness of different coefficients and evaluate the behaviour of the CV(B)-values in order to find out, if the CV(B) criteria can be used to define the minimum sample size for estimation of new sample models.

Figure 35 presents the mean CV(B)-values for all coefficients calculated as an average value of 100 random samples in each sample size category. The maximum CV(B)-value which is allowed for estimating new models is marked by an arrow. This maximum value is defined according to the sample size requirement based on the 25 percent error-limit. Hence, in principle, the CV(B)-value of the coefficient, which was most restrictive when using the VOT, or elasticity tests, was chosen as the acceptable CV(B)-value.

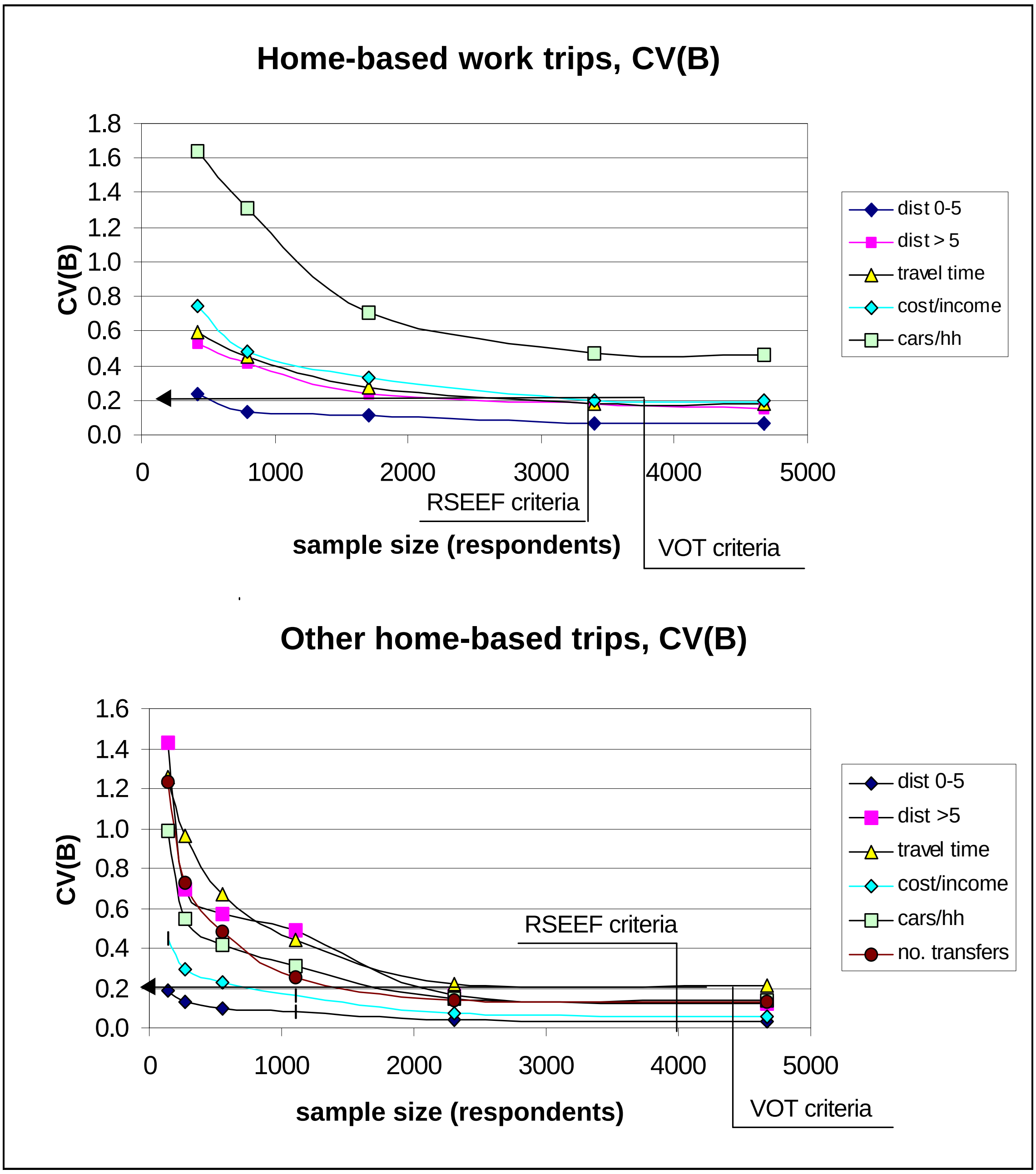


Figure 35: *The mean CV(B)-values of sample model coefficients, for home-based work trips and other home-based trips.*

Figure 35 shows, that it is not easy to determine the CV-value unambiguously because many different criteria can be used. If the CV-value is defined by the sample size determined from both the VOT-test and the RSEEF- (elasticity) tests, the VOT-test will generally be the more restrictive condition. The difficulty with this measure is that the VOT is more sensitive to the sample size than the CV value defined for the separate coefficients. Accordingly, the CV-value is nearly equal for samples over 2,500 respondents. However, the precise estimation of VOT requires a sample size of over 4,400 or 4,500 respondents. The next question is, which coefficients are used to predict changes in traffic. In this case only the effects of the 30 percent increase in public transport travel time and a 10 percent increase in car costs were evaluated. Had the sample size been based on the behaviour of some other coefficients, e.g. car ownership (cars per household), it would have been larger.

Based on the chosen 25 percent criteria, the maximum CV-value allowed in home-based trips is 0.20. In other home-based trips the maximum value is 0.21. One would expect, that the new sample models can be estimated precisely if the CV-values for all model coefficients are below this critical CV-value. This is the adequate condition required to calculate elasticities precisely; however, the estimation of VOT may require even larger sample sizes. In addition, note that CV-values are dependent on the model structure and model definitions, so they must always be defined separately for every model type used.

5.3.3.8 Discussion

The size of the required sample depends on the purpose of the model. If the main purpose is to forecast the total demand, the sample size required can be smaller than the sample size needed to precisely estimate one particular coefficient.

The investigation shows that in our case trip generation rates can be accurately cross-tabulated by using a sample size of 3,800 respondents. The sample size required is a bit larger for other home-based trips (3,800) than for home-based work trips (3,300) due to the different segmentation of these trip groups. This is because for other home-based trips three different segments were used instead of the two segments that were used for home-based work trips.

The weighted and unweighted values do not differ from each other remarkably; however, in this case, all the sample sizes are weighted, based on the five age groups. In real situations, for larger samples, it is possible to use more precise weighting procedures and obtain slightly better results.

As defining mode shares precisely proved to be much more difficult than the cross-tabulation of trip generation rates. If mode shares, based on the samples, differed significantly from those based on the whole mobility survey, the weighting of data could not correct the errors.

Because defining the mode shares correctly is one of the main purposes of mobility surveys, it will be important, in the future, to find a method by which mode shares can be determined more precisely. In principle, the process of defining mode shares is independent of mode and destination choice modelling. The mode shares are based on the cross-tabulation of mobility survey data, and are only used in the forecast process to correct alternative-specific constants to accurately present the real weighted mode shares in the area.

The sample size required for mode- and destination choice modelling has been evaluated, based on the VOT and RSEEF measures. The results are evaluated with respect to these measures, which can be obtained by using the entire set of data.

The consequences of exceeding the limits of 10, 20 and 50 percent error have been considered. According to the tests, the adequate sample size by which the models are nearly as good as the models estimated by using the entire set of data, was 4,400 respondents for home-based work trips and 4,500 respondents for other home-based trips. This is more than is used in many studies in Finland. By using only the first day diaries in the Turku region, it was concluded that a sufficient sample size would be 4,600 respondents. This requirement was based on the relative goodness with respect of the results based on the entire set of data. However, the absolute level of preciseness was much better when using two-day diaries than when using first day diaries only. So, it can be seen, that although the two-day diary does not increase the statistical accuracy of the responses (the number of respondents does not increase), it may improve the quality of the mode and destination choice models, assuming that the sample size is not greatly reduced by the use of two-day diary.

The most imprecise coefficients, in our case are “number of transfers” and “cars per household”. The variable “cars per household” and also “cost per income” are problematic because they are zonal values rather than individual values. All the variables produced by Emme/2 program are also problematic, especially the number of transfers. This is due to the program’s tendency to produce too many transfers. Therefore, there are many different factors affecting the impreciseness of the model parameters and almost all the variables have their own weaknesses. The results also show that although the sample size is an important factor in model preciseness, there is a limit beyond which the results show only a small level of improvement.

Finally, we have tried to find a unit of measurement by which the preciseness of the different model parameters can be evaluated and compared. We would also expect this measure to be used in model transfer, as a means of evaluating variable-oriented models, and to determine how the estimation and application context details have to be emphasized. This unit of measurement is the CV-value which is defined as the ratio of standard error of a coefficient divided by coefficient estimated from the whole estimation context data set. The initial premise is, that the greater the CV-values, the more imprecise the model parameters are. The maximum CV-value allowed for the models to be estimated reliably was defined as 0.2 for home-based work trips and 0.21 for other home-based trips.

6 COMPARISON OF TRANSFER METHODS

6.1 The aim and contents

This chapter presents the transfer of mode and destination choice models of internal trips in the HMA to the Turku region. The HMA database is used to estimate the models that are to be transferred. The data collected in the Turku region in 1997 represents the transfer context to which the estimated HMA models are transferred. All four methods described in Section 3.7 are examined. The methods used in the study have been described in Section 4.5.

In principle the results of spatial transferability can methodologically be generalized to temporal transferability, or vice versa. However, spatial transfer is more challenging than temporal transfer, due to the big differences between the two sets of data. Thus, if we can show that spatial transfer is possible in some contexts, we can be certain that a similar approach can be used as part of temporal transfer. The temporal aspect for HMA model transfer is discussed, e.g., in (*Karasmaa 1996a, Karasmaa 1996b, Karasmaa 1997, Karasmaa 1998, Karasmaa 2000*) and the experiments for applying SP-data in model transfer have been reported in *(Karasmaa 1995, Kurri et al. 2001*).

The results are presented in two parts. Section 6.2.1 presents the results based on the home-based work trips. Section 6.2.2 considers the results of other-home-based trips. Some of the main findings are summarized in Section 6.3.

6.2 Results

6.2.1 Home-based work trips

Table 22 contains mode and destination choice models for home-based work trips estimated from the entire set of data. These models are considered to be base (HMA) and target (Turku) models in an estimation and application context. The model specifications are presented in Section 4.5. The value of time estimated from the HMA data was 2.25euro/h (13.4 FIM/h). The value of time estimated from the Turku data was 3.60euro/h (21.4 FIM/h). The value of time is based on the assumption of an average income of 2,144 euro/month/household (12,750 FIM/month/household) in the HMA and 2,405euro/month/household (14,300 FIM/month/household) in the Turku region.

Table 22: Estimation results of home-based work trips using the entire set of data collected in the Helsinki metropolitan area in 1995 and in the Turku region in 1997.

Variable	Helsinki Metropolitan Area			Turku Region		
	Coefficient	Std_t	t-value	Coefficient	Std_t	t-value
Distance 0-5 km (Walk)	-0.9256	0.0615	-15.1	-0.8775	0.0363	-24.2
Distance > 5 km (Walk)	-0.4452	0.0390	-11.4	-0.3388	0.0308	-11.0
Total travel time (Car, Ptr)	-0.0277	0.0031	-9.0	-0.0222	0.0027	-8.2
Trip cost/income (Car, Ptr)	-1.5860	0.1400	-11.3	-0.8839	0.1220	-7.2
Cars/household (Car)	1.1470	0.2760	4.2	0.6578	0.1720	3.8
Walk dummy (Walk)	1.2900	0.2330	5.5	2.1370	0.1880	11.4
Car dummy (Car)	-1.8970	0.2590	-7.3	-0.5061	0.1950	-2.6
Log sum (Dest.)	0.5399	0.0178	30.4	0.7335	0.0170	43.2
Ln(jobs) (Dest.)	1.0000	0.0000	-	1.0000	0.0000	-
Number of trip observations	1,993			4,442		
$\rho^2(c)$	0.2336 (mode) and 0.1407 (dest)			0.2011(mode) and 0.1431 (dest)		
Walk = Walk and bicycle Ptr = Public transport Car = Car						

Table 22 shows that the coefficient for the costs/income differs significantly in the estimation and application contexts. The difference in parameters can be influenced by the differences in incomes between the two areas. The differences of parameters can also be related to the zoning system, and of course, to many other factors related to both the transportation system and the land use structure.

As an example of the comparison of the different transfer methods, Figure 36 presents the variation of travel time and cost coefficients according to the transfer method. The grey arrow marks the coefficient estimated from the entire set of estimation context data, and the black one the coefficient estimated from the entire set of application context data.

Figure 36 shows that the variation is great in new sample models, transfer scaling, joint context estimation and combined transfer estimation approaches. The differences between these methods are, on the whole, quite small. By being highly precise, the Bayesian approach differs greatly from the other methods. In spite of its stability, the Bayesian approach was never able to fully substitute the model estimated from the entire set of application data. This can be explained by the tendency of the method to emphasize the coefficients derived from the original estimation context data. The Bayesian approach emphasizes the coefficients with respect to the inverse of the variances of each coefficient. Since the number of observations is larger leading to smaller variance of coefficient, the coefficients of the estimation context data may be emphasized too strongly in model transfer. If the transfer bias is small, the Bayesian approach yields good results because it efficiently restricts the variation of coefficients. However, if transfer bias is large, as in this case, the Bayesian approach does not perform well. The combined transfer estimation method generalizes the Bayesian method in relation to transfer bias. Hence, we can see the effects of transfer bias by comparing the results from Bayes and the combined transfer estimation.

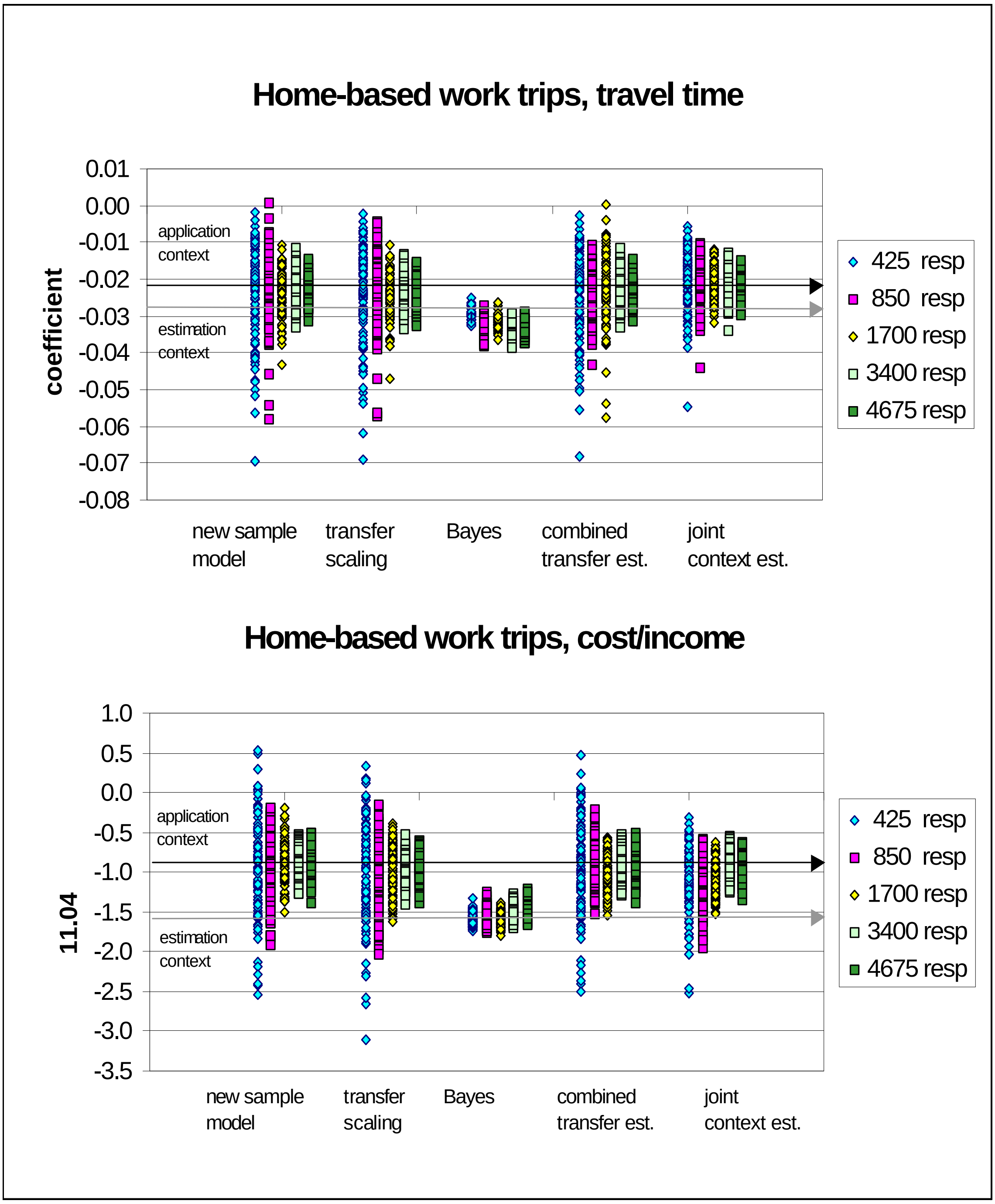


Figure 36: *The variation of coefficients of travel time and cost per income by transfer method in the Turku region in 1997; home-based work trips.*

Results based on the ratio of time and cost coefficients (VOT)

Figure 37 presents the variation of value of time (VOT). The VOT based on the whole estimation context data (2.25 euro/h) is marked by a grey arrow and the VOT estimated from the entire set of application context data (3.60 euro/h) is marked by a black arrow. Note, the VOT based on the estimation context data represents the naive transfer, that is the situation in which the estimation context models are applied in the application context. The maximum error we define as "acceptable" is rather high, 25 percent. The lower part of the figure presents the proportion of the models with an error of less than 25 percent.

The variation around "the VOT based on the entire set of data" becomes quite large in new sample models, transfer scaling or combined transfer estimation, but becomes smaller, when the sample size increases. Thus, transfer scaling, combined transfer estimation and the new sample models give quite similar results. By being highly precise, the Bayesian approach and joint context estimation differ from the other methods. However, note, in the smallest sample sizes, the joint context estimation looks worse than it actually is, because in spite of its preciseness, it slightly exceeds the allowed 25 percent error.

Due to the use of different combinations of common and data-specific variables (for samples 425 to 1,700 respondent only distance variables were estimated as data-specific, but distance variables and also time variable are data-specific for samples of 3,400 and 4,675 respondent) the variance of VOT is larger when using the sample sizes of 3,400 and 4,675 respondents than when using smaller samples. In spite of this, the observations concentrate better around the correct value when using the largest sample sizes. The next question is, how much is the improvement due to the use of more data specific variables and how much is simply a result of the larger sample size. When considering Figure 45 presented later, one can conclude, that the combination of common and data-specific variables has only a small effect in this case, and the improvement is mainly caused by the larger sample size.

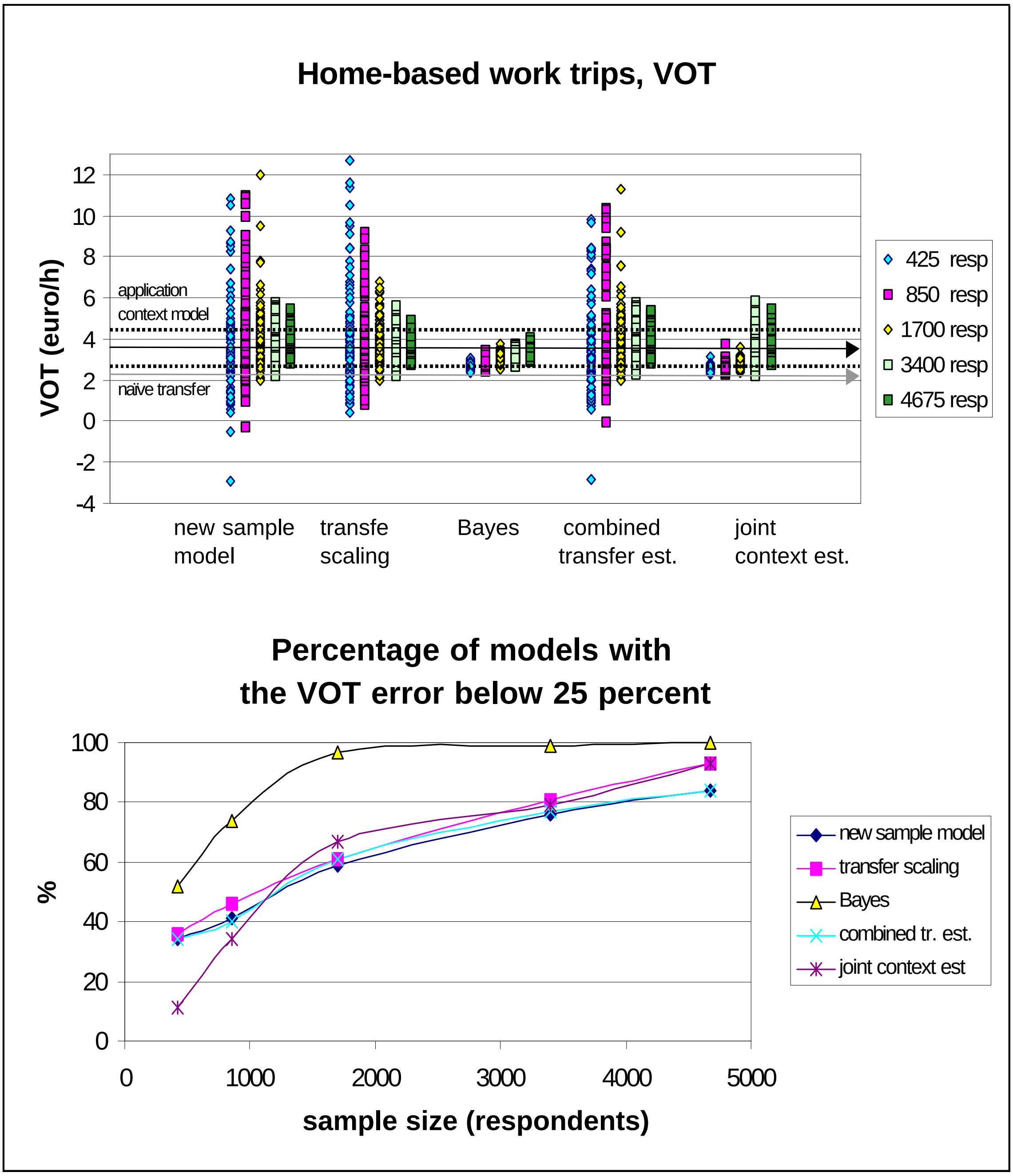


Figure 37: *The variation based on the ratio of time and cost coefficients (VOT), and the proportion of models with the error below 25 percent by transfer method in the Turku region; home-based work trips.*

Results based on the elasticity tests

Figures 38 and 39 show the effect of the variation of the transferred model parameters on the model's ability to predict the effect of changes in the transportation system. Two elasticity tests (RSEEF) are carried out on the aggregate model system. Firstly, in order to test the cost elasticity of the mode choice models, a test is carried out to examine the effect of increasing car costs by 10 %. Secondly, the effect of a 30 percent increase in public transport travel times is studied. Both tests are carried out by running the whole forecast process (based on the current land use and impedances) and including the iteration process of the alternative-specific constants (the process is described in Section 4.4.2). For the sake of comparison, the corresponding tests (RSEE) performed by applying the unweighted mobility survey data used in estimation are presented in Appendix K. The biggest difference between these two methods is the calibration of the alternative-specific constants. By calculating the elasticity with the mobility survey data, alternative-specific constants are not corrected in any way. The run of the forecast includes the iterative process in order to correct the alternative specific constants, so that the mode shares represent the weighted mode shares in the area. In addition, by calculating RSEE, only mode shares can vary and destination choices are constant, but in calculating RSEEF, destination choices can also vary.

In Figures 38 and 39, the black arrows mark the "real" decrease in car (-1.1%) or public transport share (-5.1 %) represented in the application context situation. The grey arrows mark the naive transfer. The 30 percent increase in the public transport travel time represents the situation in which the transfer bias is small but the variation of coefficients based on the small samples is large. On the other hand, the effect of a 10 percent increase in the cost of travel by car represents a situation in which the transfer bias is large (even the variation of coefficients is still quite large). The "acceptable" relative error (25 percent) in Figures 38 and 39 is marked by black dash lines.

The variation around "the correct value" becomes quite large due to the use of the minimal samples, but is smaller, when the sample size increases. The effect of a 30 percent increase in public transport travel time produces a -0.8 to -10.0 percent change in the public transport share when the minimal samples are used. The variation of the effect of a 10 percent increase in car costs, on the car mode share is +2.0 to -3.9 percent. The average RSEEF, as a function of sample size, is presented for all transfer methods in Appendix I. The RSEEF varies from 101 to -81 percent, for a 30 percent increase in public transport travel time and from 107 to -221 percent for a 10 percent increase in car costs. Thus, by considering the cost elasticities, the error is larger than the error for travel time elasticities. One reason for this might be the large transfer bias in the cost per income coefficient.

The results show, that the methods are quite similar. Only the Bayesian approach differs from the other methods. Consequently, in our data the effect of the sample size is larger than the difference between the best four methods. By comparing the RSEEF-results based on the sample enumeration tests (RSEE) presented in Appendix I, it is observed, that the iteration of alternative-specific constants is helpful.

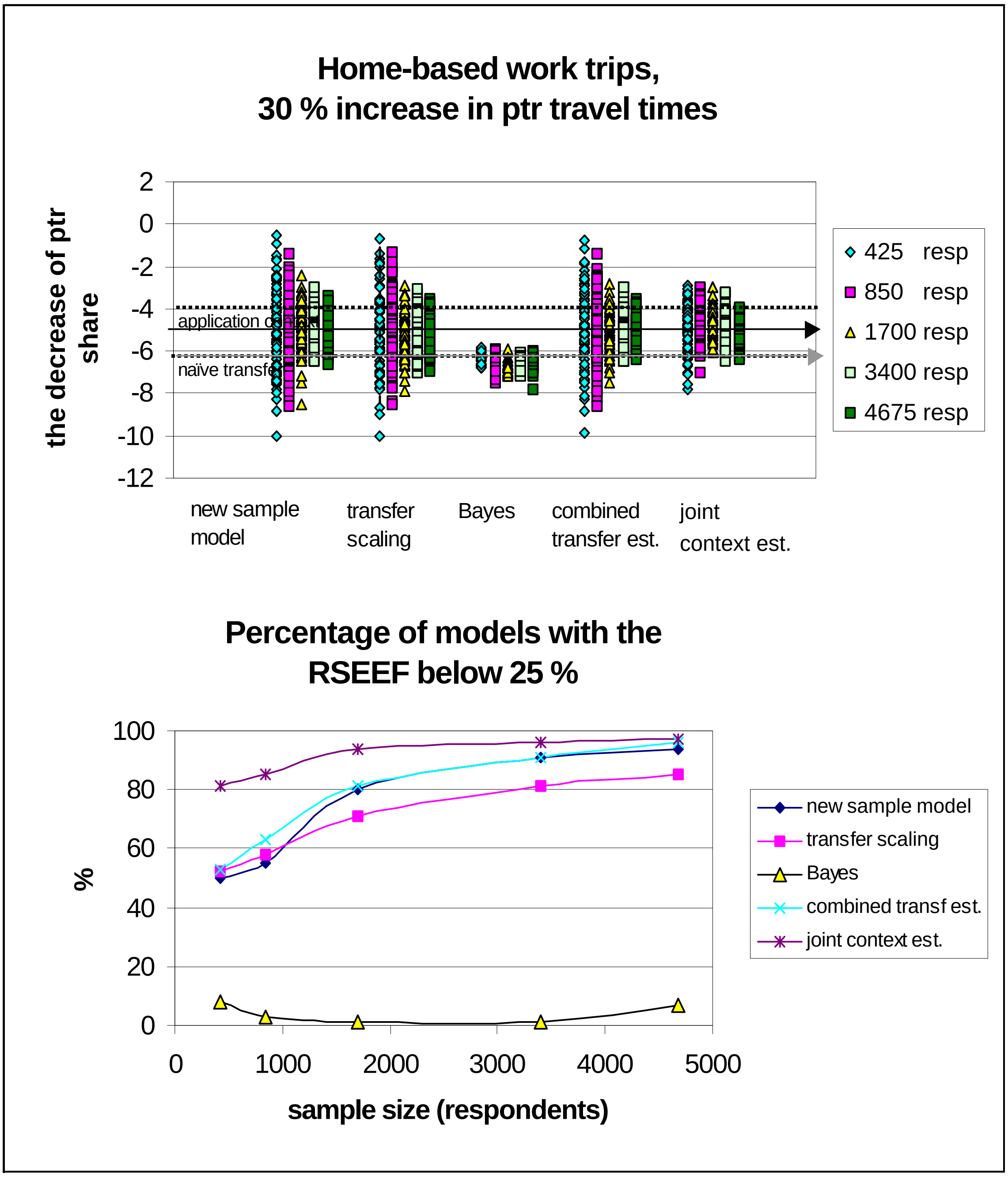


Figure 38: The variation of change in public transport (ptr) shares by transfer method, when public transport travel time is increased by 30 percent in the Turku region; home-based work trips.

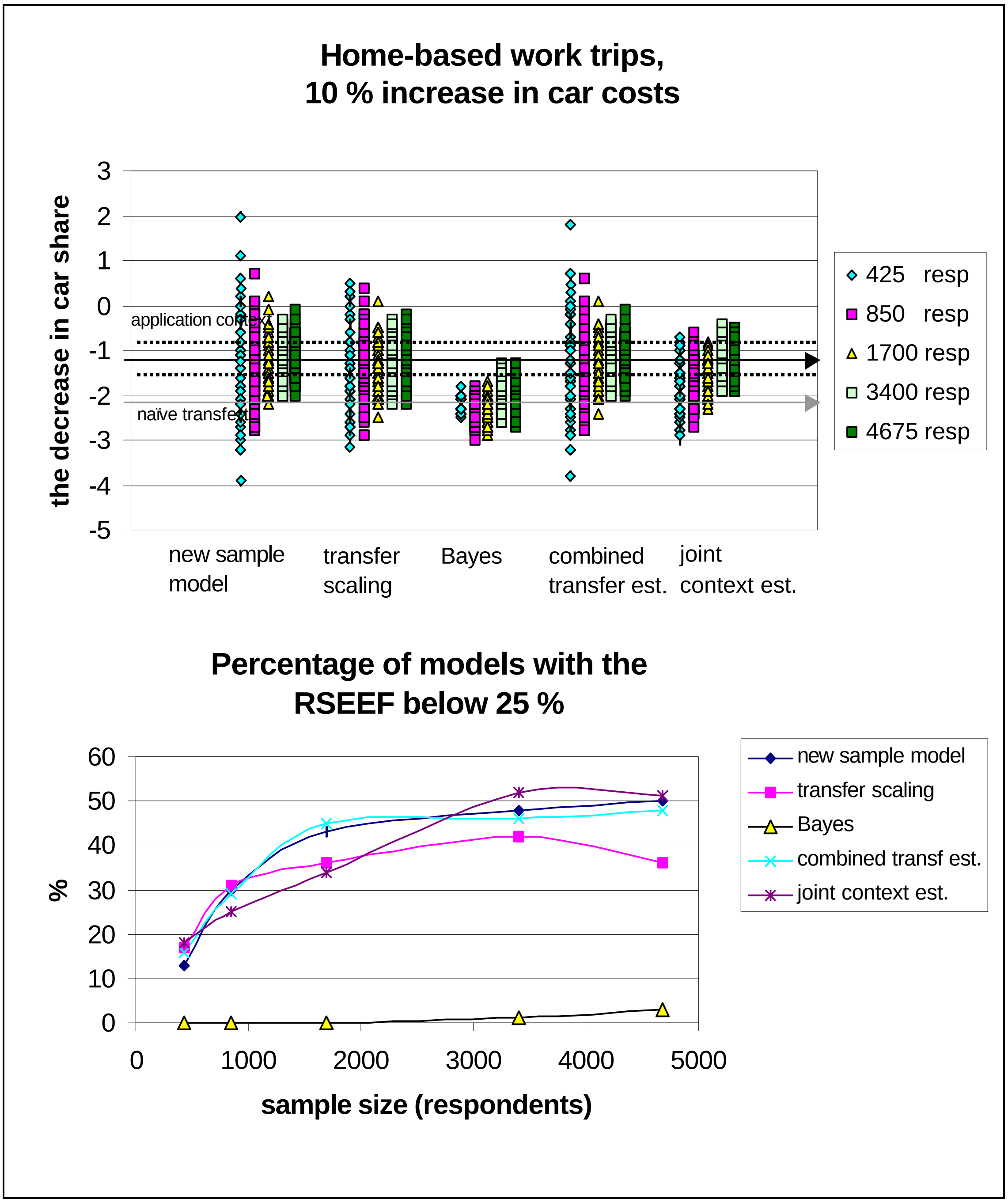


Figure 39: *The variation of change in car shares by transfer method, when car costs is increased by 10 percent in the Turku region; home-based work trips.*

Results based on the disaggregate measures of transferability

The results based on the TTS- and TI -values are summarized in Figures 40 to 43. The closer to zero the TTS-value is, the more likely the transferred model is to be similar to the application context model estimated from the entire set of data. TI has an upper bound of one which is attained when the transferred model performs as well as the model estimated using the entire set of the application context data. As sequential estimation is used, the results of the mode and destination choice level are presented separately.

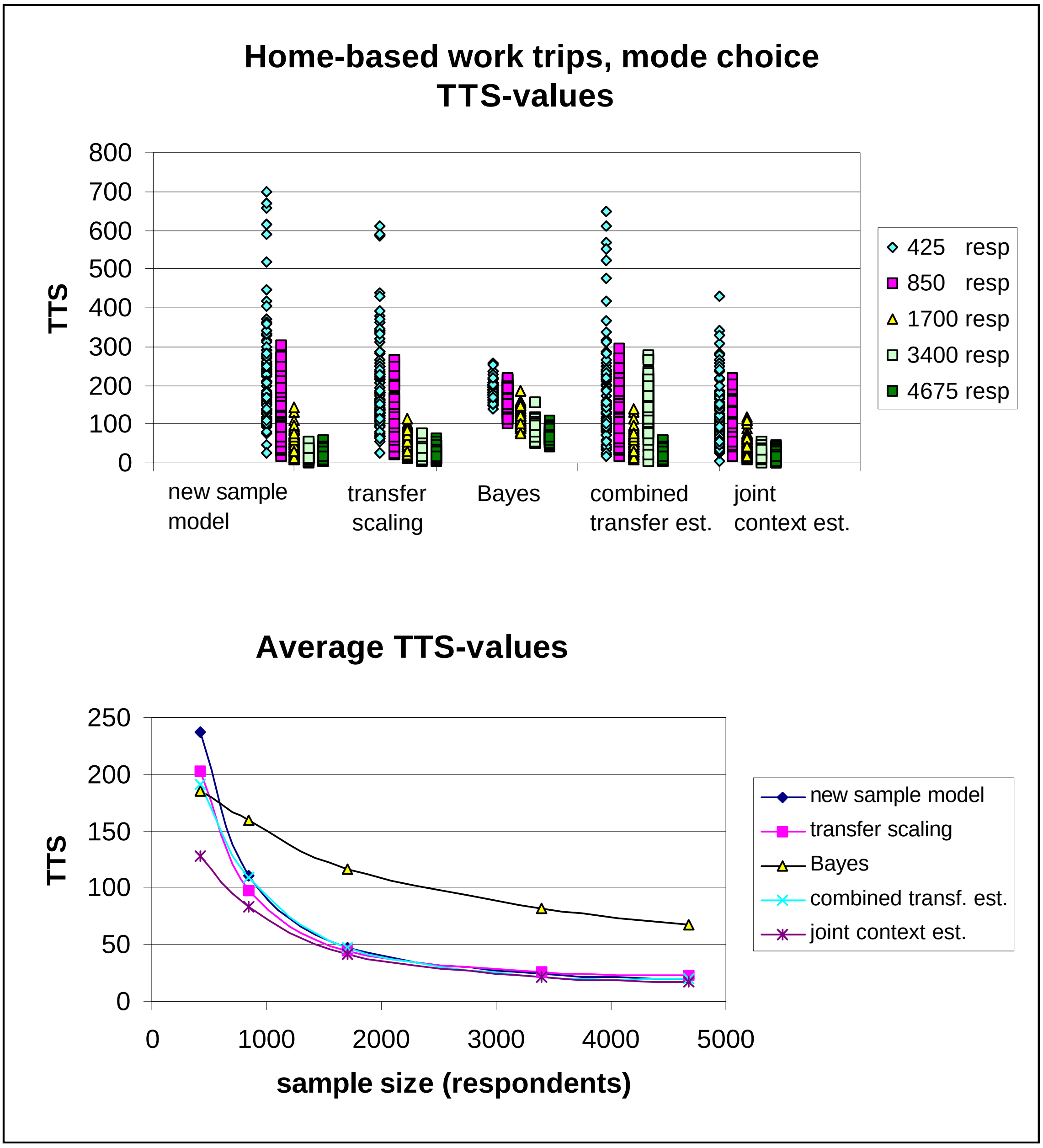


Figure 40: The variation of TTS-values and the average TTS-values for transferred mode-choice models in the Turku region; home-based work trips.

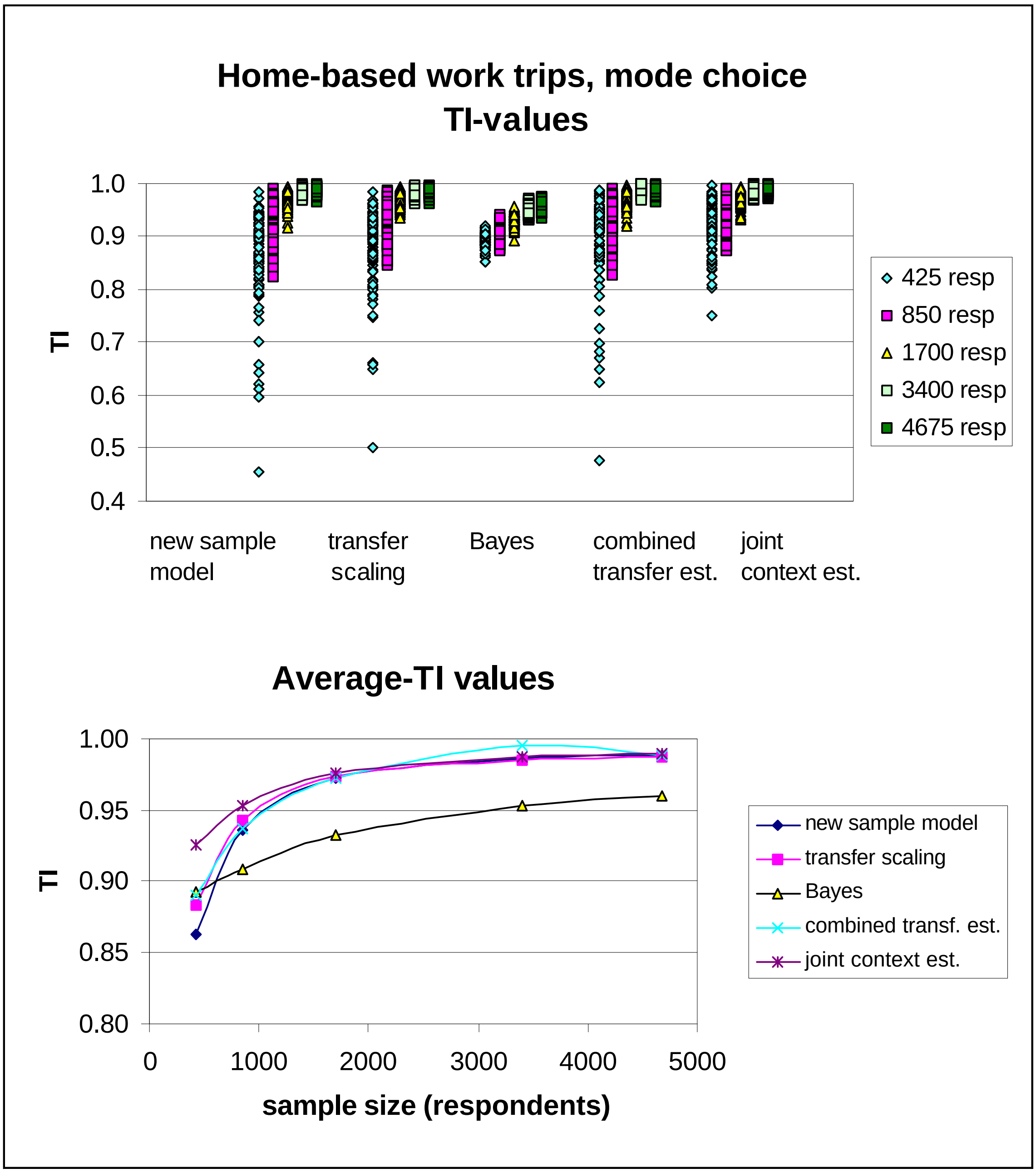


Figure 41: *The variation of TI-values and the average TI values for transferred mode-choice models in the Turku region; home-based work trips.*

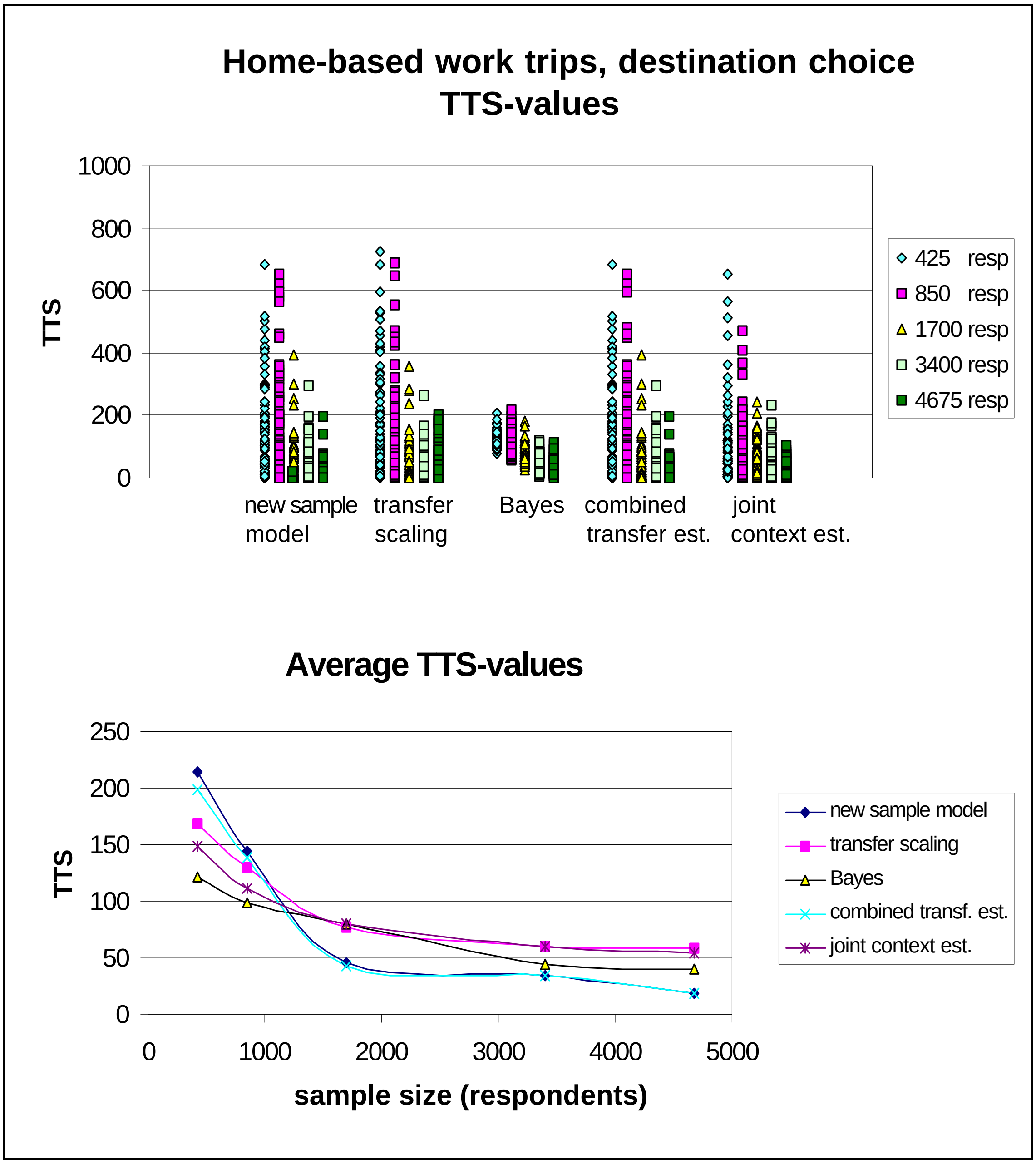


Figure 42: The variation of TTS-values and the average TTS-values for transferred destination-choice models in the Turku region; home-based work trips.

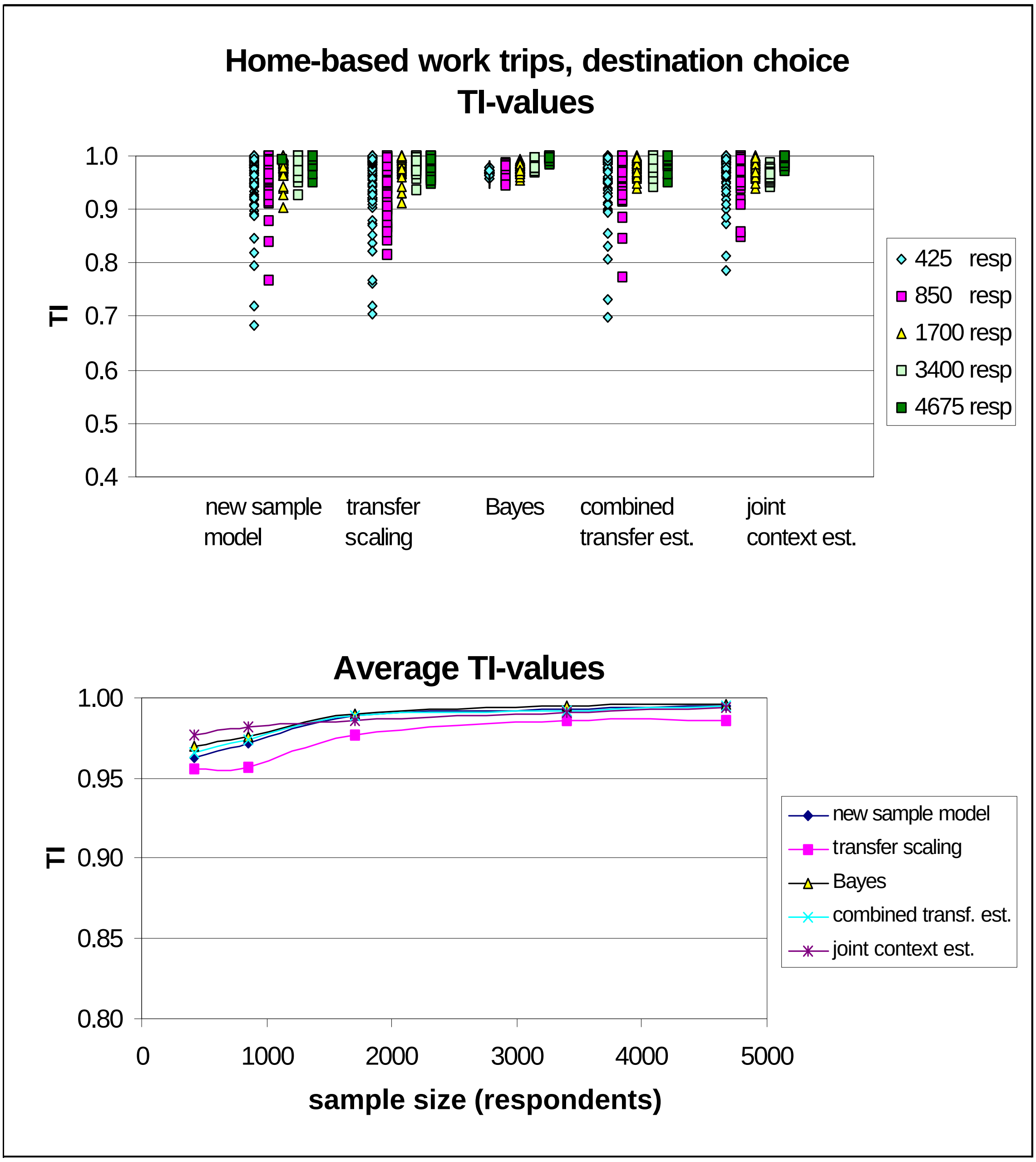


Figure 43: The variation of TI-values and the average TI -values for transferred destination-choice models in the Turku region; home-based work trips.

The illustrations show that the TTS-values are very high showing that the null hypothesis of the equality of models was rejected in most cases. On the other hand, the transfer index value TI, which is the ratio of the proportion of explained variation in estimation context data by transferred model to proportion of explained variation by the best possible estimation context model, is uniformly high, ranging in value from 0.86 to 0.99 for mode choice models and from 0.94 to 0.99 for destination choice models.

The differences between the best transfer methods are quite small. The Bayesian approach is also stable in this case, but does not yield excellent predicting performance in the mode choice level. It is to be noted that joint context estimation yields excellent results for the mode choice level, but is not so competitive at the destination choice level. This is because destination choice depends on local conditions and now, the coefficient of ln(jobs) has been estimated as a common variable. The joint context method brings such noisy information into the application context in the form of surveyed data, in contrast to other methods only bringing parameters and variances. In this case, the models based on the Bayesian method, combined transfer estimation or new sample models would be preferable for destination choice level although the joint context estimation has been used in the mode choice level.

The results based on the aggregated trip distribution

Figure 44 presents the transferred models ability to replicate observed aggregate shares. To assess predictive performance, the Turku region is subdivided into 4 regions (Central city area of Turku, suburb, Kaarina+Raisio+Naantali and Lieto+Piikkiö) Aggregate origin-destination-predictions of the travel demand (whole model system) are obtained for each of these regions. Mean absolute error (MAEF) values are computed using the aggregate predictions to assess forecast accuracy (see Appendix L). Thus, the test is carried out by running the whole forecast process and includes the iteration process of the alternative-specific constants, so that the mode shares represent the weighted mode shares in the area. The prediction efficiency is better the smaller the MAEF-index is.

Figure 44 shows that different methods give quite similar results. By being quite precise, the Bayesian method differs slightly from the other methods. The MAEF-value for naive transfer is 0.22, and the corresponding value for the model based on the entire set of data is 0.20. Thus, maybe due to the iteration of alternative-specific constants, the difference between the naive transfer and the model based on the entire set of application context data is not notable.

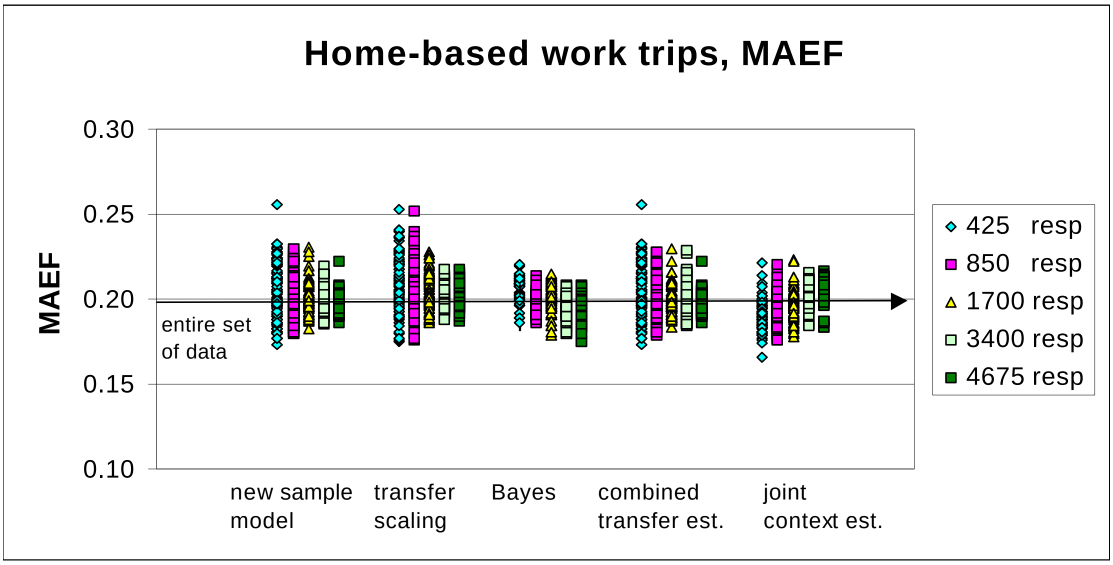


Figure 44: *The variation of MAEF-index by transfer method for home-based work trip models.*

The results relating to alternative specifications for using joint context estimation

The results relating to different ways of combining common and data-specific variables by using the joint context estimation are presented in Figures 45 and 46 and in Appendix M. The combinations used when comparing different transfer methods are presented in Table 23 (the model symbol referred to in the brackets is the model type presented in Chapter 4.5 in Table 11). The best combination was chosen based on the sum of MD-values for the three most important transport policy coefficients (travel time, costs and cars per household). Additionally, the total MD-value for all coefficients was calculated. The conclusions based on these two sum-measures are quite similar. The main principle is that the more accurate the models are (the larger the sample size is) the more efficient it is to use data-specific variables. The combinations in which the precise variables (e.g. distance) are estimated as data-specific and imprecise variables (e.g. cars per household) as common give the best results. For the destination choice level, the coefficient for ln(jobs) is fixed to one and the log sum variable is always estimated as common.

Table 23: *The best combinations of data-specific variables based on the MD in joint context estimation; home-based work trips.*

Sample size	The best combination of data-specific variables based on the MD
	mode choice
425	distance 0-5, (model B)
850	distance 0-5, (model B)
1,700	distance 0-5, (model B)
3,400	distance 0-5, distance 5, travel time, (model F)
4,675	distance 0-5, distance 5, travel time, (model F)

* Destination choice level-no data-specific coefficients

Next, the chosen combinations (that is the combinations used in Section 6.1 and presented in Table 23) are compared to the unchosen combinations to see, how well the chosen combination performs with respect to the other combinations. Figure 45 presents the variation of the VOT based on each combination of data-specific and common variables. Correspondingly, Figure 46 presents the results of the sample enumeration tests (RSEE). The RSEE is used instead of the RSEEF because it is easier and less time-consuming to calculate than the RSEEF and the relationship between the RSEE and the RSEEF is known (Appendix I)

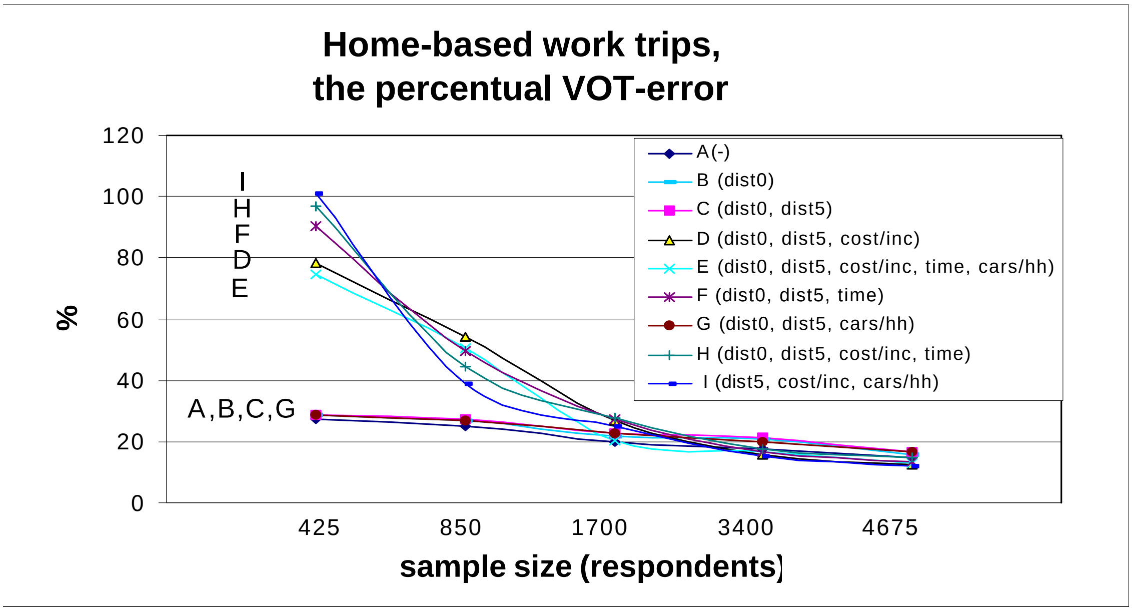


Figure 45: The comparison of the error in the ratio of time and cost coefficients (VOT error) when using different combinations of joint context estimation (the symbols A to I refer to data-specific variables in the models); home-based work trips.

The examination of VOT shows that, due to the impreciseness of the coefficients when using the smallest sample sizes, the combinations A, B, C and G are preferable when using samples of the size 425 to 1,700 respondents. This is especially, because the time and cost/income variables are estimated as common when using these combinations. Upwards of the sample size of 1,700 respondents the differences between the combinations are rather small. In two cases out of the five, the best combination based on the VOT-criteria differed from that based on the MD-criteria. This is mainly because the MD-criteria is based on the goodness of the whole model and the VOT is only concerned with the ratio of two predetermined coefficients, namely travel time and cost/income. However, the difference between the chosen and the best combination was in these cases very small. Hence, the results presented in Section 6.1 can be regarded as representing quite well the best combination of joint context estimation.

When considering the RSEE-values presented in Figure 46, the advantage of using data-specific variables can be seen. When considering the elasticities based on the 10 percent increase of travel costs, the transfer bias is large and the use of data-specific variables proved to be useful regardless of which variables were estimated as data-specific. Due to the small transfer bias, the RSEE-error is smaller for time-elasticities and the importance of the chosen combination is smaller as well. The best combination based on the RSEE-values is in three cases from five the same when based on the MD-values. However, when it is different, the difference with respect to the best model was rather small.

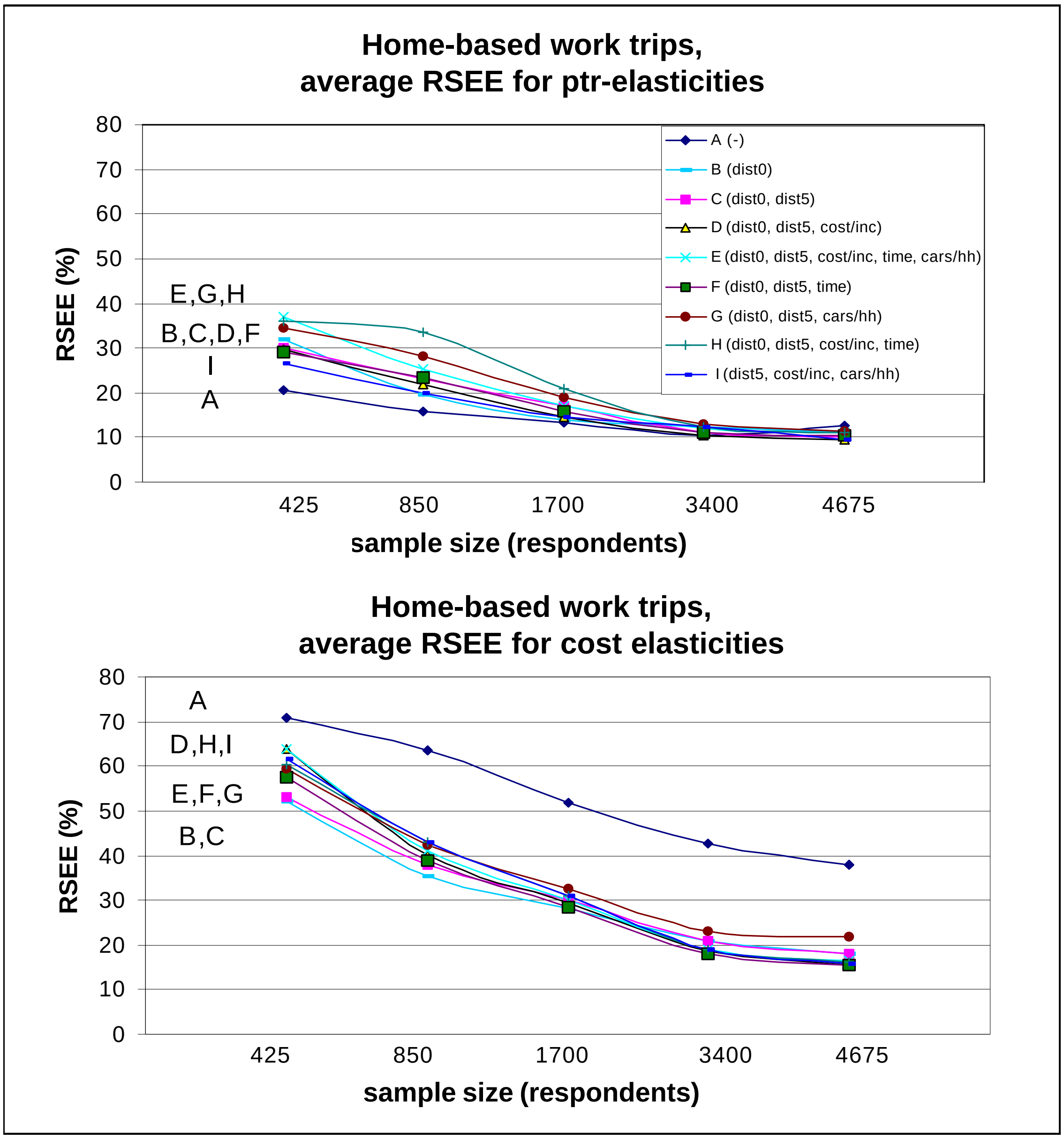


Figure 46: The comparison of the RSEE when using different combinations of common and data-specific variables for joint context estimation (the symbols A to I refer to data-specific variables in the models); home-based work trips.

Comparison of the test results

Next, the test results of the different tests are compared to each other. The main purpose of the comparison is to summarize the results based on the different test measures and to evaluate how often different test measures give similar results.

Table 24 presents the success rate for different test measures. The success rate is defined as the percentage of models where the method was the best according to each test measure. Note, the success rate does not describe the size of differences between the test measures but only how many times each transfer method was assessed to be the best one. On many occasions, in particular when considering the largest sample sizes, the differences between the methods were rather small.

Table 24: The success rate (%) of the transfer methods by the test measure in home-based work trips (subscript t and c refer to time and cost).

					425 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	10	7	16	12	19	0	15	1	0	29
transfer scaling	19	17	27	16	16	7	25	9	4	24
Bayes	55	9	5	3	5	12	8	25	11	7
combined est.	8	20	17	22	10	25	29	19	24	20
joint context est	8	47	35	47	50	56	23	46	61	20
					850 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	13	6	25	11	24	1	39	18	2	41
transfer scaling	19	19	28	28	25	31	29	38	35	21
Bayes	54	4	1	0	2	4	1	4	3	2
combined est.	11	11	18	12	8	11	19	25	10	30
joint context est	3	60	28	49	41	53	12	15	50	6
					1,700 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	12	20	15	11	29	6	38	32	9	39
transfer scaling	16	18	34	26	23	12	14	13	23	20
Bayes	61	1	0	0	0	0	2	1	0	2
combined est.	10	10	28	16	8	34	39	28	20	25
joint context es	1	51	23	47	40	48	7	26	48	14
					3,400 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	20	16	16	10	24	10	39	41	11	41
transfer scaling	18	21	24	26	26	16	17	8	17	17
Bayes	43	0	2	0	0	0	0	0	1	0
combined est.	14	17	17	15	8	14	43	23	12	42
joint context es	5	46	41	49	42	60	1	28	59	0
					4,675 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	10	30	25	18	35	5	40	49	8	43
transfer scaling	20	8	17	24	18	10	5	0	11	4
Bayes	58	0	1	0	0	0	16	2	0	13
combined est.	7	20	24	6	8	6	34	20	5	32
joint context es	5	42	33	52	39	79	5	28	76	8

The main findings relating to the characteristics of the different tests are:

- The joint context estimation was in most cases the best method. However when considering the VOT, the Bayesian method gave the best results.

- Although the joint context estimation was usually the best method at mode choice level, new sample models and combined transfer estimation gave better results at the destination choice level.

- TTS was the most sensitive to the sample size. That is, the larger the sample size the better the joint context estimation proved to be at mode choice level. At destination choice level, new sample models and combined transfer estimation gave the best results.

Next the percentage of cases in which two different tests give similar recommendations for the best method are considered. Not all the possible comparisons are made, but we have presented some basic findings made from the comparison.

Table 25: The number of cases in which each test-pair recommend the same transfer method when considering home-based work trips.

	sample size (respondents)				
	425	850	1700	3400	4675
VOT*/TTS_{total}	21	16	15	18	16
TTS_{mode}/TI_{mode}	96	96	95	96	97
VOT/RSEEFt	13	14	7	13	12
VOT RSEEFc	25	20	10	19	19
VOT/RSEEt	14	15	7	13	12
VOT/RSEEc	21	16	13	24	14
RSEEFc/RSEEc	60	49	27	42	60
RSEEFt/RSEEt	66	54	57	47	69

Table 25 shows that the results based on the TTS and TI are nearly identical. The RSEE and RSEEF also give quite similar results. However, the comparison of other tests indicated more contradictory results. This is mainly due to the fact that the Bayesian method gave the best results when considering the VOT whereas the joint context estimation proved to be the best method when using the other tests.

6.2.2 Other home-based trips

Table 26 contains mode and destination choice models for other home-based trips estimated from the entire set of data. The model specifications are presented in Section 4.5. The value of time estimated from the Turku data is 1.1euro/h (6.7 FIM/h). The value of time estimated from the HMA data is 1.5 euro/h (8.9 FIM/h). The value of time is based on the assumption of an average income of 2,144 euro/month/household (12,750 FIM/month/household) in the HMA and 2,405 euro/month/household (14,300 FIM/month/household) in the Turku region.

Table 26: Estimation results of other home-based trips using the entire set of data collected in the Helsinki metropolitan area in 1995 and in the Turku region in 1997.

Variable	Helsinki Metropolitan Area			Turku Region		
	Coefficient	Std.	t-value	Coefficient	Std.	t-value
Distance 0-5 km (Walk)	-0.7812	0.0325	-24.0	-1.0380	0.0212	-49.0
Distance > 5 km (Kv)	-0.3539	0.0290	-12.2	-0.2470	0.0183	-13.5
Total travel time (Car, Ptr)	-0.0197	0.0032	-6.2	-0.0135	0.0019	-7.3
Number of transfers (Ptr)	-0.1189	0.0615	-1.9	-0.4372	0.0415	-10.5
Trip cost/income (Car, Ptr)	-1.8910	0.1170	-16.1	-1.7390	0.0853	-20.4
Cars/household (Car)	1.3700	0.2010	6.8	1.1730	0.0881	13.3
Walk dummy (Walk)	0.7118	0.1310	5.4	2.1570	0.1060	20.4
Car dummy (Car)	-1.7690	0.1850	-9.5	-0.8932	0.1150	-7.7
Log sum (Dest.)	1.5670	0.0187	83.8	1.2020	0.0089	134.8
Scale-factor	1.0000	0.0000	-	1.0000	0.0000	-
Number of trip observations	4,149			13,989		
ρ²(c)	0.2008 (mode), 0.3667 (dest)			0.2265 (mode), 0.2806 (dest)		
Walk = Walk and bicycle Ptr = Public transport Car = Car	Scale factor: - inhabitants 1.00 - retail employment 28.73 - service employment 2.29			Scale factor: - inhabitants 1.00 - retail employment 14.89 - service employment 2.52		

Table 24 shows that the coefficients for the “number of transfers” and distance variables differ significantly in estimation and application context. The difference for number of transfers between can at least be partially explained by the uncertainty of the “number of transfers” variable due to the difficulties in controlling this variable using the Emme/2 assignment program.

Figure 47 presents the variation of travel time and cost coefficients by transfer method. The coefficients based on the entire set of estimation and application context data are marked by the arrows.

The figure shows that the variation is great when using models based on small samples, transfer scaling or combined transfer estimation approaches. However, the differences between these methods are, on the whole, quite small. By being highly precise, the Bayesian approach differs greatly from the other methods. Due to the small transfer bias between the original coefficients in the estimation and application context, the Bayesian approach performs especially well when considering cost per income variable. The joint context estimation also gives quite precise results for the time variable.

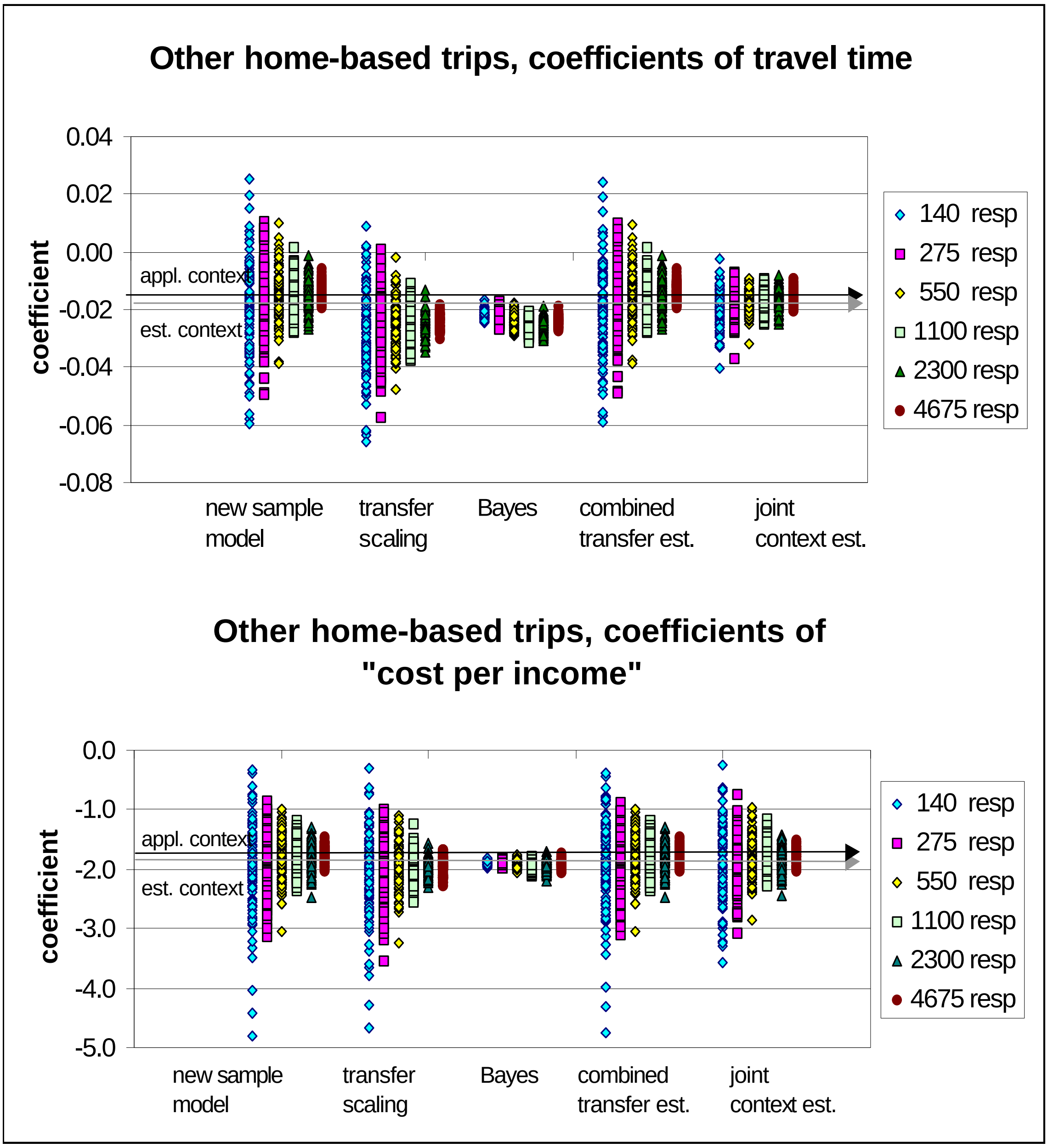


Figure 47: The variation of coefficients of travel time and cost per income by transfer method in the Turku region in 1997; other home-based trips.

Results based on the ratio of time and cost coefficients (VOT)

Figure 48 presents the variation of value of time (VOT) for other home-based trips. The grey arrow marks the VOT, calculated from the entire set of estimation context data (naive transfer) and the black arrow from the entire set of application context data. The bottom of the Figure presents the proportion of the models with the error below 25 percent.

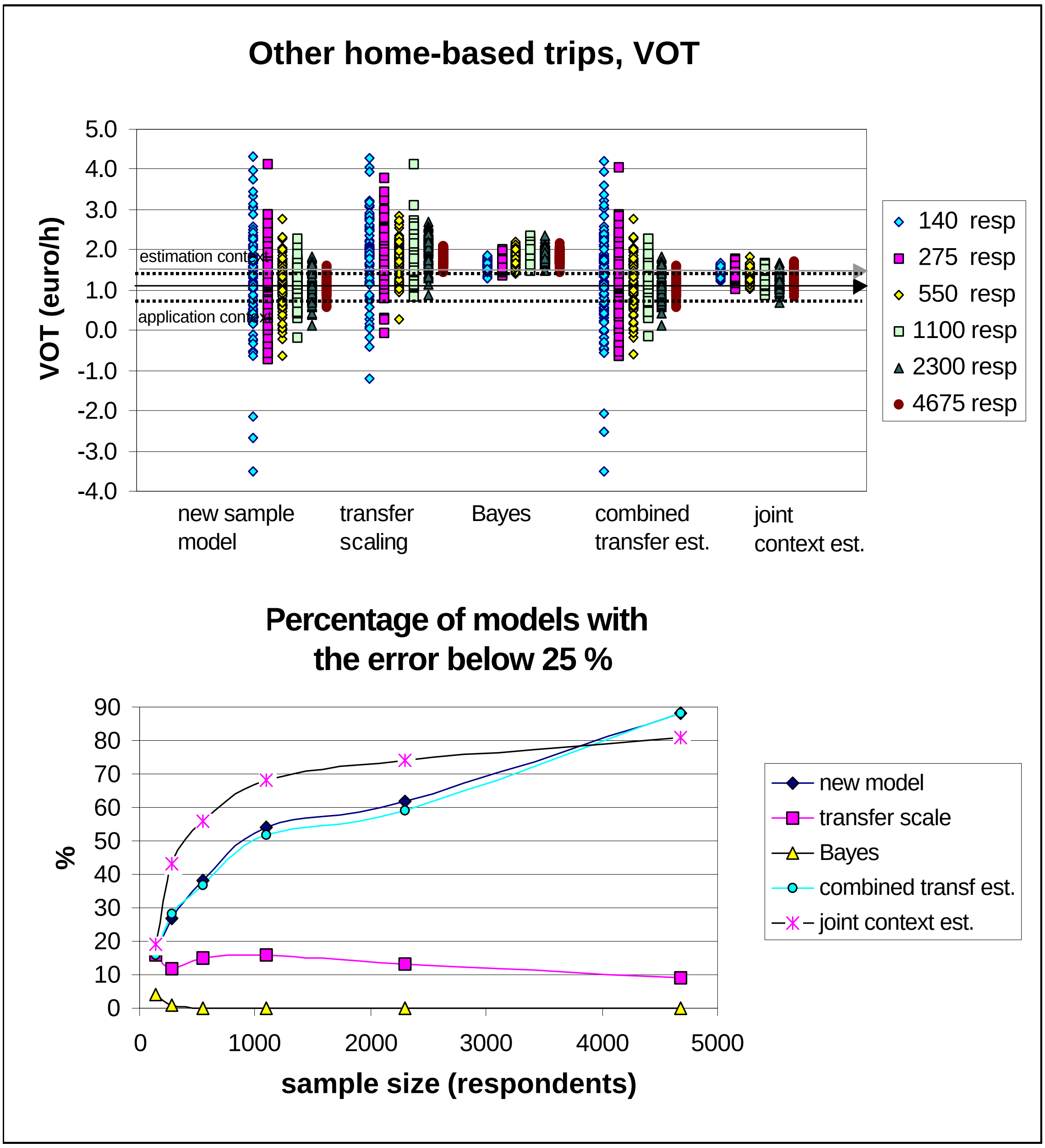


Figure 48: The variation based on the ratio of time and cost coefficients (VOT), and the proportion of models with the error below 25 percent by transfer method in the Turku region;other home-based trips.

The variation around “the VOT based on the entire set of data” becomes quite large due to the use of the minimal samples, but this is smaller, when the sample size increases. The joint context estimation yields the best results excluding the full sample models for which the combined transfer estimation and new sample models yield best results giving the 88 percent proportion of models with an error below 25 percent. By using the smallest sample size, the differences between the methods are initially rather small, but become larger as the sample size increases. The joint context estimation performs poorly for the smallest sample size, because many coefficients are so imprecise that they have to be estimated as common coefficients (Tables 11 and 27). This means that the estimation context data is emphasized too strongly. On the other hand, also when the largest sample size is used, the models based on the joint context estimation are not as good as one could expect. This is because three data-specific coefficients (distance variables and cost/income) were utilized. The use of data-specific coefficients improves the quality of the number of transfers in public transport (due to the large transfer bias for this coefficient) (Appendix N). However, generally, it increases the variation of the coefficients and gives poorer estimates for VOT than the combination used for the smaller sample sizes. If distance variables had only been estimated as data-specific, the best result for VOT would have been much better (88 percent). Due to the common scale-factor used for “travel time” and “cost per income” variables, transfer scaling does not yield good results.

Results based on the elasticity tests

The top of Figures 49 to 50 show the effect of the variation of the transferred model parameters on the model’s ability to predict the effect of changes in transportation system. Arrows mark the "real" decrease in the car or public transport share, calculated from the forecasts based on the entire sets of estimation and application context coefficients. This real value was -3.0 for the decrease in the public transport share (in Figure 49) and -2.2 for the decrease in the car share (in Figure 50).

The 30 percent increase in the public transport travel time represents a situation in which the transfer bias is rather large, and also the variation of coefficients based on the small samples is large. On the other hand, the effect of a 10 percent increase in the cost of travel by car represents a situation in which the transfer bias is small, and the variation of coefficients is also quite reasonable. The “acceptable” relative error (25 percent), in Figure 49 and 50, is marked by black dash lines. The lower part of Figure 49 and 50 presents the probability that the RSEEF error will be less than 25 percent. The corresponding test results achieved by applying the unweighted mobility survey data are presented in Appendixes J and O.

The effect of a 30 percent increase of public transport travel time produced a +8.9 to –7.4 percent change in the public transport share when the minimum samples were used. The variation of the effect of a 10 percent increase in car costs, on the car mode share was +0.7 to –4.5 percent. The RSEEF varied from -397 to +147 percent, for a 30 percent increase in public transport travel time and from –132 to +105 percent for a 10 percent increase in car costs. Thus, the coefficients of cost per income variable are more precise than the coefficients for travel time, and the transfer bias is also smaller for cost per income variable.

The new sample models and the models based on the combined transfer estimation approaches are highly sensitive to the sample size. In addition, transfer scaling gives systematically biased estimates especially for the travel time coefficient, which is scaled with the "number of transfers". Joint context estimation yields quite safe results, being the best method for the smallest samples. In spite of its stability, the Bayesian method is never able to predict the effects of a 30 percent increase in public transport travel time. However, by predicting the effect of a 10 percent increase in car costs on the car mode share, the Bayesian approach proved superior for all sample sizes due to the small transfer bias between the estimation and application context. However, this is only in the case of the alternative-specific constants for mode choice models are re-estimated iteratively. If the alternative-specific constants are not calibrated to represent the mode shares in the observed weighted mobility survey data, the Bayesian approach is unable to predict changes in the transportation system at all (Appendixes J and O).

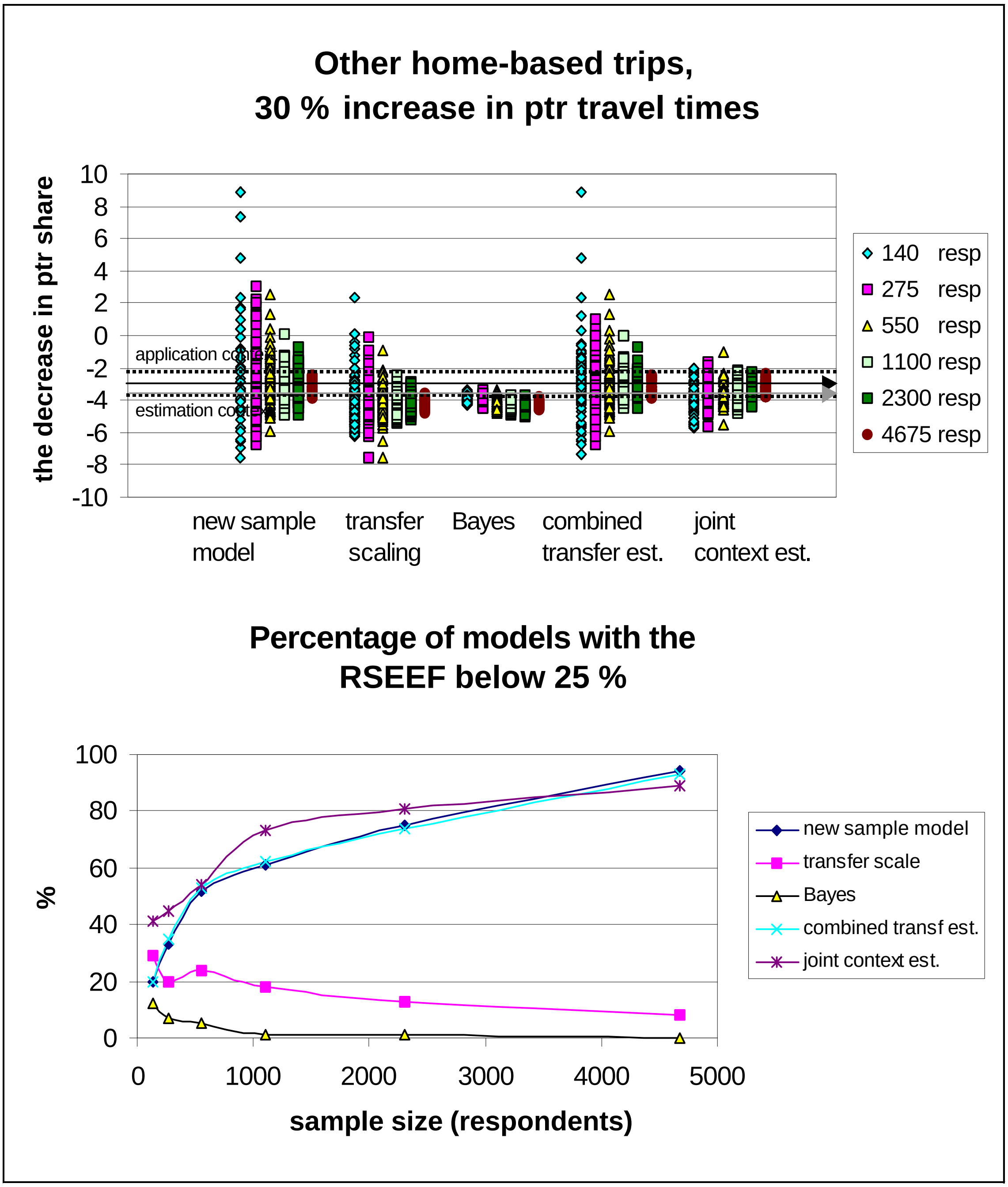


Figure 49: The variation of change in public transport (ptr) shares by transfer method, when public transport travel time is increased by 30 percent in the Turku region; other home-based trips.

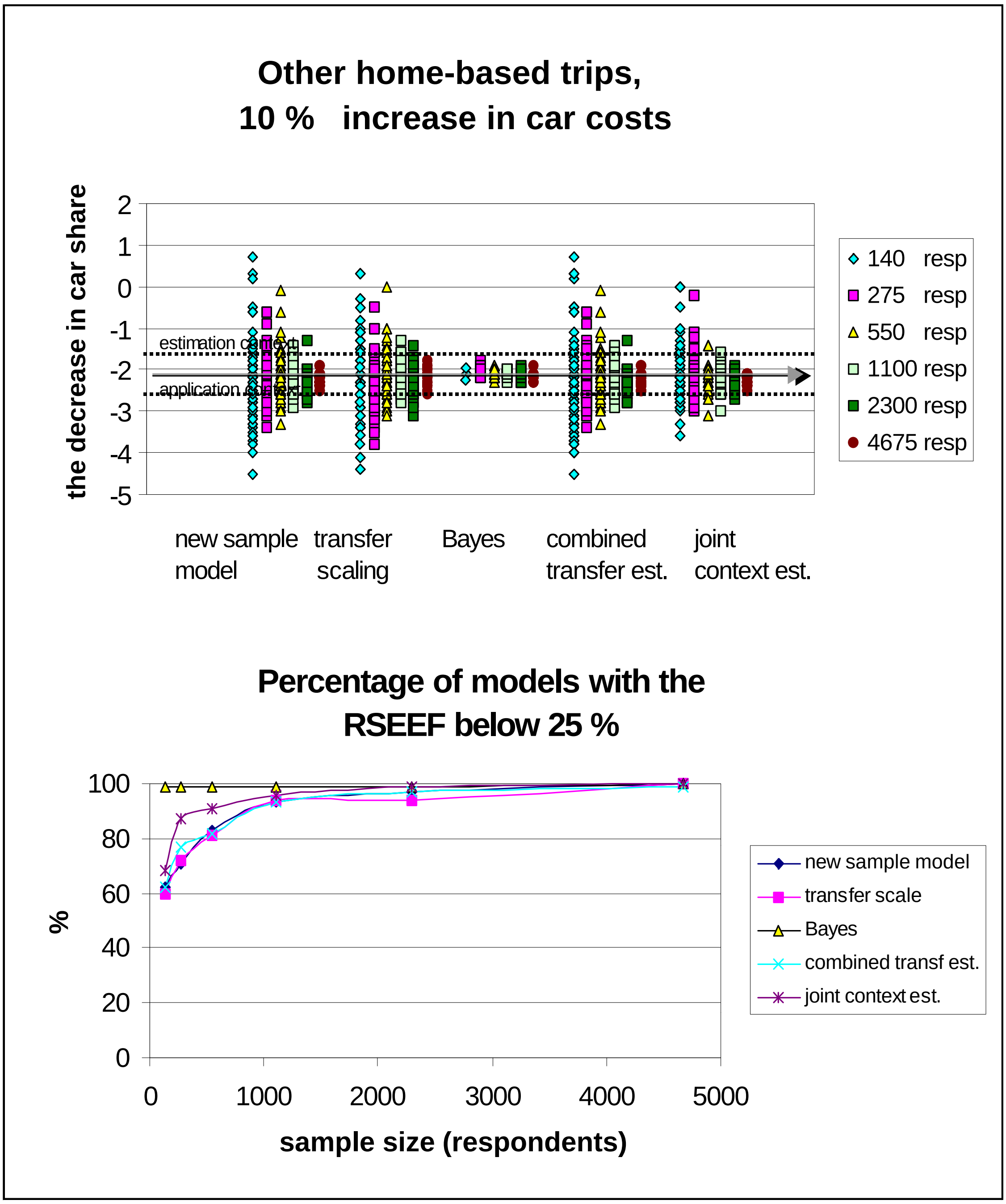


Figure 50: *The variation of change in car shares by transfer method, when car costs are increased by 10 percent in the Turku region; other home-based trips.*

Results based on the disaggregate measures of transferability

Figures 51 to 54 present the results based on the TTS- and TI-values. In this case, the magnitude of the TTS-values of transferred models or new sample models indicates that transferred models are never able to fully substitute the best application context model. On the other hand, Transfer Indexes (TI) for transferred models are in almost all cases in excess of 80 percent, suggesting that the transferred models provide a significant component of the information obtained from the application context model. The methods differ only slightly from each other; however, joint context estimation, new sample models and combined transfer estimation can be regarded as being slightly better than the other methods.

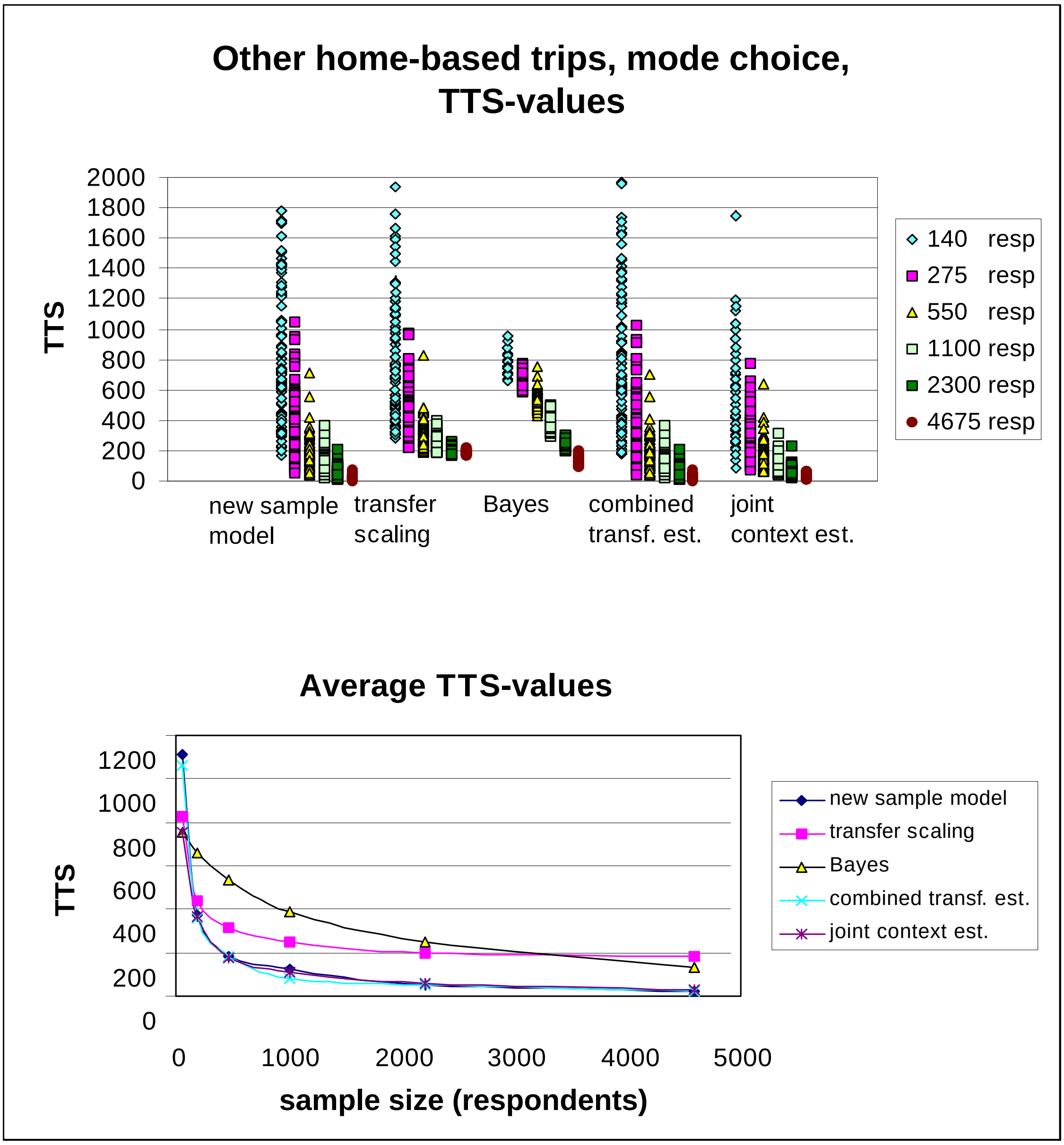


Figure 51: The variation of TTS-values and the average TTS-values for transferred mode choice models in the Turku region; other home-based trips.

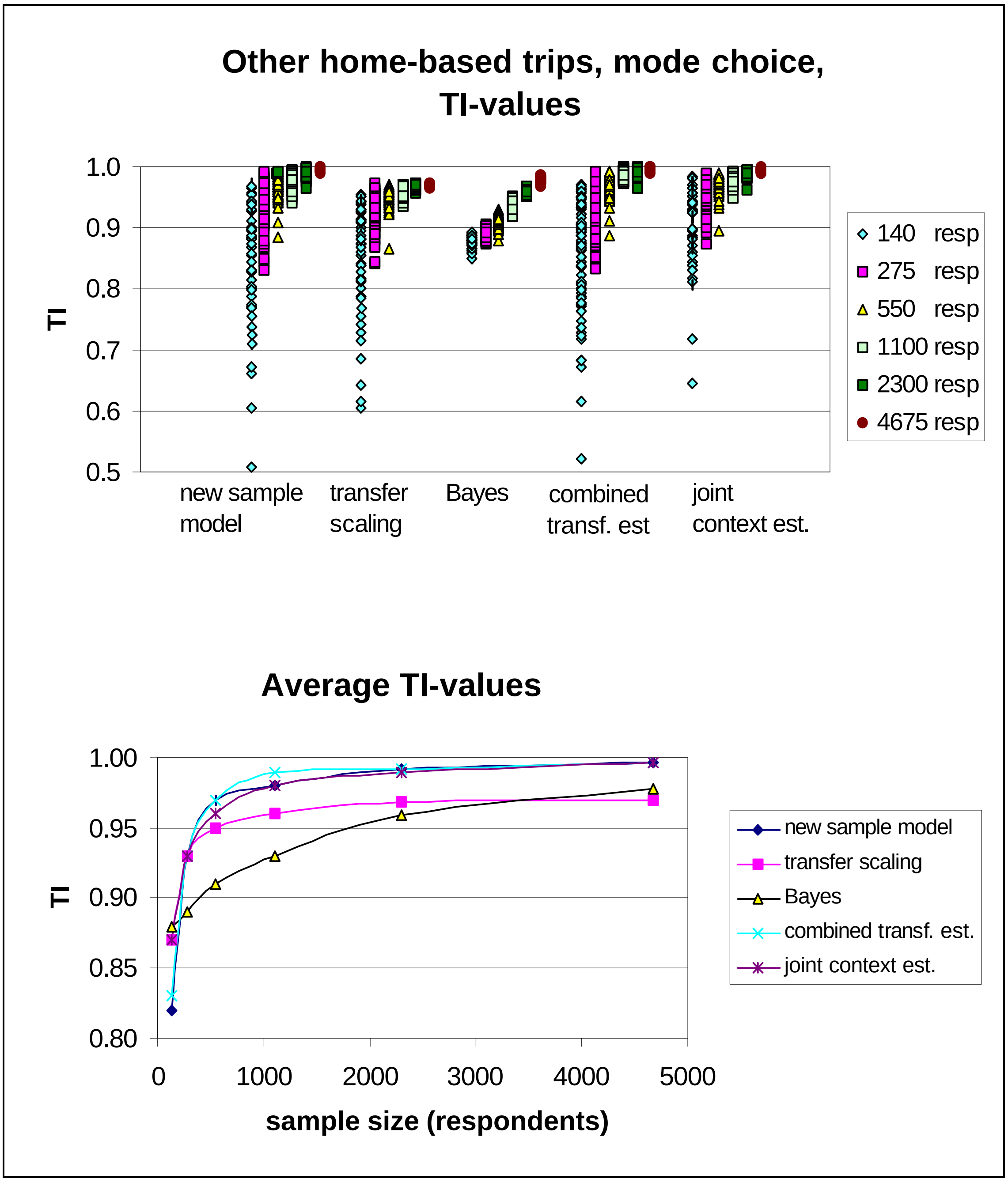


Figure 52: The variation of TI -values and the average TI-values for transferred mode choice models in the Turku region; other home-based trips.

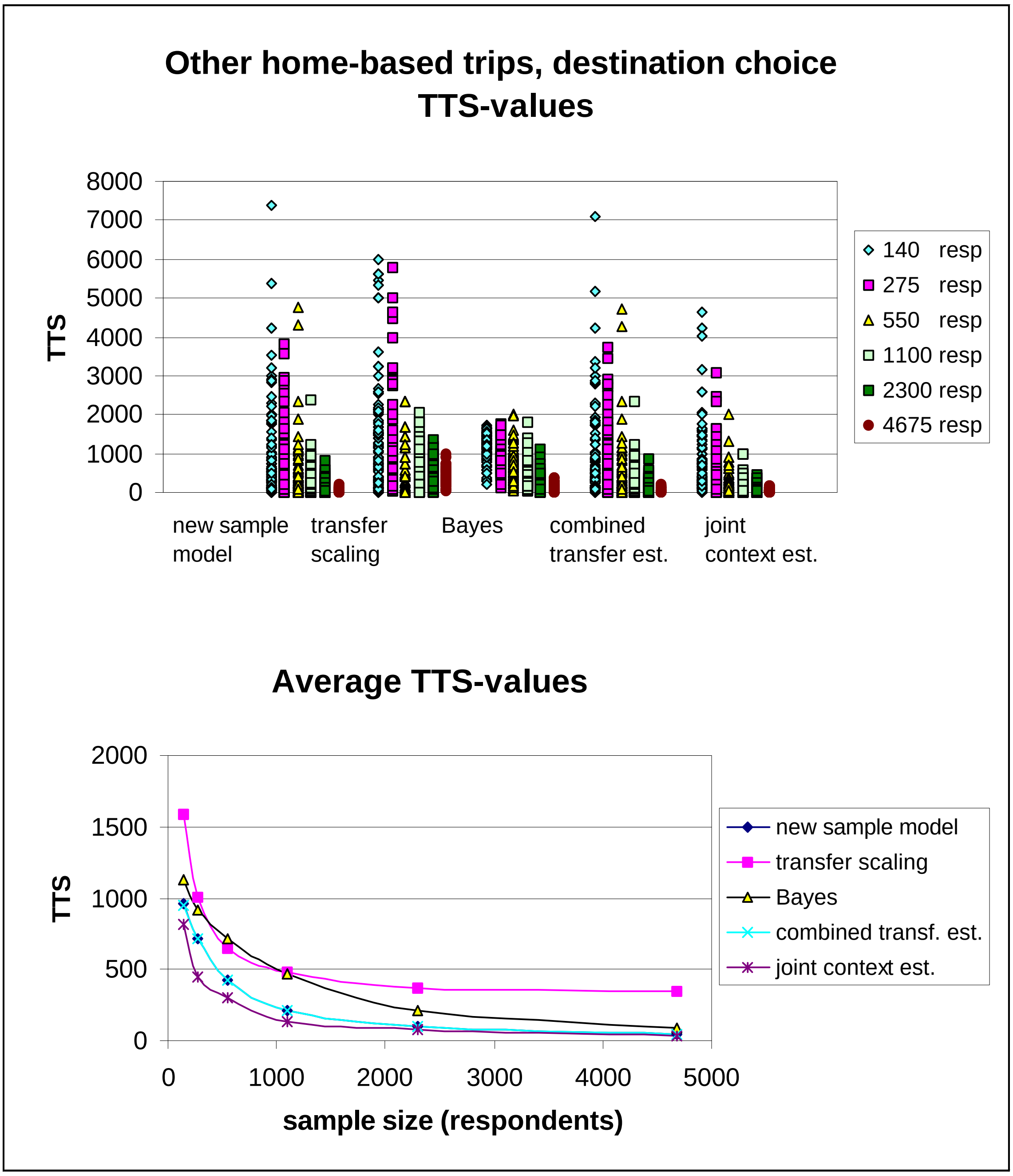


Figure 53: The variation of TTS-values and the average TTS-values for transferred destination choice models in the Turku region; other home-based trips.

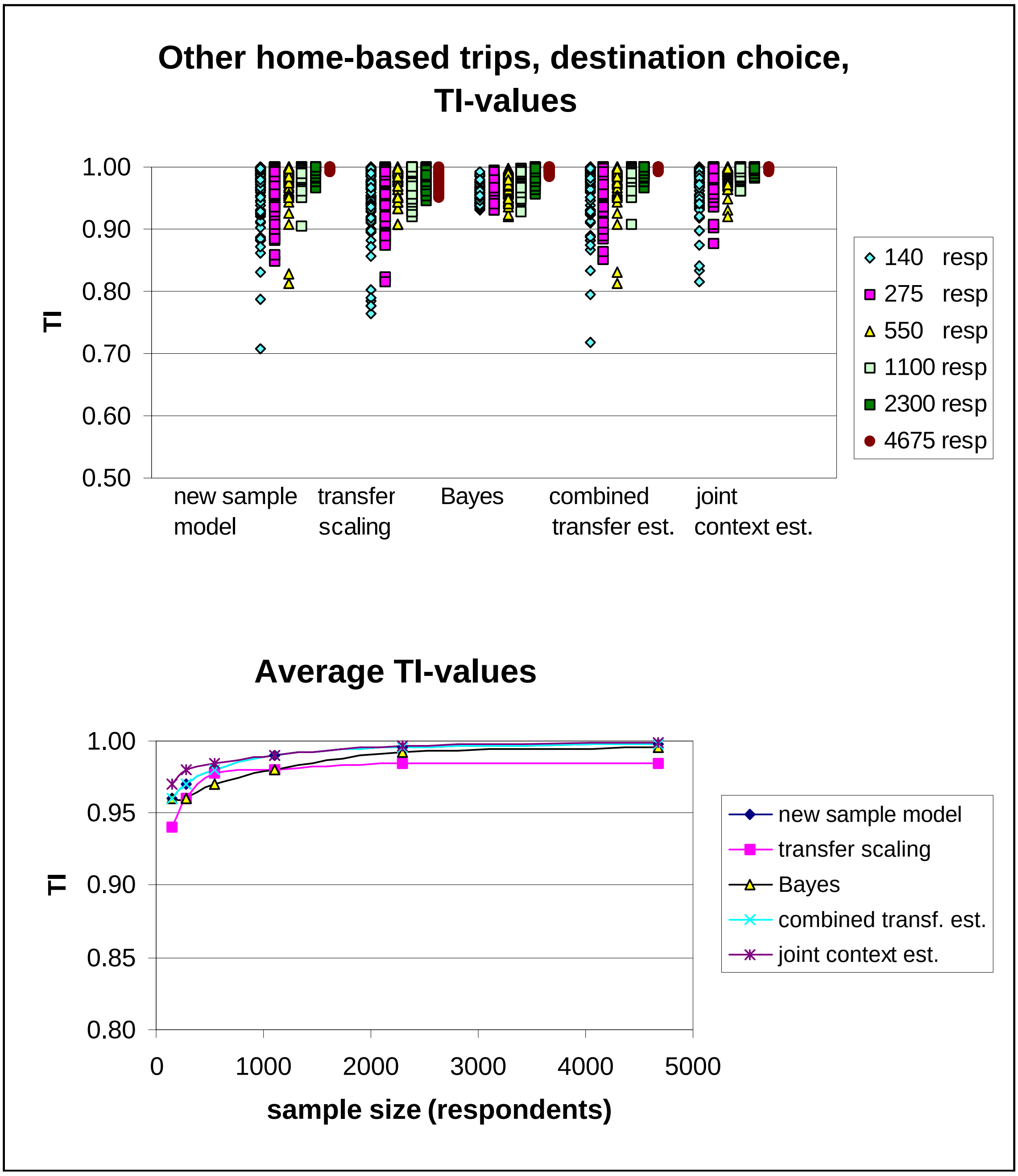


Figure 54: The variation of TI-values and the average TI-values for transferred destination choice models in the Turku region; other home-based trips.

The results based on the aggregated trip distribution

Figure 55 presents the MAEF-index calculated for the other home-based trips**.** As in Section 6.2.1 the Turku region is subdivided into 4 regions, and aggregate OD-predictions are obtained for each of these regions. The MAEF-index is calculated as a sum of the absolute differences of the observed and the predicted values. As for home-based work trips, the results based on the entire set of data give a reference point when assessing the predictive performance of the transferred models. Based on this, the acceptable MAEF-value can be regarded as being approximately 0.25.

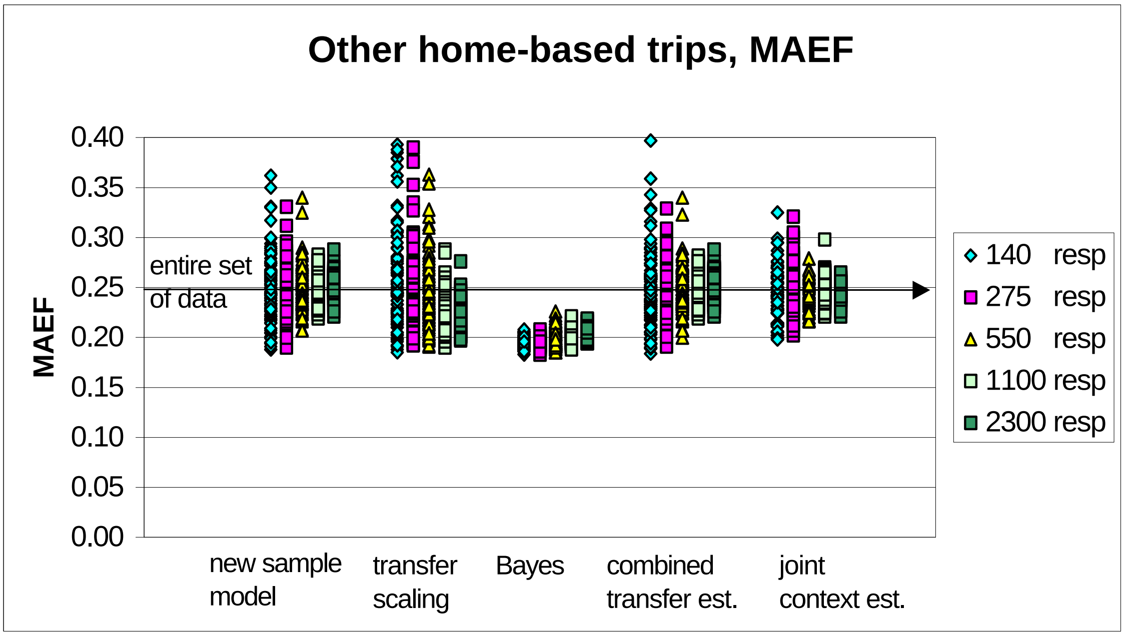


Figure 55: The variation of MAEF-index for other home-based trips.

Figure 55 shows that, as was the case with home-based work trips, most methods give quite similar results. By being highly precise, the Bayesian method differs from the other methods in giving a smaller MAEF-error. It can also be seen that the increase in the sample size does not necessarily improve the forecast prediction accuracy. Thus, sometimes the forecast based on a small sample is better than the forecast based on the entire set of data. The possible reasons for this might be the following:

- The models are applied to the weighted mobility survey data for all possible origin-destination pairs, not to the data for which they are estimated. Thus the model coefficients are concentrated around the values based on the entire set of data, not around the external observed values. Thus, the MAEF describes the effect of sampling error, if the MAEF for entire set of data would be perfect, but not the sample size effect with respect to the observed data.

- The iteration of alternative-specific coefficients correct the original mode shares to represent the mode shares based on the weighted mobility survey.

- The lack of alternative-specific coefficients in destination choice level. When comparing the rho-squared $\rho^2(0)$ and $\rho^2(c)$ the difference is very small indicating that the alternative-specific constants are in fact, the best determining variables and the level-of service and socio-economic attributes only explain a small part of the travel behaviour. In destination choice level when there are 118 different alternatives the problem is worse than in the mode choice level.

The results relating to alternative specifications for using joint context estimation

The results relating to different ways of combining common and data-specific variables by using the joint context estimation are presented in Appendix N. The best combinations are presented in Table 27. As with home-based trips, the best combination is chosen based on the sum of Mean Deviations (Equation 36) for the three most important transport policy coefficients (travel time, cost and cars per household). The coefficient for “number of transfers” was always estimated as common in this study.

The choice between common and data-specific variables in the destination choice level is not made by considering the absolute values of size coeffients but by making certain that the relationship between the coefficients of size variables are as accurate as possible. This is because, in this case, the relationship is more important than the absolute values.

Table 27: The best combinations of data-specific variables based on the MD in joint context estimation; other home-based trips.

SAMPLE SIZE	THE BEST COMBINATION OF DATA-SPECIFIC VARIABLES
	mode choice
140	Distance 0-5, (model B)
275	Distance 0-5, (model B)
550	Distance 0-5, distance 5, (model C)
1,100	Distance 0-5, distance 5, (model C)
2,300	Distance 0-5, distance 5, (model C)
4,675	Distance 0-5, distance 5, cost/income (model D)
	destination choice
140	Retail employment, service employment (model B)
275	Retail employment, service employment (model B)
550	Retail employment, service employment (model B)
1,100	Retail employment, service employment (model B)
2,300	Retail employment, service employment (model B)
4,675	Retail employment, service employment (model B)

Figures 56 and 57 present the sensitivity analysis for different combinations of data-specific and common variable.

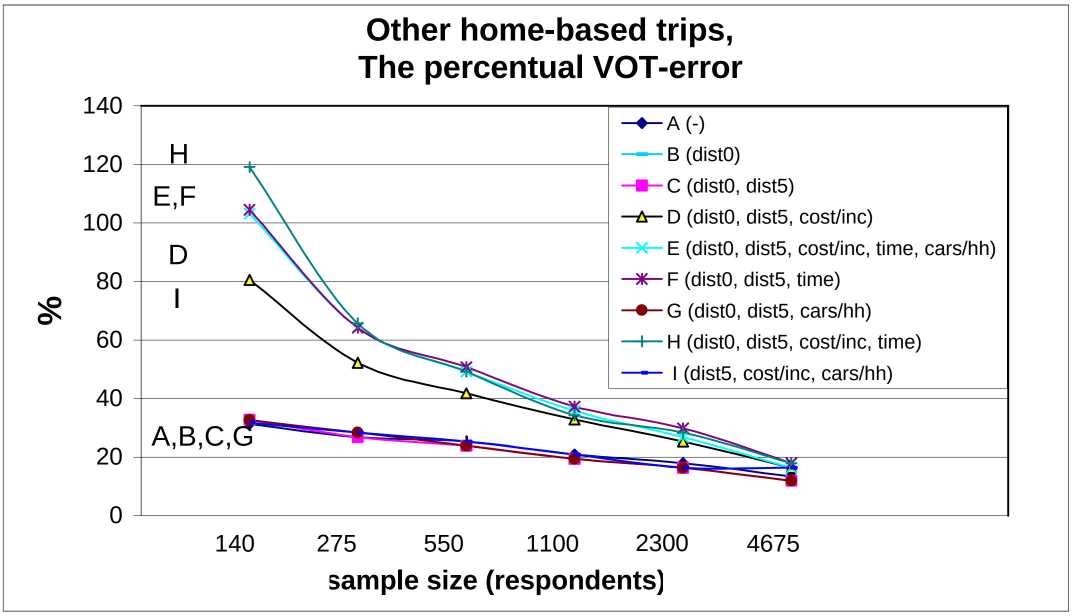


Figure 56: The comparison of the error in the ratio of the time and cost coefficients (VOT-error) when using different combinations of joint context estimation; other home-based trips (the symbols A to I refer to data-specific variables in the models).

The examination of VOTs shows that the differences between the combinations becomes quite large when using the smallest sample sizes, but they will be smaller, when the sample size increases. When using the smallest sample sizes, the combinations A, B, C and G were clearly better than the other combinations, in which the data-specific variables were used for time or cost/income variables. When using the largest sample size the choice of combination is only of minor importance.

In 4 cases out of 6 (6 sample sizes), the best combination based on the VOT-criteria differed from that based on the MD-criteria. However, the difference between the chosen and the best combination was in these cases very small.

When considering the RSEE, the differences of combinations are slightly smaller than with the VOT. In this case only in 4 cases out of the 12, the best combination is achieved when using the RSEE-test. On the other hand, the difference between the best and the chosen combination was always negligible. Although the transfer bias is small in both cases (when considering the increase of ptr time or car costs), the examination of the RSEE gives clear evidence that the use of data-specific variables is useful when predicting changes in the transportation system.

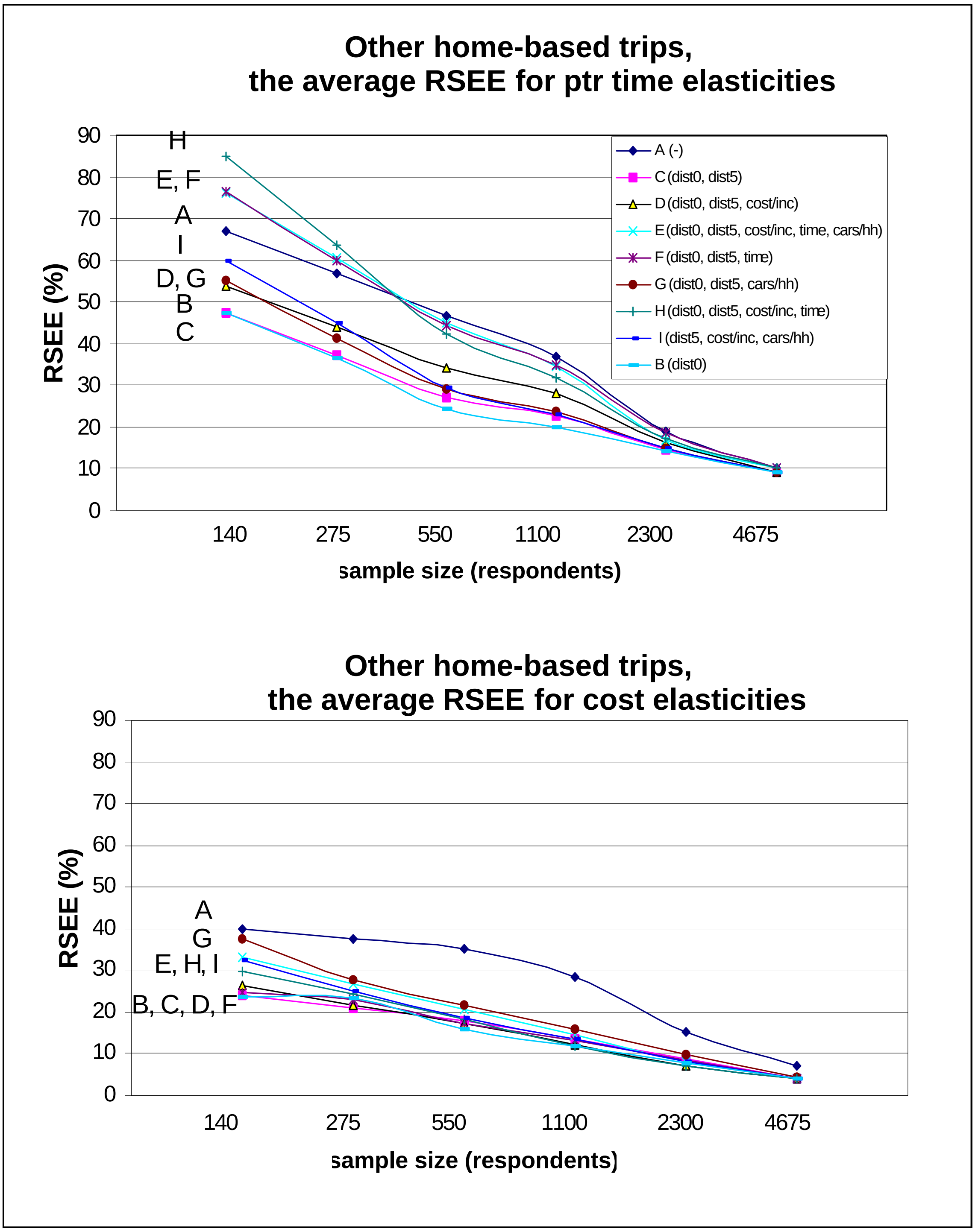


Figure 57: The comparison of the RSEE-error when using different combinations of joint context estimation; other home-based trips (the symbols A to I refer to data-specific variables in the models).

Comparison of the test results

Table 28 presents the success rate for different test measures. The success rate is defined as the percentage of cases when the method was the best according to each test measure.

Table 28: The success rate (%) of the transfer methods by the test measure (subscript t and c refer to time and cost)

					140 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	12	11	17	17	4	3	24	6	1	21
transfer scaling	12	21	10	12	25	2	17	9	1	18
Bayes	4	24	43	6	40	16	7	13	17	8
combined tr. est	5	1	3	8	11	10	21	18	10	24
joint context est	67	43	27	57	20	69	31	54	71	29
					275 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	17	12	14	27	7	8	10	5	7	13
transfer scaling	5	20	16	6	21	3	28	13	0	29
Bayes	0	12	42	2	29	3	5	5	3	7
combined tr. est	13	16	5	9	5	23	12	20	24	6
joint context est	65	40	23	56	38	63	45	57	66	45
					550 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	20	19	12	28	14	7	17	5	9	16
transfer scaling	8	15	10	12	20	1	25	8	1	26
Bayes	0	5	50	0	23	0	8	2	0	7
combined tr. est	12	18	9	19	5	41	17	28	40	17
joint context es	60	43	19	41	38	51	33	57	50	34
					1100 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	29	20	6	30	16	4	15	1	6	16
transfer scaling	5	7	16	10	20	0	16	1	0	15
Bayes	0	2	50	0	20	0	8	0	0	9
combined tr. est	11	22	9	18	11	67	12	44	62	12
joint context es	55	49	19	42	33	29	49	54	32	48
					2300 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	22	19	16	38	14	0	18	15	0	19
transfer scaling	2	5	5	0	27	0	6	4	1	5
Bayes	1	0	46	0	20	0	15	2	0	14
combined tr. est	17	23	20	13	14	86	15	30	86	14
joint context es	58	53	13	49	25	14	46	49	13	48
					4675 respondents					
	VOT	RSEEFt	RSEEFc	RSEEt	RSEEc	TTS mode	TTS dest	TTS total	TI mode	TI dest
new model	24	26	20	25	18	0	45	19	0	46
transfer scaling	0	0	6	10	19	0	0	0	0	0
Bayes	0	0	40	7	20	0	16	0	0	15
combined tr. est	24	22	20	18	16	95	11	37	94	11
joint context es	52	52	14	40	27	5	28	44	6	28

The main findings relating to the characteristics of the different tests are:

- Regardless of the sample size, the joint context estimation was the best method in 40 to 70 percent of cases when testing VOT, RSEE or RSEEF.

- TTS was the most sensitive to the sample size (especially at the mode choice level). That is, for small sample sizes joint context estimation was the best method, but when the sample size was increased this test preferred combined transfer estimation.

- Joint context estimation was usually the best method for mode choice, whereas the distribution of the best method was more uniform for the destination choice.

Table 29 presents the percentage of cases in which two different tests give similar recommendations for the best method.

Table 29: The number of cases in which each test-pair recommend the same transfer method when considering other home-based trips.

	sample size (respondents)					
	140	275	550	1100	2300	4675
VOT*/TTS_{total}	58	45	43	18	25	15
TTS_{mode}/TI_{mode}	95	91	95	95	95	96
VOT/RSEEFt	42	50	58	65	78	85
VOT/RSEEFc	26	28	30	29	33	32
VOT/RSEEt	64	55	47	61	64	63
VOT/RSEEc	19	20	22	18	21	20
RSEEFc/RSEEc	43	44	45	52	41	43
RSEEFt/RSEEt	51	57	61	81	87	93

Table 29 shows that the results based on the TTS and TI are nearly identical. The relationship between the RSEEF and the RSEE depends on the transfer bias. In principle, the RSEEF and the RSEE measure the same thing. However, the RSEEF differs from the RSEE by including the effect of destination choice models and the iteration of alternative-specific constants. In the case of cost-elasticities, the transfer bias was small giving a good prediction performance for the Bayesian method. The goodness of the Bayesian method was better still, when the alternative-specific constants were corrected via the iteration process. This effect is larger in the case of the Bayesian method than for the other transfer methods, which explains the differences between the RSEE and the RSEEF. The effect of transfer bias can also be seen when comparing the VOT to the RSEE or the RSEEF. When comparing the VOT and the TTS, it was noteworthy that the increase in sample size decreased the similarity of the test results of these two tests.

6.3 Conclusions

The main purpose of this chapter was to investigate the applicability of different transfer methods as a function of transfer bias and the preciseness of model coefficients, and to further examine, if the joint context estimation method could be developed to control the effect of transfer bias more efficiently than previously.

Generally, the smaller the transfer bias and the more precise the coefficients are, the better the transferred models performed. On the other hand, the research indicated that although the coefficients in the HMA and in the Turku region differed quite a lot, the model transfer performed usually better than the new application context models estimated by using the same sample size.

The joint context estimation proved to be the best method. This is greatly due to the utility scaling components which make it possible to take areal differences into account. The results were still improved given that the imprecise variables (such as “number of transfers”, “travel time” and “cars per household”) were estimated as common and the precise variables as data-specific. That is, the estimation context was emphasized for the imprecise variables and the new application context for precise variables.

The combined transfer estimation and the models based on the small samples were highly sensitive to the sample size. The coefficients based on these two methods were rather equal with each other. This is largely the result of the dominance of the transfer scaling component of the combined transfer procedure which effectively results in a procedure corresponding to a simple re-estimation of the model using the small application context dataset.

Transfer scaling also did not yield good results. The problem is that for example, in our case, a common scale factor was used for travel time and the number of transfers when considering other home-based trips. If the ratio of the coefficients of these variables does not remain the same in the estimation and application context, the models ability to predict the effects of changes in the transportation system may become rather weak. Both of these coefficients are normally quite imprecise, thus the risk of having biased coefficients is high.

As a conclusion, Table 30 presents the results listed in order of quality. The joint context estimation performed well especially if the transfer bias was large and only some of the coefficients were precise. However, in situations of small transfer bias, the Bayesian method yielded good results as well. However, due to its inability to explicitly take account of the transfer bias, it was often biased. The Bayesian approach emphasizes the coefficients with respect to the inverse of the variances of each coefficient. This means that the coefficients based on the estimation context data will be emphasized too strongly in the model transfer, since the number of observations is larger in this context. The Bayesian method also suffers from a repeated measurement problem when the normal Finnish data is used. Repeated measurements imply that there will be correlation between the answers provided by the same individual, and therefore in the error terms in the utility function. This means, that when estimating the new sample models, the variation of model coefficients represents the real variation, but due to the incorrect standard deviations calculated by estimation program, the estimated coefficients based on the Bayesian method or combined transfer estimation are no longer correct. This bias is investigated more precisely in Appendix H. Appendix H shows that the Bayesian estimates especially may be biased due to the correlations between the answers.

In the case of large transfer bias and imprecise coefficients, there is actually not very much to do. However, in this scenario joint context estimation is seen as the best (safest) method.

Table 30: The summary of the order of quality of transfer methods according to the transfer bias and precision of model parameters

	IMPRECISE COEFFICIENTS	PRECISE COEFFICIENTS
Small transfer bias	Bayes	Joint context estimation
	Joint context estimation	Transfer scaling
	Transfer scaling	Combined transfer estimation
Large transfer bias	(Joint context estimation)	Joint context estimation
		Combined transfer estimation
		Transfer scaling

The differences between the methods were larger for other home-based trips than for home-based work trips. This was especially so with regard to VOT and elasticity tests for travel time. One reason for this might be that the transfer scaling method performed less well for other home based trips than for home-based trips due to the common scale factor used for the travel time and the number of transfers in modelling other home based trips. Another noteworthy point is, that the repeated measurement problem was larger for home-based work trips than for other home-based trips (see Section 5.3.3.3). As a consequence of that the real variation of coefficients in home-based work trips was larger than in other home-based trips.

The research also indicated that the effects that are produced by sample size and many other factors connected to the modelling is often greater than that produced by the differences between the best methods.

Table 31 presents the sensitivity analysis of the sample size required in model transfer. The sample size requirements were evaluated in terms of two differerent hypothetical error limits using the tests already applied in Chapter 5. TTS is not considered due to its tendency always reject the hypothesis of the equality of the model parameters in two contexts.

The results indicate that the mode and destination choice models made in the HMA were well transferable if the decrease of 15 percent of the best possible test measure was accepted. However, if only the 5 percent error was allowed, the advance of model transfer was smaller. The sample size requirement was 2,500 respondents for home-based work trips when 5 percent error limit was used and 1,400 respondents when 15 percent error limit was used. The corresponding values for other home-based trips were 4,450 and 2,400 respondents, respectively. In almost all cases the best transferred model yielded better results than the corresponding new sample model

Table 31: The sensitivity analysis of the sample size requirements in model transfer.

		the results compared to the value, which is 5 percent below the maximum				the results compared to the value, which is 15 percent below the maximum		
	trip group	required sample size for new sample model based on 25 % error limit	the percent of models for which the error of test value is below 25 %	required sample size for the best transfer method*	required sample size for joint context estima-tion**	the percent of models for which the error of test value is below 25 %	required sample size for the best transfer method*	required sample size for joint context estima-tion**
VOT	HBW	3,800	80	900	3,100	71	800	2,000
	OHB	4,450	84	-	-	75	2,300	2,300
RSEEF ptr time	HBW	2,800	90	2,400	2,400	80	400	1,700
	OHB	4,000	90	-	-	81	2,400	2,400
RSEEF car costs	HBW	3,200	48	2,500	2,500	43	1,400	1,400
	OHB	1,200	94	(400)	1,200	84	0[x]	250

* based on the test-value presented in previous column

** the sample size which is needed to get the same test value as required sample size produces for new sample models

\- can not be achieved when using model transfer

x the HMA model explains this test measure good enough

When comparing the different test measures it can be said, that joint context estimation was the best approach in 40 to 70 percent of cases when testing RSEE or RSEEF and also when testing VOT for other home-based trips. The Bayesian method was the best method when testing VOT for home-based work trips. TTS gave partially contradictory results with respect the other methods. It was also the most sensitive for the sample size. The problem when using TTS is, that two models having totally different coefficients may have the same TTS. Although these two models can be similar in some ways, their ability to predict the effect of changes in a transportation system is nevertheless different. It was also shown, that TTS test has strong tendency to reject transferebility. On the other hand Transfer Indexes (TI) for transferred models were in almost all cases in excess of 80 percent, suggesting that the transferred models provide a significant component of the information obtained from application context model.

The greatest problem in model transfer is the large range of fluctuation caused by imprecise coefficients. The more new application context parameters are emphasized (e.g. by using data-specific coefficients), the greater the confidence interval of the coefficients and the larger the risk of arriving at biased coefficients is. So, when choosing the model definitions, we are trying to find a model, which does not exceed the predefined error limit, but which produces better results more often than the new sample model. Next, a short interpretation of a real model transfer situation and the different choices which are to be made are presented:

- The first step in model transfer is data collection. The data collection methods should be similar in estimation and application contexts. In addition, it is important to ensure, that the sample size is large enough to define mode shares accurately (if there is not any other statistical source).

- The network description should be made similarly as well. This means, for example, that similar demand functions in traffic assignment are used in both contexts.

- In principle, the same variables used in the estimation context should be used in the application context as well. However, if both datasets are available, the estimation context models can be re-estimated and different variable combinations can be tested. When using joint context estimation, it is also possible to add additional attributes which are not represented in the estimation context model.

- The model specifications should be made similarly. That is, the variables, which are differently defined, are totally different variables and thus are not to be transferred. The sole exception is, if data-specific coefficients are used in a joint context estimation.

- The use of alternative-specific dummy-variables in the destination choice level is not normally recommended due to the difficulty of forecasting the importance of these kind of attributes. However, if they are needed, they can be applied by estimating new sample models at the destination choice level and by using some other transfer method at the mode choice level.

- The choice of transfer method depends on the kind of data available. If both the estimation and application context data are available, the best transfer method is joint context estimation. When using the Bayesian method or the combined transfer estimation, the parameter vectors and covariance-matrices are needed in both contexts. Note, if the Bayesian approach is used, the repeated measurements should be ignored (e.g. drawing randomly only one observation per respondent from the data). Transfer scaling is the easiest way; only the coefficients of estimation context model and the application context data are needed.

- When using joint context estimation, the best choice between common and data-specific variables depends on the transfer bias and the preciseness of the coefficients. It can be stated that the more common variables the model includes the smaller the variance of all the coefficients is. On the other hand, the model which only includes common variables, strongly emphasizes the estimation context model parameters. This is problematic when the transfer bias is large. Thus, the common variables are most suitable for situations in which the observed variance of the application context coefficients is rather high and the transfer bias is known to be small.

In Chapter 5 we defined the CV measure as being able to compare the preciseness of different coefficients analytically, and in such a way as to be able to evaluate which variables should be estimated as common and which as data-specific, when using joint context estimation. The CV(A) value is defined as the ratio of the standard error of a new sample model coefficient divided by the coefficient estimated from the whole estimation context data set. The estimation context parameters must be used as the denominator, because in a real transfer situation the application context parameters, based on the entire set of data, are unknown. The initial premise is that the greater the CV(A)-values, the more imprecise the model parameters are.

The CV(A)-values for empirical cases are presented in Figure 58. The variation of CV(A) is considered in appendices P and Q. We have not defined any absolute value for the acceptable CV(A), because the sample size requirements depend on the purpose for which the models are to be estimated. However, CV(A)-values can be used as a tool for selecting the variables which are to be estimated as data-specific. One can formulate a hypothetical confidence limit to describe the minimum preciseness of the coefficients by which the data-specific coefficients can be used. That is, the data-specific coefficients should be used for coefficients which are under the confidence limit and common coefficients should be used for other variables in the model. In this case, the coefficients that exceed the limit will be fixed to the well-known estimation context situation and the data-specific coefficients will only be used for precise coefficients. If all the coefficients exceed the confidence limit, the model should be estimated using only common coefficients. This is the way this process is generally done.

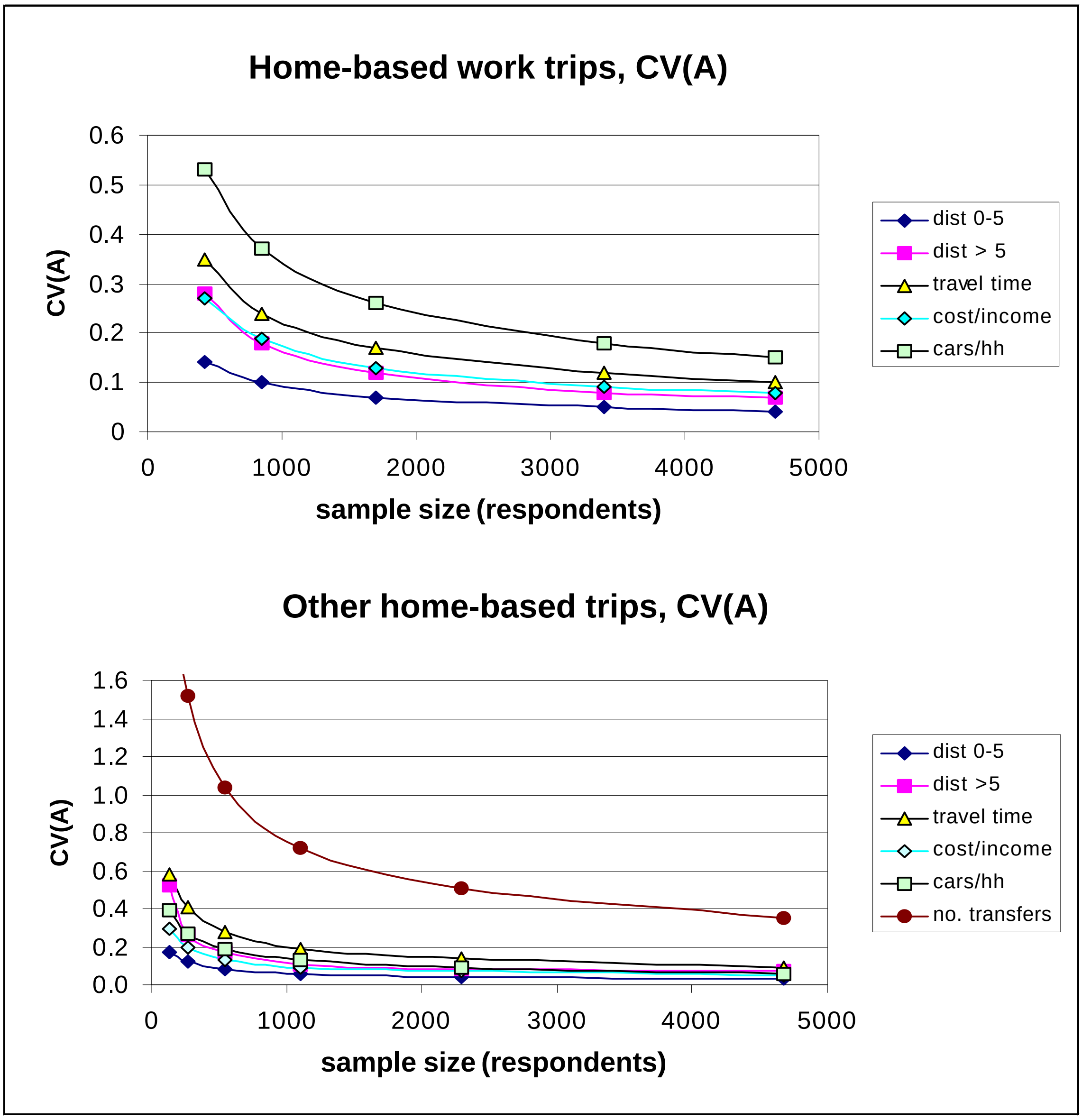


Figure 58: The mean CV(A)-values estimated for models based on the small samples.

7. DISCUSSION

The main purpose of the study was to compare alternative methods of spatial transfer as a function of sample size, and identify the factors affecting the model quality and the preciseness of model parameters. As a large part of the difference in coefficients and predictions is due to the discrepancies in the formulation of the initial data, or random variation, and only partly due to the real differences in the estimation and application context, the main objective of this study was to investigate the relationship between the transfer bias (difference between the estimation and application context parameter estimates) and the impreciseness caused by the sample size. In addition, different test measures for studying model transferability were compared and the applicability of the traditional statistical tests, with respect to those based on the prediction accuracy of sample enumeration tests and forecasts, were assessed.

The approach of primarily studying model transfer as a function of sample size rather than transfer bias proved to be successful. This is because in a real transfer situation the preciseness of the model coefficients can be roughly evaluated, but it is not known whether the transfer bias is caused by the random variation in the estimated model parameters, or whether it actually presents real differences in travel behaviour.

The use of bootstrap *(Efron and Tibshirani 1993)* when resampling observations made it possible to effectively illustrate the variability in the test measures that arise as a function of decreasing sample size. It was shown that the results based on individual samples may greatly differ from each other, giving a totally different model transfer performance. It was also indicated that the effect that is produced by sample size and many other factors connected to the modelling are often greater than that produced by the differences between the best transfer methods.

Data quality

The first step in modelling is data collection. The research highlighted the fact that telephone interview is usually a more reliable method of collecting information than the postal questionnaire. The walk and bicycle trips, in particular, were better reported in telephone interviews, and also people without trips were better captured in the telephone survey. It seemed to be that the representativeness of mode shares is more dependent on the data collection method than sample size even if the sample size is a very important factor as well. We also obtained evidence that the data gathering method can have an effect in the mode- and destination choice level, thereby suggesting that some of the differences between the coefficients in the estimation and the application context may be caused by the data gathering method and not just by real differences in local conditions in the HMA and in the Turku region. Although these effects are very important in real transfer situation, they do not affect the conclusions drawn from the model transferability, because studying the real differences was not the focus of this thesis but rather the main purpose was to study model transferability by using models which include different levels of impreciseness and transfer bias.

Model Transferability

On the whole, the comparison of different transfer methods by using many different kinds of tests highlighted effectively the differences between the methods. The goodness of the models is not an unambiguous issue but is dependent on the purpose for which model is estimated. In most cases, the joint context estimation proved to be the best transfer method. As *Badoe (1994)* has concluded this is largely due to the utility scaling components. This raises the potential for a "universal choice model", in which utility attribute parameters are common to all the contexts, and the scales and constants are expressed as functions of an urban structure. In this study it was shown that the results could still be improved further when the imprecise variables (such as "number of transfers", "travel time" and "cars per household") are estimated as common and the precise variables as data-specific.

The combined transfer estimation and the new sample models were highly sensitive to the sample size. The coefficients based on these two methods were very similar with each other. This is largely the result of the dominance of the transfer scaling component of the combined transfer procedure which effectively results in a procedure corresponding to a simple re-estimation of the model using the small application context dataset.

Transfer scaling did not yield good results either. The problem is that for example, in our case, a common scale factor was used for travel time and the number of transfers when considering other home-based trips. If the ratio of the coefficients of these variables does not remain the same in the estimation and application context, the models ability to predict the effects of changes in the transportation system may become rather weak. Both of these coefficients are normally quite imprecise, thus the risk of having biased coefficients is high.

The sample size needed in model transfer is highly dependent on the accuracy required for each step in the forecast system. No definite advice can be given as to the required precision of the system, but the need for the accuracy depends on the purpose for which models are used. In this research we did not consider, when the models are precise enough, but the results were mainly compared to the results based on the entire set of data, which was assumed to represent the best achievable situation. At the mode and destination choice level two different accuracy criteria were tested: 5 and 15 percent decreases from the best possible result. When considering the 5 percent error limit in model transfer, we were not essentially able to decrease the sample size requirement. However, when considering the 15 percent error limit, the sample size requirement was only half the number of observations needed to estimate the best possible model.

We also obtained evidence that although the mode and destination choice models could be transferred using a quite small data set, an appropriate definition of mode shares and also defining trip generation rates correctly requires much more data. In principle, the process of defining mode shares and trip generation rates is independent of mode and destination choice modelling. The mode shares and trip generation rates are based on the cross-tabulation of mobility survey data, and mode shares are only used in the forecast process to correct alternative-specific constants to accurately present the real weighted mode shares in the area. The 1 percent absolute error limit that we defined to be the maximum allowable deviation from the correct mode share was quite strict. However, in real modelling the 1 percent deviation is regarded to be definitely the maximum error to be allowed. Due to the large sample size needed to define mode shares accurately, the use of separate OD-surveys, instead of a mobility survey, would be preferable for this purpose.

The use of empirical data highlighted problems which are not connected to the transfer method itself, but greatly affect the model transfer. Of these, the most important was the repeated measurement issue. Repeated measurements imply that there will be a correlation between the answers provided by the same individual, and therefore in the error terms in the utility function. Consequently, the estimated standard deviations are underestimated and t-values based on these standard deviations are incorrect because they are too high. However, repeated measurements do not affect the parameter estimates of new sample models, the transfer scaling or joint context estimation, but they do affect the results based on the Bayesian method and combined transfer estimation. The correlation between the observations means that when using the Bayesian method in particular, the results may be strongly biased.

The comparison of different tests

Model transferability has traditionally been evaluated on the basis of how well transferred models reproduce existing behaviour rather than on their ability to adequately forecast changes in travel demand. The comparison of VOT and elasticities to the corresponding TTS-values indicated, however, that statistical tests are not able to evaluate the goodness of transferred models with a high enough degree of versatility. For example, two models that have totally different values for their coefficients may have the same likelihood-value and hence the same TTS. Consequently, their ability to predict the effect of changes in a transportation system may differ greatly. Hence, because the main purpose of the policy models is to predict the changes in the transportation system, as was the case for example in the HMA and in the Turku region, the conclusions based on the RSEEF and VOT should be emphasized.

Another related problem both in calculating the TTS as well as in calculating elasticities is that in modelling the scale parameter μ in the choice probability function (Equation 3) is fixed to one. This means that the absolute level of the coefficients is actually dependent on the real value of μ. Consequently, the ratio of the model parameters to each other can be estimated reliably but the absolute level of the parameters is unknown. Consequently the comparison of model parameters on an equal basis, as has been the case in many transferability studies, is not justified. Another problem is that in spite of the fact that the absolute level of the model parameters is unknown, the elasticities are highly dependent on the absolute level of the coefficients, which can be regarded as a weakness of the logit model formulation.

Conclusions

The research confirmed the general view that model transfer is useful in situations in which a large data set can not be collected. Generally, the smaller the transfer bias and the more precise the coefficients are, the better the transferred models perform. On the other hand, it was shown that due to the difficulty in defining mode shares accurately based on small samples, spatial transfer cannot be recommended as a primary method, in spite of the fact that mode and destination choice models can be transferred quite reliably. Consequently, the largest utility for model transfer can be achieved when it is applied to model updating rather than to spatial transfer. Model updating using only a small set of new application context data makes it possible to update models more frequently and economically than used be the case.

Generally, spatial transfer tends to succeed better when there is a significant number of equal transport conditions in the estimation and application context areas. This means that depending on the mode shares, two or three different areal models should be used as the basis of modelling in Finland.

In future, it will be important to improve the quality of the data and the travel demand models. For example, research into the use of individual variables, different modelling techniques and developing totally new modelling approaches at the destination and route choice level would be of benefit. This is because many factors connected to the modelling itself are greater than those produced by the differences between the best transfer methods.

Improving the quality of the model also helps to develope the ultimate goal whereby the data-specific coefficients can be used as much as is possible, to improve model transferability. This is especially important in cases where the transfer bias is large.

8. REFERENCES

Abdelwahab W.M., (1991). Transferability in intercity disaggregate mode choice models in Canada. Canadian Journal of Civil Engineering, No. 18:1991, pp. 20 – 26.

Algers S., Colliander J., Widdlert S., (1987). Logitmodellen - Användbarhet och generaliserbarhet. Byggforskningsrådet, rapport R30:1987, pp. 184-213.

Algers S., Lindqvist J., Tretvik T., Widlert S., (1994). Överföring av trafikmodeller i Norden. Nordisk Ministerrådet. Tema Nord 1994:552, 137 p.

Amemiya T., (1978). On a two-step estimation of a multivariate logit model. Journal of Econometrics 8:1978, pp. 13-21.

Armstrong, P., Garrido, R., Ortúzar, J. de D. (2001). Confidence intervals to bound the value of time. Transportation Research Part E 37, 143-161.

Atherton T.J., Ben-Akiva M.E., (1976). Transferability and updating of disaggregate travel demand models. Transportation Research Record, No. 610, Transportation Research Board. National Research Council. Washington, D.C., 1976, pp. 12 - 18.

Badoe D.A., (1994). An Investigation into the long range transferability of work-trip discrete mode choice models. A thesis in conformity with the requirements for the degree of Doctor of Philosophy Graduate, Department of Civil Engineering, University of Toronto, 273 p.

Badoe D.A., Miller E.J., (1995a). Analysis of temporal transferability of disaggregate work trip mode choice models. Transportation Research Record, No. 1493, Transportation Research Board. National Research Council. Washington, D.C., 1995, pp. 1 - 11.

Badoe D.A., Miller E.J., (1995b). Comparison of alternative methods for updating disaggregate logit mode choice models. Transportation Research Record, No. 1493, Transportation Research Board. National Research Council. Washington, D.C., 1995, pp. 90 - 100.

Badoe D.A., Miller E.J., (1998). Modelling mode choice with data from two independent cross-sectional surveys: an investigation. Transportation Planning and Technology, Vol.21:1998, pp. 235-261.

Ben-Akiva M.E., (1973). Structure of passanger travel demand models. PhD dissertation, Department of Civil Engineering, MIT, Cambridge, Massachusetts.

Ben-Akiva M.E., (1974). Structure of passanger travel demand models. Transportation research Record, No. 526, Transportation Research Board. National Research Council. Washington, D.C., 1974, pp. 26-42.

Ben-Akiva M.E., (1981). Issues in transferring and updating travel behavior models, In: Stopher P.R., Meyburg A.H. ja Brög W. (eds), New Horizons in Travel Behaviour Research, D.C. Heath and Co. Lexington, Massachusetts, pp. 665-686.

Ben-Akiva M.E., Bolduc D., (1987). Approaches to model transferability and updating: the combined transfer estimator. Transportation Research Record No. 1139, Transportation Research Board. National Research Council. Washington, D.C., 1987, pp. 1 - 7.

Ben-Akiva M.E., Bolduc D., Bradley, M., (1993). Estimation of travel choice models with randomly distributed values of time. Transportation Research Record 1413, pp. 88-97.

Ben-Akiva M.E., Lerman S., (1985). Discrete choice analysis - theory and application to travel demand. M.I.T. Press, Cambridge, Massachusetts, 384 p.

Ben-Akiva M.E., Morikawa, (1990). Estimation of travel demand models from multiple data sources. Proceedings 11th International Symposium on Transportation and Traffic Theory, Yokohama, 16 p.

Bergström P., (1999). Bootstrap methods and applications in econometrics-a brief survey. Department of Economics Uppsala University, Working paper 1999:2. 27 p.

Bradley M.A., Daly A.J., (1991). Estimation of logit choice models using mixed stated preference and revealed preference information. The 6th International Conference on Travel Behaviour, Quebec. 16 p.

Brand D., Cheslow M.D., (1981). Spatial, temporal, and cultural transferability of travel-choice models. In: Stopher P.R., Meyburg A.H. ja Brög W. (eds), New Horizons in Travel Behaviour Research, D.C. Heath and Co. Lexington, Massachusetts, pp. 687 - 691.

Brundell-Freij K., (1995). The logit model applied to the modal split of regional commuting, a case study and Monte Carlo simulations. PhD, Tekniska högskolan i Lund, bulletin 123, 127 p.

Brög W., Meyburg A.H., (1981). Consideration of nonresponse effects in large-scale mobility surveys. Transportation Research Record 807. Transportation Research Board. National Research Council. Washington, D.C., pp. 39 - 45.

Chester G.W., (1983). The impact of market segmentation on the transferability of discrete choice models. PhD dissertation, Evanston, Illinois, August 1983, 243 p.

Cirillo C., Daly A., Lindveld K., (2000). Eliminating bias due to the repeated measurements problem in SP data. In: Ortúzar, J de D., Stated preference modelling techniques, A compination of major papers selected from PTRC's vast bank of meeting and conference material, pp. 137-154.

Dateline (2003). Design and Application of a travel survey for European long-distance trips. European Commission under the Competitive and Sustainable Growth Programme of the Fifth Framework Programme. http://www.ncl.ac.uk/dateline (viitattu 24.3.2003).

Dehghani Y., Talvitie A., (1983). Forecasting accuracy and transferability of simple work trip mode choice models, a paper presented at PTRC Annual Meeting England, (july 1981).

Efron B., Tibshirani R.J., (1993). An introduction to the Bootstrap, Chapman and Hall, New York, 430 p.

Elmi A.M., Badoe D.A., Miller E. J., (1999). Transferability analysis of work-trip distribution models. Prepared for presentation at the 1999 TRB Annual Meeting, Washington, D.C., 1999, 19 p.

Galbraith R. A, Hensher D.A., (1982). Intra-Metropolitan transferability of mode choice models, Journal of Transport Economics and Policy, XVI(1), pp. 7-29.

Gunn H., Pol H., (1986). Model transferability: The potential for increasing cost-effectiveness. Behavioural Research for Transport Policy. pp. 217 - 243.

Hague Consulting Group, (1990). An investigation of alternative methods to update forecasting models: Including the use of stated preference data, Hague.

Hague Consulting Group, (1992). ALOGIT, users' guide, version 3.2:September 1992, The Netherlands.

Hallipelto H., (1993). Strukturoitu logittimalli ja otosenumeraatio liikenne-ennusteissa. Pääkaupunkiseudun yhteistyövaltuuskunta YTV, Pääkaupunkiseudun julkaisusarja C 1993:22. Helsinki. 90 p.

Hansen S., (1981). In favor of cross-cultural transferability of travel-demand models. In: Stopher P.R., Meyburg A.H. and Brög W. (eds), New Horizons in Travel Behaviour Research, D.C. Heath and Co., Lexington, Massachusetts, pp. 637 - 651.

Hensher D. A., Johnson L.W., (1991), Applied discrete-choice modelling. Croom-Helm, London, 468 p.

Horowitz J.L., (1981). Sources of Error and uncertainty in behavioral travel-demand models. In: Stopher P.R., Meyburg A.H. ja Brög W. (eds), New Horizons in Travel Behaviour Research, D.C. Heath and Co., Lexington, Massachusetts, pp. 543-557.

Horowitz J.L., (1982). Specification tests for probabilistic choice models. Transportation Research A, Vol. 16A, No. 5-6, pp. 383-394.

Kaartokallio M., (1997). Liikkumistottumukset ja niiden muutokset pääkaupunkiseudulla. Pääkaupunkiseudun yhteistyövaltuuskunta YTV, Pääkaupunkiseudun julkaisusarja C 1997:7. Helsinki. 129 p.

Kalfs N., Meurs H., Saris W., (2000). Transport surveys, raising the standard. Proceedings of an International Conference on Transport Survey Quality and Innovation, May 24-30, 1997, Grainau, Germany. Transportation Research Circular, No. E-C008, Transportation Research Board. National Research Council, Washington, D.C., II-E-pp. 1-22.

Karasmaa N., (1995). Pääkaupunkiseudun kulkutapamallien siirrettävyys Ouluun. Tielaitoksen selvityksiä 1995/40. Helsinki 1995. 82 p.

Karasmaa N., (1996a). Liikenne-ennustemallien päivitettävyys pääkaupunkiseudulla, esiselvitys. Pääkaupunkiseudun yhteistyövaltuuskunta YTV, Pääkaupunkiseudun julkaisusarja C 1996:8. Helsinki. 160 p.

Karasmaa N., (1996b). Liikenne-ennustemallien siirrettävyys. TKK:n rakennus- ja yhdyskuntatekniikan osaston liikennelaboratoriossa tehty lisensiaatintyö. Helsinki. 114 p.

Karasmaa N., Kurri J., Pursula M., (1997). Pääkaupunkiseudun liikennemallit 1995. Pääkaupunkiseudun yhteistyövaltuuskunta YTV, Pääkaupunkiseudun julkaisusarja C 1997:12. Helsinki. 74 p.

Karasmaa N., (1998). Liikenne-ennustemallien päivitettävyys pääkaupunkiseudulla, Pääkaupunkiseudun julkaisusarja B 1998:3. Helsinki. 157 p.

Karasmaa, N., (2000). The temporal transferability of the Helsinki Metropolitan Area mode choice models. 3rd KFB Research Conference, June 13-14, 2000. 24p.

Karasmaa, N., (2001). The spatial transferability of the Helsinki Metropolitan Area mode choice models. 9th WCTR in Soul, July 22-27, 2001.

Karasmaa N., Pursula M., (1995). Liikenne-ennustemallien alueellinen siirrettävyys, kirjallisuustutkimus. Tielaitoksen selvityksiä 1995/2. Helsinki. 100 p.

Karasmaa N., Pursula M., (1997). Empirical studies on the transferability of the Helsinki Metropolitan Area travel forecasting models. Transportation Research Record No.1607, Transportation Research Board. National Research Council, Washington, D.C., 1997. pp. 38-44.

Kivari M., Pastinen V., Pentilä T., Kurri J, Laakso J-M., Karasmaa N., Vesajoki T., Korhonen A., Leinonen R., Murto R., (2000). Suositukset alueellisten henkilöliikenne-tutkimusten yhtenäistämiseksi. Liikenneministeriön julkaisuja 28/2000. Helsinki. 144p.

Koppelman F.S., Chu C., (1983). Effect of sample size in disaggregate choice model estimation and prediction. Transportation Research Record No. 944, Transportation Research Board. National Research Council. Washington, D.C., 1983, pp. 60-69.

Koppelman F.S., Wilmot C.G., (1982). Transferability analysis of disaggregate choice models. Transportation Research Record, No. 895, Transportation Research Board. National Research Council. Washington, D.C., 1982, pp. 18 - 24.

Koppelman F.S., Kuah G-K., Wilmot C. G., (1985). Transfer model updating with disaggregate data. Transportation Reseach Record, No. 1037, Transportation Research Board. National Research Council. Washington D.C., 1985, pp.102-107.

Koppelman F.S., Pas E.I., (1986). Multidimensional choice model transferability, Transportation Research B. Vol. 20 B**,** No. 4 pp. 321-330.

Koppelman F.S., Rose G., (1983). Geographic transfer of travel choice models: evaluation and procedures, optimization and discrete choice in urban systems, edited by B.G. Hutchinson, P. Nijkamp and M. Batty, pp. 272-309.

Kurri J., Karasmaa N., (1999). Tutkimusmenetelmän vaikutus henkilöliikennetutkimuksen luotettavuuteen. Liikenneministeriön mietintöjä ja muistioita B 1/99. Helsinki. 66p.

Kurri J., Mikola J., Karasmaa N., (2001). Stated preference study of mode choice in the Helsinki metropolitan area. In: Pursula M., Niittymäki J. (eds), Mathematical methods on optimization in transportation system, Kluwe Academic Publishers, Dordrect, the Netherlands 2001, pp. 203-223.

Lerman S.R., (1984). Recent advances in disaggregate demand modelling. In: Florian M., (eds), Transportation planning models, proceedings of the course given at the International Center for Transportation Studies (ICTS), Amalfi, Italy-October 11-16,1982, Elsevier Science Publishers B.V., Amsterdam, the Netherlands, pp. 3-18.

Louviere J.J., (1981). Some comments on premature expectations regarding spatial, temporal and cultural transferability of travel-choice models. In: Stopher P.R., Meyburg A.H. and Brög W. (eds), New Horizons in Travel Behaviour Research, D.C. Heath and Co. Lexington, Massachusetts, pp. 653-663.

Luce R., Suppes P., (1965). Preference, utility and subjective probability. In: Luce R.D, Bush R.R, Galanter E. (eds), Handbook of mathematical psychology. Wiley, New York.

McCarthy G.M., (1982). Further evidence on the temporal stability of disaggregate travel demand models. Transportation Research B, Vol. 16B, No. 4, pp. 263-278.

McCoomb L.A., (1986). Analysis of the transferability of disaggregate demand models among ten Canadian cities, Transportation Forum, Vol 3-1: pp. 19-32.

McFadden, D., (1974). Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P., (eds.), Frontiers in Econometrics, Academic Press, New York.

McFadden, D., (1978). Modelling the choice of residential location. In: Karlquist A., Lundquist L., Snickars F., Weibull J.W., (eds.), Spatial interaction theory and planning models, North-Holland, Amsterdam, pp. 75-96.

Nielsen O.A., Jovicic (2003). The AKTA road pricing experiment in Copenhagen. 10th International Conference on Travel Behaiour Research in Lucerne 10.15. August 2003.

Ortúzar J. de D., Willumsen L.G., (1994). Modelling transport, John Wiley & Sons Ltd., Chichester, West Sussex, 439 p.

Ortúzar J. de D., Willumsen L.G., (2001). Modelling transport, John Wiley & Sons Ltd., Chichester, West Sussex, Third edition, 499 p.

Parody T., (1977). An analysis of disaggregate mode choice models in prediction. - Transportartation Research Record, No. 637, Transportation Research Board. National Research Council. Washington, D.C., pp. 51 - 57.

Pursula M., Kanner H., (1992). Traffic forecasting in the Helsinki metropolitan area transportation study 1988. Transportation Research Record No. 1357, Transportation Research Board, National Research Council, Washington, D.C., 1982, pp. 29-39.

Pursula M., Widlert S., (1990). Nordiska efterfrågemodeller, Översikt och projektförslag.

Rao C., (1973). Linear statistical inference and its applications 2. New York, John Wiley & Sons, 625 p.

Rich, J., (2002). Long-term Travel Demand Modelling. Ph.d -thesis. In press CTT/DTU.

Richardson A.J, Ampt E.S, Meyburg A.H., (1995). Survey methods for transport planning. Eucalyptus Press, Australia. 459 p.

Sermons M. W., (2000). Assessing the impact of variance scaling, weighting, and systematic taste variation on the transferability of residential location choice models. 79th Annual Meeting of the Transportation Research Board, Washington, D.C., 2000.

Smith M.E., (1979). Design of small-sample home-interview travel surveys. Transportation Research Record 701. Transportation Research Board, National Research Council, Washington, D.C., 1979, pp. 29-35.

Sørensen, M.V. (2003a). Discrete Choice Models Estimation of Passenger Traffic. Ph.d.-thesis. In press.CTT/DTU.

Sørensen, M.V. (2003b). MSL for mixed logit model estimation-on shape of distributions. European Transport Conference in France, October 8-10, 2003.

Stopher P.R., (1982). Small-sample home-interview travel surveys: Application and suggested modifications. Transportation Research Record No. 886, Transportation Research Board, National Research Council, Washington, D.C., 1982, pp. 41-47.

Stopher P.R., Meyburg A. H., (1979). Survey sampling and multivariate statistics for social scientists and engineers, D.C Heath and Co., Lexington Books, Lexington, Mass, 385 p.

Stopher P.R., (1991). Conventional Wisdom in building mode-choise models: Some fallacies and other myths, 33rd Annual Forum of the Transportation Research Forum, New Orleans, Louisiana, October 1991.

Talvitie A., (1981). Refinement and application of individual choice models in travel demand forecasting. In: A guide to the development and application of disaggregate mode choice models. State University of New York at Buffalo.

Talvitie A., (1983). Liikenteen ennustemenetelmät ja niiden käyttö. In: Liikenne-ennusteet ja edullisuusvertailut. RIL K11-1983. pp. 45-80.

Talvitie A.P, Kirshner D., (1978), Specification transferability and the effect of data outliers in modelling the choice of mode in urban travel. Transportation 7, No3, pp. 311-328.

Tardiff T.J., (1979), Specification analysis for quantal choice models, Transportation Science Vol. 13, No.3 pp. 179-190.

Thurstone L.L., (1927). A law of comparative judgement. Psychological Review 34.

Tilastokeskus, (1996). Suomen tilastollinen vuosikirja, Gummerus Kirjapaino Oy, Jyväskylä 1996. pp 54-55.

Tilastokeskus, (1998). Suomen tilastollinen vuosikirja, Karisto, Hämeenlinna 1998. pp 57-58.

Train K., (1978). A validation test of disaggregate mode choice model. Transportation Research, Vol. 12, No.3, pp. 167-174.

Train K., (1986). Qualitative choice analysis. Theory, econometrics, and an application to automobile demand. The MIT Press, Cambridge, Massachusetts. London, England, second printing 1990.

Tretvik T., (1989). Logitmodeller for transportplanlegging - Matematiska modeller for reiseatferd basert på kvalitativ valghandlingsteori (In Norwegian), Doctor ingeniøravhandling 1989:27, Norges Tekniske Høgskole, Trondheim, 262 p.

Turun kaupunki, Raision kaupunki, Naantalin kaupunki, Kaarinan kaupunki, Liedon kunta, Varsinais-Suomen liitto, Turun tiepiiri, Liikenneministeriö, Ympäristöministeriö, Ratahallintokeskus (1999a). Liikennetutkimukset ja mallit 1997, yhteenvetoraportti. Turku 1999, 45 p.

Turun kaupunki, Raision kaupunki, Naantalin kaupunki, Kaarinan kaupunki, Liedon kunta, Varsinais-Suomen liitto, Turun tiepiiri, Liikenneministeriö, Ympäristöministeriö, Ratahallintokeskus (1999b). Liikennetutkimukset ja mallit 1997, osaraportti 1: henkilöhaastattelut. Turku 1999.

Turun kaupunki, Raision kaupunki, Naantalin kaupunki, Kaarinan kaupunki, Liedon kunta, Varsinais-Suomen liitto, Turun tiepiiri, Liikenneministeriö, Ympäristöministeriö, Ratahallintokeskus (1999c). Liikennetutkimukset ja mallit 1997, osaraportti 2: kenttätutkimukset. Turku 1999.

Turun kaupunki, Raision kaupunki, Naantalin kaupunki, Kaarinan kaupunki, Liedon kunta, Varsinais-Suomen liitto, Turun tiepiiri, Liikenneministeriö, Ympäristöministeriö, Ratahallintokeskus (1999d). Liikennetutkimukset ja mallit 1997, osaraportti 3: sisäisen henkilöliikenteen mallit. Turku 1999, 69 p.

Vasama P., Vartia Y., (1979). Johdatus tilastotieteeseen, osa II. 3. painos. Helsinki, Oy Gaudeamus Ab. 387 p.

Walker W.T., Brady S. H., Taylor C., (1998). Updating existing travel simulation models with small-sample survey data using parameter scaling methods. Transportartation Research Record, No.1607, Transportation Research Board. National Research Council. Washington, D.C., pp. 55 - 61.

Watson P. L., Westin R, B., (1975). Transferability of disaggregate mode choice models. Regional Science and Urban Economics 5 (1975), North-Holland, pp. 227-249.

Widlert S., (1990). Överföring av modeller till Hälsingborg, TRANSEK AB. 25 p.

YTV, (1989). Linja-auto- ja raideliikenteen palvelutasojoustojen merkitys ja kuvaaminen liikennemalleissa. Pääkaupunkiseudun yhteistyövaltuuskunta YTV, Pääkaupunkiseudun julkaisusarja B 1989:10. Helsinki, 23 p.

YTV, (1990a). Ajoneuvoliikenteen määräpaikkatutkimus 1987-88. Pääkaupunkiseudun liikennetutkimus 1988. Pääkaupunkiseudun yhteistyövaltuuskunta YTV, Pääkaupunkiseudun julkaisusarja B 1990:3. Helsinki, 147 p.

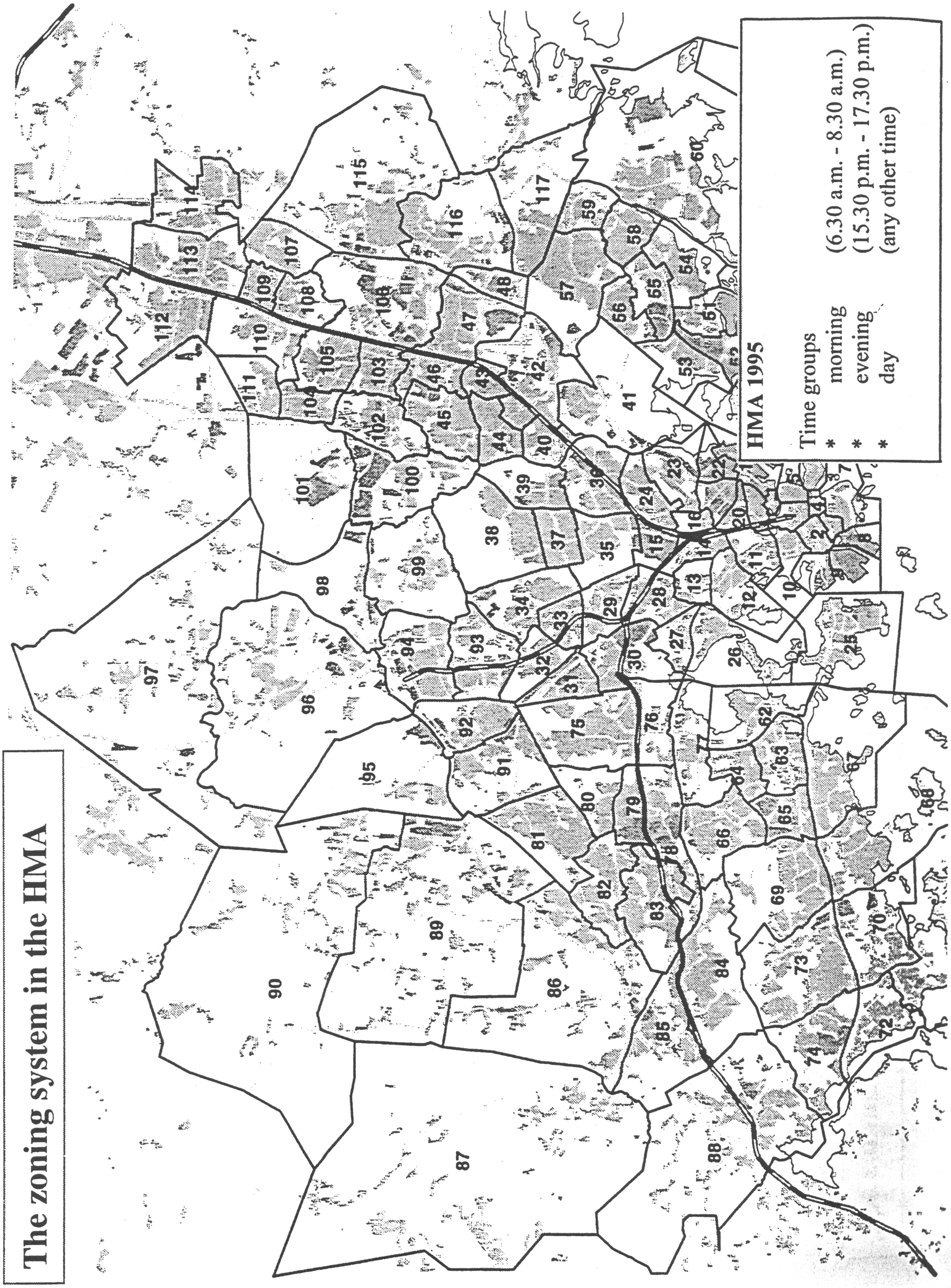
YTV, (1990b). Joukkoliikenteen määräpaikkatutkimus 1988. Pääkaupunkiseudun liikennetutkimus 1988. Pääkaupunkiseudun yhteistyövaltuuskunta YTV, Pääkaupunkiseudun julkaisusarja B 1990:5. Helsinki.

YTV, liikenneministeriö (1990a). Liikennemallit. Pääkaupunkiseudun liikennetutkimus 1988. Pääkaupunkiseudun yhteistyövaltuuskunta YTV, Pääkaupunkiseudun julkaisusarja B 1990:15. Helsinki. 230 p.

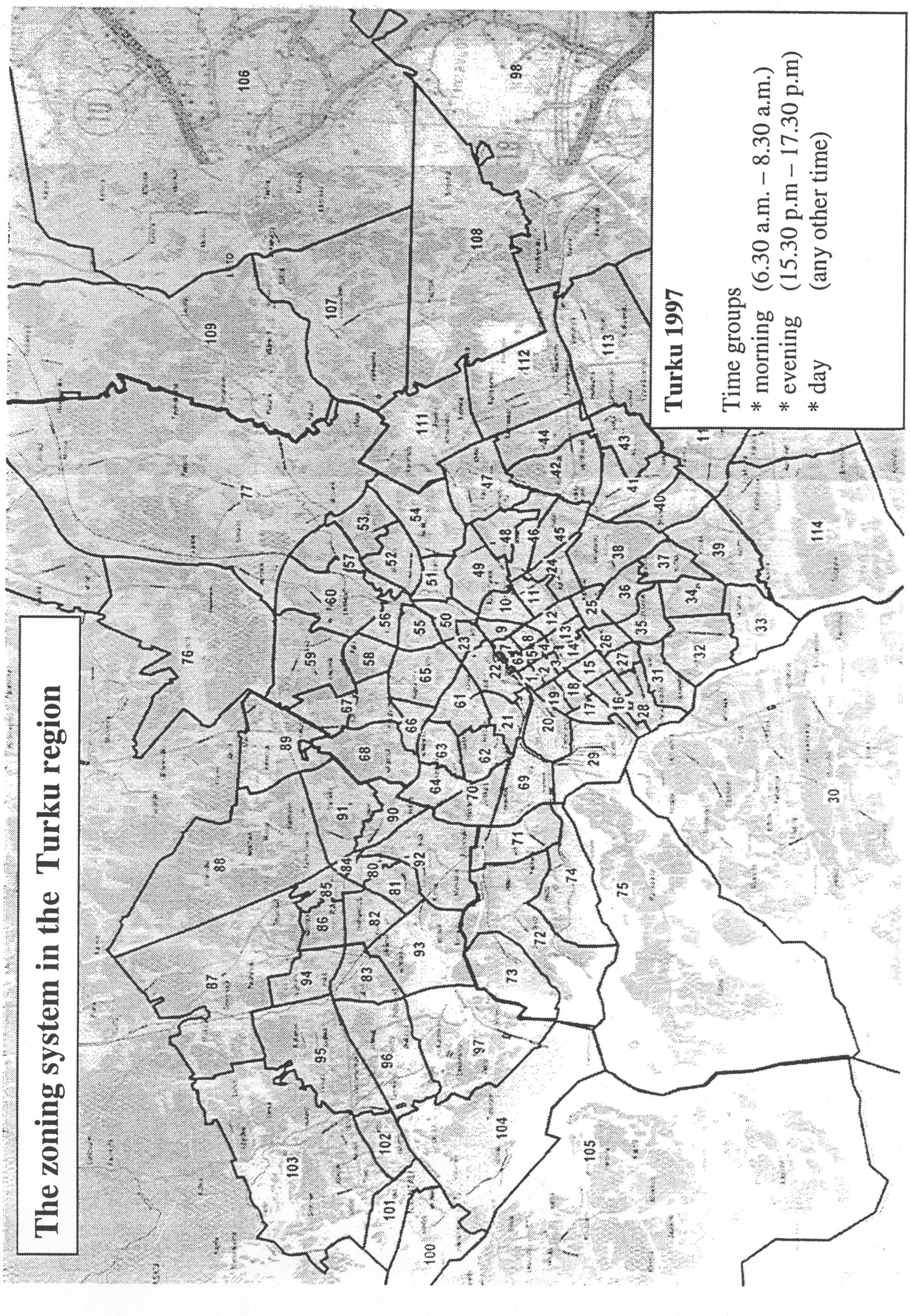
YTV, liikenneministeriö, (1990b). Liikkumistottumukset pääkaupunkiseudulla, Henkilöhaastattelut v. 1988. Pääkaupunkiseudun liikennetutkimus 1988. Pääkaupunkiseudun yhteistyövaltuuskunta YTV, Pääkaupunkiseudun julkaisusarja B 1990:2. Helsinki. 92 p.

YTV, liikenneministeriö, (1991). Pääkaupunkiseudun liikennetutkimus 1988, Yhteenveto. Pääkaupunkiseudun yhteistyövaltuuskunta YTV, liikenneministeriö, Pääkaupunkiseudun julkaisusarja B 1991:7. Helsinki. 130 p.

APPENDIX A


The zoning system in the HMA
HMA 1995
Time groups
* morning (6.30 a.m. - 8.30 a.m.)
* evening (15.30 p.m. - 17.30 p.m.)
* day (any other time)

APPENDIX B


The zoning system in the Turku region
Turku 1997
Time groups
* morning (6.30 a.m. – 8.30 a.m.)
* evening (15.30 p.m – 17.30 p.m)
* day (any other time)

APPENDIX C

The comparison of sequential and simultaneous estimation

1. Problem description

In this chapter the effect of estimation method in estimating nested logit models is examined. In principle the simultaneous estimation is more efficient than sequential estimation, due to the full information used in simultaneous estimation. It is also well known, that even the the sequential estimation method is simple, and produce estimates which are consistent; it has potentially strong problems. For example, if there are not sufficient data to estimate lower nest models, the estimates may be inefficient both in the sense that information is omitted at the lower levels and that the errors thus obtained are passed on to the superior level *(Ortúzar and Willumsen 1994).*

In our case, the results based on the estimations in the HMA and in the Turku region indicated that the coefficients of mode- and destination choice models differ greatly depending on the estimation method. As an example of that, Table C.1 presents the simultaneously and sequentially estimated models for home-based work trips and other home- based trips when using the entire set of empirical Turku data. For other home-based trips we have presented two different cases depending on if mode or destination choice is on the upper level. This has been made because when using the same order than used in sequential estimation, the coefficient of logsum is over 1.

Table C.1 shows that the models of home-based work trips did not differ very much from each other. The value of time is 3.6 euro/h (21.4 FIM/h) using sequential estimation for home-based work trips and 2.3 euro/h (13.6 FIM/h) by simultaneously estimated model. However, the differencies for other home-based trips are larger. The value of travel time coefficient, in particular, seems to be biased when using simultaneous estimation. The value of time for other home-based trips is 0.02 euro/h (0.1 FIM/h) by simultaneously estimated models and 1.2 euro/h (6.9 FIM/h) for sequentially estimated models.

Next the possible reasons causing differences between the sequential and simultaneous estimation are studied. Because in preliminary studies the variable “number of transfers” affected strongly the coefficient of travel time, the examination is made both for the models with the variable “number of transfer” as well as for the models without this variable. The effect of the logsum variable is studied because when empirical data was used the parameter for the logsum was over one when estimating the models for other home-based trips. Therefore simultaneous estimation also was problematic, in particular, for other home-based trips.

Table C1: *Estimation results of mode choice models using simultaneous or sequential estimation of home based work trips and other home based trips Estimated from data collected in the Turku region in 1997. The shaded models are used in model transfer in Chapter 6.*

	Home Based Work Trips		**Other Home Based trips**		
	Sequential estimation	**Simultaneous estimation**	**Sequential estimation**	**Simultaneous estimation A**	**Simultaneous estimation B**
Distance 0-5 (Walk)	-0.878 (-24.2)	-0.812 (-23.0)	-1.038 (-49.0)	-1.153 (-82.8)	-0.914 (-46.5)
Distance > 5 (Walk)	-0.339 (-11.0)	-0.344 (-13.3)	-0.247 (-13.5)	-0.282 (-17.6)	-0.215 (-14.3)
Total travel time (Ptr, Car)	-0.022 (-8.2)	-0.017 (-7.7)	-0.014 (-7.3)	-0.006 (-5.9)	-0.0002 (-0.2)
Number of transfers (Ptr)	-	-	-0.437 (-10.5)	-0.600 (-16.1)	-0.591 (-17.7)
Walk-dummy (Walk)	2.137 (11.4)	2.041 (15.7)	2.157 (20.4)	2.903 (28.4)	2.461 (42.1)
Car-dummy (Car)	-0.506 (-2.6)	-0.903 (-5.5)	-0.893 (-7.7)	-0.872 (-8.9)	-0.624 (-7.3)
Cars/household Car)	0.658 (3.8)	1.288 (10.3)	1.173 (13.3)	1.768 (24.2)	1.399 (24.4)
Trip cost/income (Ptr, Car)	-0.884 (-7.2)	-1.073 (-12.7)	-1.739 (-20.4)	-2.311 (-52.6)	-1.709 (-31.0)
Ln(nr of jobs) (Dest)	1.000 -	1.000 -	-	-	-
Log sum (Dest)	0.734 (43.2)	0.826 (16.8)	1.202 (134.8)	0.738 (35.1)	1.354 (36.4)
Scale factor	-	-	1.000	1.000	1.000
- service employment	-	-	14.89	14.70	14.53
- retail employment	-	-	2.52	2.43	2.31
Number of trip observations	4,442	4,442	13,989	13,989	13, 989
ρ^2(c) (mode and dest)	0.2011 and 0.1431	0.1486	0.2265 and 0.2802	0.2689	0.2690
VOT (FIM/h)	21.4	13.6	6.9	2.2	0.1

Walk=walk and bicycle
Ptr=bus and train
Car =driver or passanger

A mode choice on the upper level
B destination choice on the upper level

2. Methodology

The importance of estimation method is studied by using simulated data, because there is no real certainty as to which of these two methods better corresponds to the real situation. In the simulated data the true model coefficients are known, and can be compared with the results of the two estimation methods.

The main steps of the study are:

1. The artificial data are created using Monte Carlo simulation, assuming the true coefficients in Table C.2 and C.3. Choice probabilities are approximated by repeating Monte Carlo simulation process 100 times with the parameters held constant. The generation of attribute values is mainly based on the attribute values taken from the data set (HBW) used in the case study. However, two different logsum parameters are used to study the effect of logsum variable. That is, totally 200 samples are taken, 100 with the logsum of 0.8 and 100 with the logsum of 2.0.

   The empirical values of the origin and the starting time has been used as the basis of the simulation. In the empirical data, trips are directed from home to destination, or from destination to home. However, when simulating data, all the trips are directed from home to destination. Therefore, when the trip in empirical data is directed from destination to home, the home is used to represent an origin, and the starting time of the trip is also changed to correspond to the reversed situation in the empirical data.

2. A set of alternatives is defined (354 alternatives, 3 modes and 118 destinations), and the respective utilities are conducted from the attributes of that alternative. After the computation of utilities, choices are simulated by drawing a uniform random number from 0 to 1 for each individual and choice set, and comparing it to the logit probabilities to determine which alternative was chosen. All the coefficients and the individual-specific error term are assumed to have zero variation across the sample. These assumptions correspond to those of the basic linear logit model.

3. The models are estimated for each simulated sample by using both the simultaneous and sequential estimation. The results are presented in Tables C.2 and C.3. We have calculated the average coefficients for 100 mode and destination choice models to compare, how well different estimation methods can repeat the coefficients used in data simulation.

Table C.2: The average coefficients for 100 mode choice models using simultaneous or sequential estimation and simulated data.

		Simulated data (logsum=0.8)		**Simulated data (logsum=2.0)**	
	Coefficients used in data generation	**Sequential estimation**	**Simultaneous estimation**	**Sequential estimation**	**Simultaneous estimation**
Distance 0-5 (Walk)	-0.878 (-)	-0.882 (-2.3)	-0.872 (-2.4)	-0.854 (-2.3)	-0.814 (-2.5)
Distance > 5 (Walk)	-0.334 (-)	-0.340 (-10.8)	-0.338 (-12.8)	-0.311 (-8.7)	-0.344 (-8.7)
Total travel time (Ptr, Car)	-0.022 (-)	-0.022 (-7.7)	-0.021 (-8.8)	-0.018 (-6.5)	-0.014 (-6.7)
Walk-dummy (Walk)	2.137 (-)	2.167 (11.7)	2.216 (16.4)	2.392 (12.0)	2.401 (10.5)
Car-dummy (Car)	-0.506 (-)	-0.302 (-1.8)	-0.480 (-2.5)	-0.256 (-1.9)	-0.211 (-1.7)
Cars/household (Car)	0.658 (-)	0.681 (3.7)	0.703 (4.0)	0.722 (3.8)	0.683 (4.7)
Trip cost/income (Ptr, Car)	-0.884 (-)	-0.885 (-6.6)	-0.897 (-9.4)	-0.755 (-5.5)	-0.940 (-10.6)
Logsum (Dest)	0.8/2.0 (-)	0.812 (45.0)	0.820 (16.5)	2.166 (90.8)	2.1990 (21.4)
Scale factor	1.000	1.000	1.000	1.000	1.000
Number of trip observations	4,442	4,442	4,442	4,442	4,442
VOT (FIM/h)	21.4	21.1	18.1	22.0	17.9

Walk=walk and bicycle
Ptr=bus and train
Car =driver or passanger

Table C.3: The average coefficients for 100 mode and destination choice models using simultaneous or sequential estimation and simulated data for the model including the "number of transfers".

		Simulated data (logsum=0.8)		**Simulated data (logsum=2.0)**	
	Coefficients used in data generation	**Sequentially estimation**	**Simultaneous estimation**	**Sequential estimation**	**Simultaneous estimation**
Distance 0-5 (Walk)	-0.878 (-)	-0.872 (-2.3)	-0.869 (-2.4)	-0.854 (-2.3)	-0.811 (-2.5)
Distance > 5 (Walk)	-0.334 (-)	-0.343 (-10.8)	-0.337 (-12.8)	-0.309 (-8.7)	-0.342 (-8.7)
Total travel time (Ptr, Car)	-0.022 (-)	-0.024 (-7.7)	-0.021 (-8.8)	-0.017 (-6.5)	-0.012 (-6.7)
Nr of transfers (Ptr)	-0.437 (-)	-0.25 (-4.0)	-0.27 (-6.0)	-0.32 (-3.4)	-0.38 (-5.1)
Walk-dummy (Walk)	2.137 (-)	2.118 (11.7)	2.206 (16.4)	2.417 (12.0)	2.445 (10.5)
Car-dummy (Car)	-0.506 (-)	-0.280 (-1.8)	-0.478 (-2.5)	-0.173 (-1.9)	-0.256 (-1.7)
Cars/household (Car)	0.658 (-)	0.658 (3.7)	0.687 (4.0)	0.716 (3.8)	0.659 (4.7)
Trip cost/income (Ptr, Car)	-0.884 (-)	-0.869 (-6.6)	-0.872 (-9.4)	-0.742 (-5.5)	-0.973 (-10.6)
Logsum (Dest)	0.8/2.0 (-)	0.848 (44.1)	0.826 (16.5)	2.135 (88.5)	2.321 (21.4)
Scale factor	1.000	1.000	1.000	1.000	1.000
Number of trip observations	4,442	4,442	4,442	4,442	4,442
VOT (FIM/h)	21.4	21.1	18.1	22.0	17.9

Walk=walk and bicycle
Ptr=bus and train
Car =driver or passanger

Tables C.2 and C.3 show that the variance in the coefficients is larger for sequentially estimated models than for those estimated simultaneously. Although simultaneous estimation is more efficient than corresponding sequential estimation, sequential estimation was, unlike the simultaneous estimation, always able to repeat the coefficients used in the data creation. That is, the models based on the sequential estimation seem to be identifiable. However, when using simultaneous estimation, the coefficient of travel time, in particular, is biased in the situation where the logsum is 2.0. The problem is less significant when using models without the variable "number of transfers" than when using models with them.

Generally, simultaneous estimation has been thought of as being better than sequential estimation, because simultaneous estimation optimize the entirety of mode-and destination choice. However, this strength of the method can also be regarded as weakness. The probability that a certain destination (from all of the 118 destinations) is really chosen, is small. The lack of alternative-specific constants at the destination choice level makes it even more difficult to find an optimal solution.

The abovementioned simulation tests indicated that both the logsum and the variable number of transfers affect those results that are based on the simultaneous estimation. However, in essence, the problem is mainly caused by the multicollinearity, as is described in *Rich (2002)* and *Horowitz (1981).*

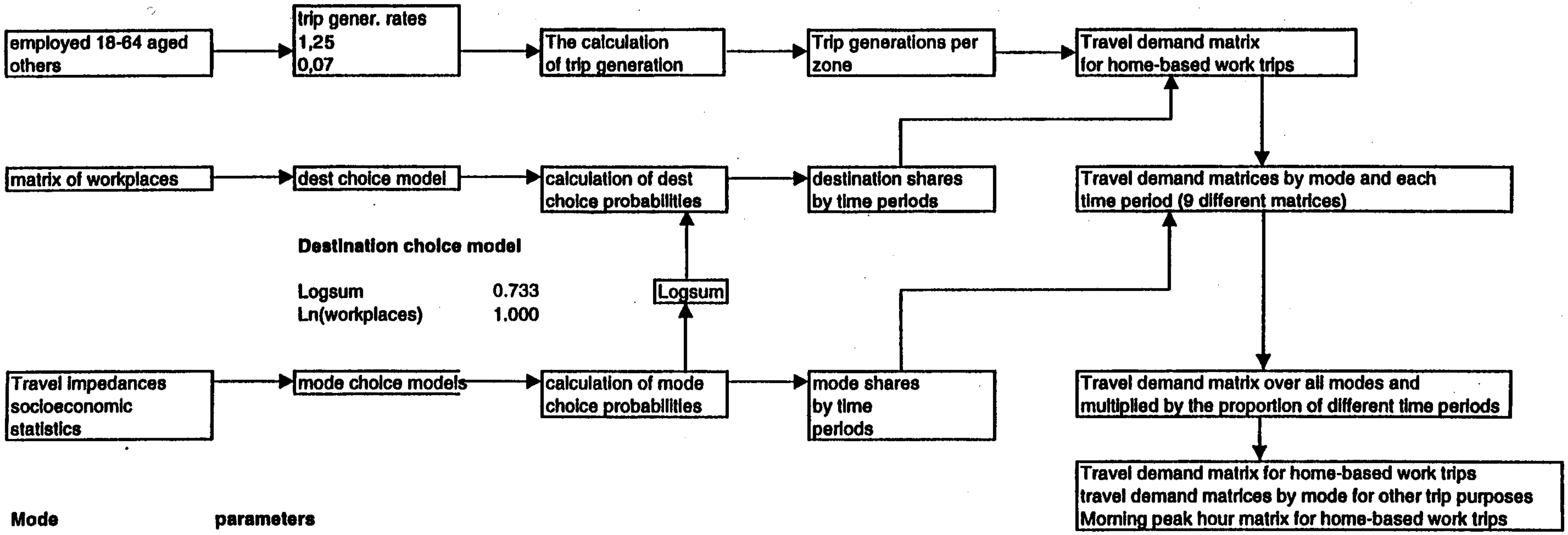
Horowitz (1981) has shown, that the use of zonally averaged variables in maximum-likelihood estimation will normally yield inconsistent estimates of disaggregate choice probabilities. It will also produce inconsistent estimates of zonal-average choice probabilities, unless the following conditions are fulfilled:

- The zonally averaged explanatory variables are not correlated with any disaggregate explanatory variables that are included in the model's specification
- The zonally averaged variables have the same joint distribution function in each zone, both in the estimation data set and in the data sets used for forecasting.

These conditions are not fulfilled in this study. The principal component analysis (not presented here) also suggested the use of generic coefficients for time and cost. That is, due to the identification problem that arises when using simultaneous estimation, sequential estimation was used in this research. The inverse order of mode and destination choice was also tested when using simultaneous estimation. However, the prediction performance for these models was not good.

APPENDIX D

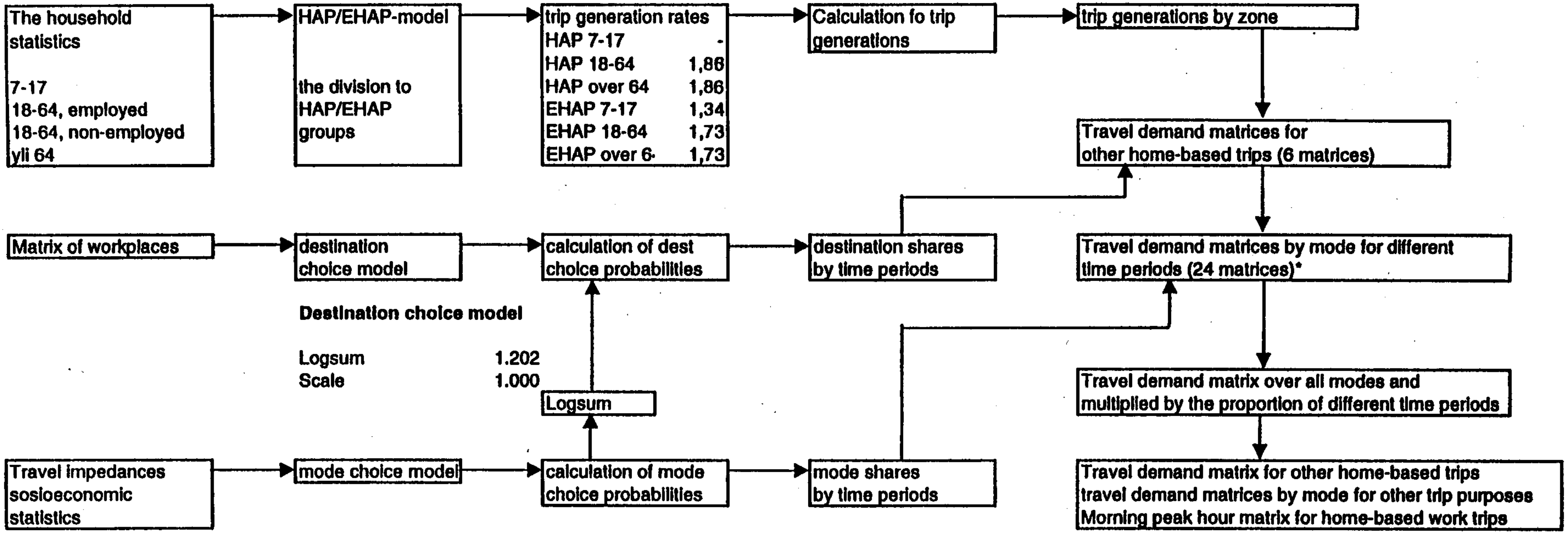
APPENDIX D: The forecast process for home-based work trips in the Turku region



Mode	parameters	
car	travel time	-0.0222
	cost/income	-0.8839
	cars/hh	0.658
	car dummy	-0.506
ptr	travel time	-0.022
	cost/income	-0.884
walk and bicycle	distance 0-5 km	-0.878
	distance over 5 km	-0.339
	kv-dummy	2.137

APPENDIX E

APPENDIX E: The forecast process for other home-based trips in the Turku region



Mode	Parameters	
car	travel time	-0.0135
	cost/income	-1.739
	cars/hh	1.173
	car dummy	-0.893
ptr	travel time	-0.0135
	cost/income	-1.739
	number of transfers	-0.437
walk and bicycle	distance 0-5 km	-1.038
	distance over 5 km	-0.247
	walk dummy	2.157

* matrices are calculated separately for HAP/EHAP groups by using the common mode choice model estimatedfor other home-based trips

APPENDIX F

The formulation of cost and parking cost variables

Travel costs for car

Travel costs for cars are based on 0.10 euro/km in the HMA in 1995 and 0.12 euro/km in the Turku region in 1997. These costs include both the fuel costs and other variable costs and the parking costs. Fixed charges are not taken into account.

Parking costs (for one-way trips) have been cross-tabulated from the mobility survey based on the monthly costs car drivers had paid, and are included in the car costs.

The average cost for home-based work trips was assumed to be the price of a monthly cost divided by 25, which was the average number of work trips made in a month. When calculating the average cost for other home-based trips, only those people who were paid their parking as a single payment, were included to the examination.

Below is presented the average values of parking costs according to the destination zone in the case of the trip is directed from home to destination. In other case, when home is destination, the corresponding parking cost is defined according to the origin area. Parking costs are presented only those areas, in which the cost is different from zero.

	Area	Parking cost (eur)	
		HBW	OHB
1	Kluuvi	1.00	1.36
2	Kamppi	0.68	0.68
3	Punavuori	0.91	0.91
4	Kaartinkaupunki	0.97	0.97
5	Kruununhaka	1.23	1.23
6	Katajanokka	0.45	0.45
7	Ullanlinna, Eira, Kaivopuisto	0.34	0.34
8	Munkkisaari	0,00	0.00
9	Ruoholahti, Jätkäsaari	0,00	0.00
10	Etu-Töölö	0.35	0.35
11	Taka-Töölö	0.12	0.01
12	Meilahti	0.12	0.12
13	Ruskeasuo, Pikku Huopalahti	0,00	0.00
14	Länsi-Pasila	0.02	0.25
15	Pohjois-Pasila	0.00	0.00
16	Itä-Pasila	0.02	0.03
17	Hakaniemi	0.19	0.19
18	Kallio	0.09	0.19
19	Sörnäinen	0.01	0.17
20	Alppiharju	0.00	0.14
21	Vallila	0.00	0.06
22	Hermanni	0.00	0.00
23	Kumpula, Toukola	0.00	0.00
24	Käpylä, Koskela	0.00	0.00
63	Tapiola	0.00	0.28
101	Lentoasema	0.00	0.32

Area		Parking cost (eur)
		HBW, OHB
4	Turku city centre	0.31
5	Turku city centre	0.15
6	Turku city centre	0.13
7	Turku city centre	0.15
8	Turku city centre	0.23
56	Kärsämäki-Runosmäki	1.38
72	Pansio-Perno	0.13
98	Piikkiö city centre	0.09

Travel costs for public transport

Travel costs for public transport are zonal values, which have been cross-tabulated from the mobility survey data. The travel cost in public transport were calculated as an average cost based on the ticket types respondents indicated they owned. The average costs were calculated in divisions of four zones.

HMA

The average cost for home-based work trips was assumed to be the daily price of a monthly ticket divided by 70, which was the average number of trips made in a month. The average cost for other home-based trips was calculated as an average of travel costs approximated to the ticket type respondents indicated they owned. The monthly daily price of a monthly ticket was divided by 45. For people under 17 years old the price of a children's ticket was used. In addition, some discount tickets were also taken into account. The average costs of public transport are as followed:

	Hki city centre	Hki suburb	Espoo+Kauniainen	Vantaa
Hki city centre	0.84	0.84	1.77	1.77
Hki suburb		0.84	1.77	1.77
Espoo+Kauniainen			1.09	1.93
Vantaa				1.09

Turku region

The average cost of the internal trips made in the Turku city area was defined to be 0.79 euros/trip. The price between the Kaarina and Turku was approximated to be 0.84 euros/trip and between the Raisio and Turku 1.68 euros/trip. Other prices are based on the distance rates. The factor 0.75 when using distance rates is based on the assumption that approximately 60 percent of all passangers have season ticket and approximately 45 percent of all passangers are children. The final ticket price was defined as follows:

```
 if (((origin=Turku) and (destination=Turku))) then price:=0.79
else if((origin=Kaarina) and (destination=Turku)) then price:=0.84
else if((destination=Turku) and (destination=Kaarina)) then price:=0.84
else if((origin=Raisio) and (destination=Turku)) then price:=1.68
else if((destination=Turku) and (destination=Raisio)) then price:=1.68
(*according to distance rate*)
else if ((0<=dist) and (dist<6)) then price:=0.75*1.68
else if ((6<=dist) and (dist<9)) then price:=0.75*1.85
else if ((9<=dist) and (dist<12)) then price:=0.75*2.19
else if ((12<=dist) and (dist<16)) then price:=0.75*2.52
else if ((16<=dist) and (dist<20)) then price:=0.75*2.86
else if ((20<=dist) and (dist<25)) then price:=0.75*3.36
else if ((25<=dist) and (dist<30)) then price:=0.75*3.70
else if ((30<=dist) and (dist<35)) then price:=0.75*4.20
else if ((35<=dist) and (dist<40)) then price:=0.75*4.70
else if ((40<=dist) and (dist<45)) then price:=0.75*5.05
else if ((45<=dist) and (dist<50)) then price:=0.75*5.55
hav[j]:=2*(price);
```

APPENDIX G

The description of the bootstrap technique

This appendix gives a brief review of the bootstrap sampling methodology. The procedure of calculating bootstrap quantities is presented more specifically in *Efron and Tibshirani (1993).*

The bootstrap was introduced in 1979 as a computer-based method for estimating the standard error of parameter θ (*Efron and Tibshirani 1993).* The basic idea of bootstrap is to take that sample which we are interested in and use it as a population and then by resampling create a new sample, a bootstrap sample, which we use to compute the quantities that we are interested in. If we repeat this several times, obtaining lots of bootstrap samples, we can use the mean of the computed quantities as an estimate of the expected value of this bootstrapped quantity (*Bergström 1999*).

According to Efron and Tibshirani *(1993)* bootstrap methods depend on the notation of a boostrap sample. Let F be the empirical distribution, putting probability 1/n on each of the observed values x_i, i=1,2,...,n. A bootstrap sample is defined to be a random sample of size n drawn from F, say

$$\begin{aligned} &x^*=(x_1^*, x_2^*, ..., x_n^*) \\ &F \rightarrow (x_1^*, x_2^*, ..., x_n^*). \end{aligned} \tag{1}$$

The star notation indicates that x* is not the actual data set x, but rather a randomized, or resampled, version of x.

There is another way to say: the bootstrap data points $x_1^*, x_2^*, ..., x_n^*$ are a random sample of size n drawn with replacement from the population of n objects $(x_1, x_2,..., x_n)$. Thus, we might have $x_1^*=x_7$, $x_2^*=x_3$, $x_3^*=x_3$, $x_4^*=x_{22}$, ..., $x_n^*=x_7$. The bootstrap data set $(x_1^*, x_2^*,..., x_n^*)$ consists of members of the original data set $(x_1, x_2,..., x_n)$, some appearing zero times, some appearing once, some appearing twice, etc.

The bootstrap estimate of $se_F(\theta)$ is the standard error of θ for data sets of size n randomly sampled from F. Computationally, the bootstrap algorithm works by drawing many independent bootstrap samples, evaluating the corresponding bootstrap replications, and estimating the standard error of θ by the empirical standard deviation of the replications. The result is called the bootstrap estimate of standard error, denoted by se_B, where B is the number of bootstrap samples used.

$$\hat{se}_B = \sqrt{\frac{\sum_{b=1}^{B} [\hat{\theta}^*(b) - \hat{\theta}^*(.)]^2}{(B-1)}}, \textit{ where}$$
$$\hat{\theta}^*(.) = \sum_{b=1}^{B} \frac{\hat{\theta}^*(b)}{B} . \tag{2}$$

The ideal bootstrap estimate se_∞ has the smallest possible standard deviation among nearly unbiased estimates of $se_F(\theta)$, at least in an asymptotic ($n \to \infty$) sense. It is not hard to show that se_B always has greater standard deviation than se_∞. The practical question is "how much greater?"

The increased variability due to the stopping after B bootstrap replications, rather than going on to infinity, is reflected in an increased coefficient of variation,

$$cv(\hat{se}_B) = \sqrt{cv(\hat{se}_\infty)^2 + \frac{E(\Delta)+2}{4B}}. \tag{3}$$

Here Δ is a parameter that measures how long-tailed the distribution of θ^* is; Δ is zero for the normal distribution, it ranges from -2 for the shortest-tailed distributions to arbitrarily large values when F is long-tailed. Table G1 compares $cv(se_B)$ with $cv(se_\infty)$ for various choices of B, assuming Δ=0.

Table G1. The coefficient of variation of se_B as a function of the coefficient of variation of the ideal bootstrap estimate se_∞ and the number of bootstrap samples B (Efron and Tibshirani 1993).

		B→				
		25	50	100	200	∞
$cv(se_\infty)$	0.25	0.29	0.27	0.26	0.25	0.25
	0.20	0.24	0.22	0.21	0.21	0.20
	0.15	0.21	0.18	0.17	0.16	0.15
	0.10	0.17	0.14	0.12	0.11	0.10
	0.05	0.15	0.11	0.09	0.07	0.05
	0.00	0.14	0.10	0.07	0.05	0.00

APPENDIX H

The effect of repeated measurements in model transfer

1. Introduction

This appendix concerns the repeated-measurement issue. As stated in Chapter 5.3.3.3, repeated answers provided by the same individual do not affect the real variation (std_p), but decrease the estimated standard error (std_t-values). The repeated measurement issue is considered because, although some of the methods react directly with observed variation, others (Bayesian approach and combined transfer estimation) only react to the estimated standard deviations (that is, the coefficients will be emphasized with respect to the inverse of their estimated variances) in model transfer. Thus the differences between these two standard deviations and correlations may cause differences in the comparison of the model transfer methods.

2. Repeated measurement

2.1 Problem describtion and methodology

The aim of this appendix is to investigate the degree to which the repeated measurement error might affect the results found in Chapter 6.

First, we look at what happens, if the repeated measurement problem, which occurs in estimation and application context, is removed. That is, the Turku data, including only one randomly chosen observation per respondent, is used in model transfer. In this case, the whole set of one observation per respondent HMA data (1,133 trip observations) is used as the estimation context data. The samples of 650 respondents in the Turku data are used in the application context (the sample size is the same as for the whole one observation per respondent data in the Turku region).

Using normal two-day data creates the reference situation. The sample size is chosen so that the observed standard deviation std_p for the estimation and application context is the same as the std_p-value gathered by using the one observation per respondent set of data. In addition, the estimation context samples are taken so that the coefficients are as similar as possible to those using the one observation per respondent HMA data. In estimation context, the coefficients should be similar to those estimated by using the one observation per respondent set of data. However, there are some small differences between these models (Table H.1).

The studied combinations are presented in Figure H.1. The model coefficients for estimation context data sets and for seed samples of application context samples are presented in Table H.1.

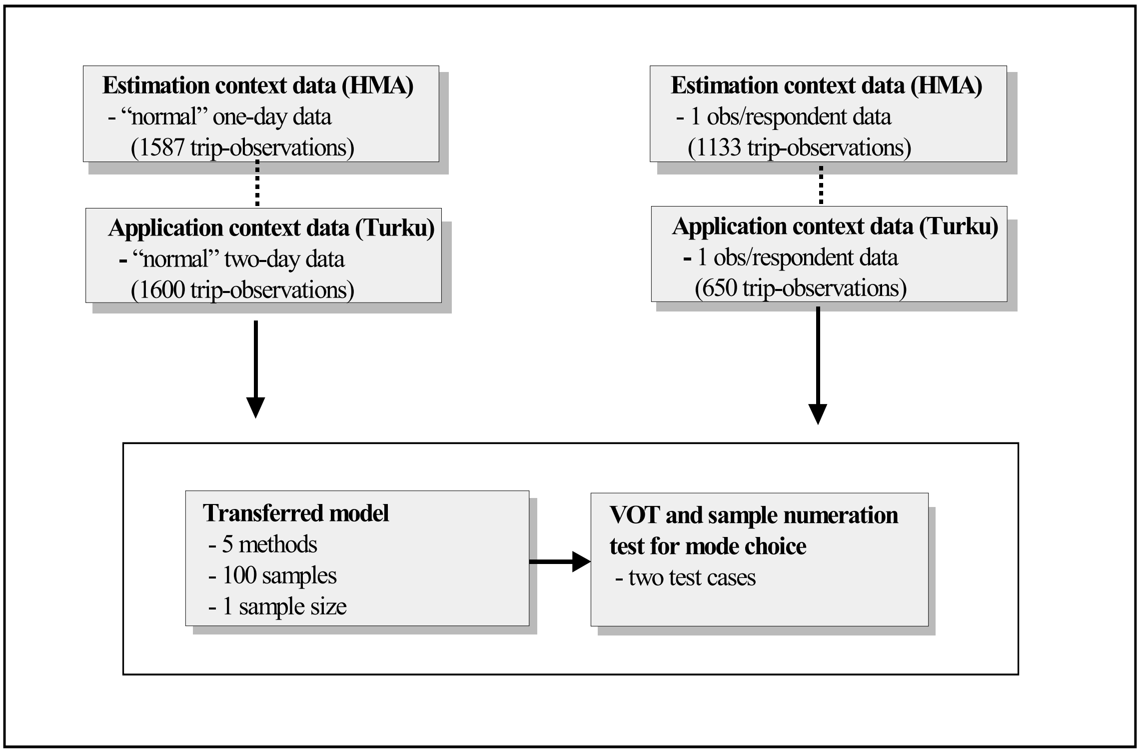


Figure H.1: The process when studying the effect of the repeated measurement issue.

The model transfer is investigated by all five methods studied in Chapter 6. All transferability tests have been made by using 100 samples (resampled from the Turku 1997 dataset) for each transfer method and studied combination. The combination B (only coefficient for distance 0-5 is estimated as data-specific) of data-specific coefficients is used for the joint context estimation. This is the combination, which proved to best for the corresponding std_p variation in empirical data.

The "normal" two day data represents the empirical data used in Chapter 6. The one observation per respondent data has been thought to represent the data, in which the repeated measurement problem is undoubtedly absent. In principle, the unbiased data could be formulated by merely eliminating the common observations.

Note, due to the two-day diaries used in the Turku area, the repeated measurement issue is more serious for application context (Turku) data than for the HMA data. Consequently, the repeated measurement correction is of a different size for the estimation and application context data sets. That is, when using the one day per respondent data, the effect of using two day diaries is also removed.

Table H.1: The coefficients for estimation context models and for seed samples used in the estimation context when studying repeated measurement issue.

	HMA* 1587 resp	HMA** 1 obs/resp	Turku *** entire data	Turku **** 1obs/resp
Distance 0-5 km	-0.952 (-13.1)	-0.911 (-11.0)	-0.878 (-24.2)	-0.873 (-14.1)
Distance >5 km	-0.474 (-9.5)	-0.449 (-8.3)	-0.339 (-11.0)	-0.321 (-6.6)
Total travel time (Car, Ptr)	-0.027 (-7.9)	-0.028 (-6.9)	-0.022 (-8.2)	-0.023 (-5.1)
Trip cost/income (Car, Ptr)	-1.364 (-9.1)	-1.397 (-7.8)	-0.884 (-7.2)	-0.984 (-4.8)
Cars/household (Car)	1.022 (3.3)	1.229 (3.4)	0.658 (3.8)	0.633 (2.2)
Walk dummy (Walk)	1.499 (5.5)	1.283 (4.1)	2.137 (11.4)	1.925 (6.2)
Car dummy (Car)	-1.858 (-6.5)	-1.973 (-5.7)	-0.506 (-2.6)	-0.584 (-1.8)
$\rho^2(c)$	0.2336	0.2235	0.2011	0.1845
No of observations	1,587	1,133	4,442	1,530

* estimation context data and model for reference data

** estimation context data and model for one observation / respondent data

*** data from which the reference samples of 1,600 trip-observations have been taken

**** data from which the "one observation per respondent" samples including 650 trip-observations have been taken

2.2 Results

Figures H.2 and H.3 present the main findings. Figure H.2 shows that the VOT error of the new sample models is the same for both data sets. This is plausible, because the sample size for the one observation per respondent data was chosen so that std_p would be the same as that from using the reference data. However there are differences between the results concerning the Bayesian method and the combined transfer estimation. One could expect that the Bayesian method would be better for the empirical data than for the one observation per respondent data, because the std_t is more highly overestimated in the application context (due to the 2-day data) than in the estimation context, causing the application context coefficients to be a bit more highly emphasized as well. Nonetheless, due to the violated correlations of empirical data, the results are quite the opposite. The Bayesian method, in particular give better results when using one observation per respondent data.

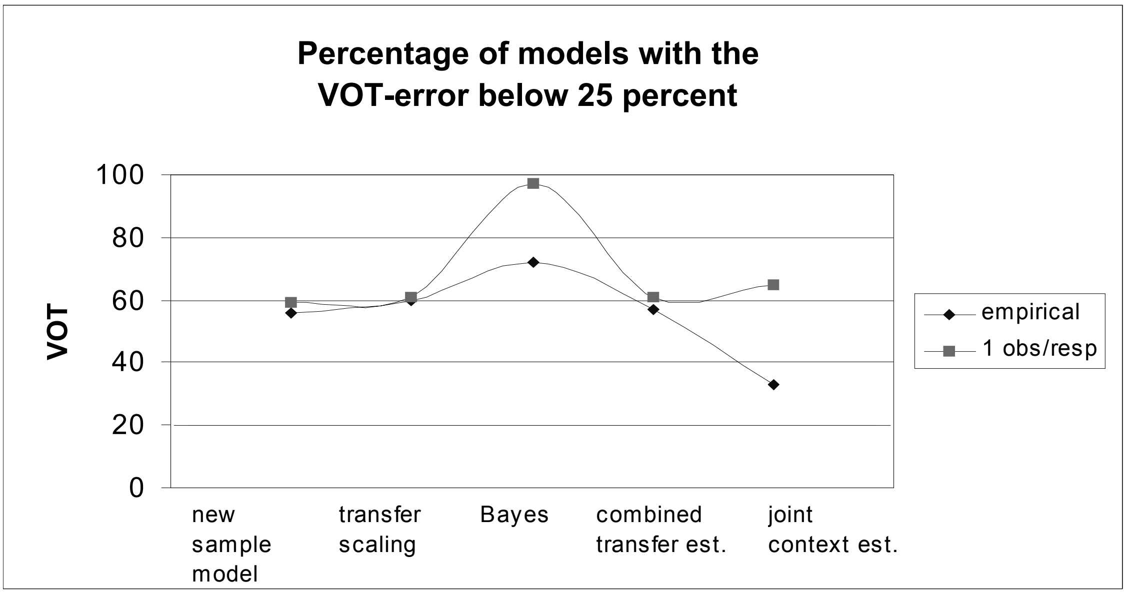


Figure H.2: *The VOT for samples taken one observation/respondent data and empirical reference data.*

Table H.2: *The average-VOT-error based on the 100 models estimated using empirical reference data and one observation per respondent data.*

Home-based work trips, average VOT-error		
	empirical	**1 obs/resp**
new model	25.5	25.5
transfer scaling	21.2	21.3
Bayes	22.7	7.8
combined transfer est	28.5	28.9
joint context est.	14.1	13.2

When considering the RSEE error (Figure H.3), the results concerning the Bayesian method and the combined transfer estimation look quite similar. However, the more precise examination indicates that there are differences by considering the average RSEE values presented in Table H.3. Also, in this case, the combined transfer estimation and the Bayesian procedure, in particular, give better results when using the one observation per respondent data. Due to the larger transfer bias, the differences are greater for cost-elasticities than for time-elasticities. The differences between these two cases are smaller for the combined transfer estimation, because this method emphasize, in any ways, the application context data.

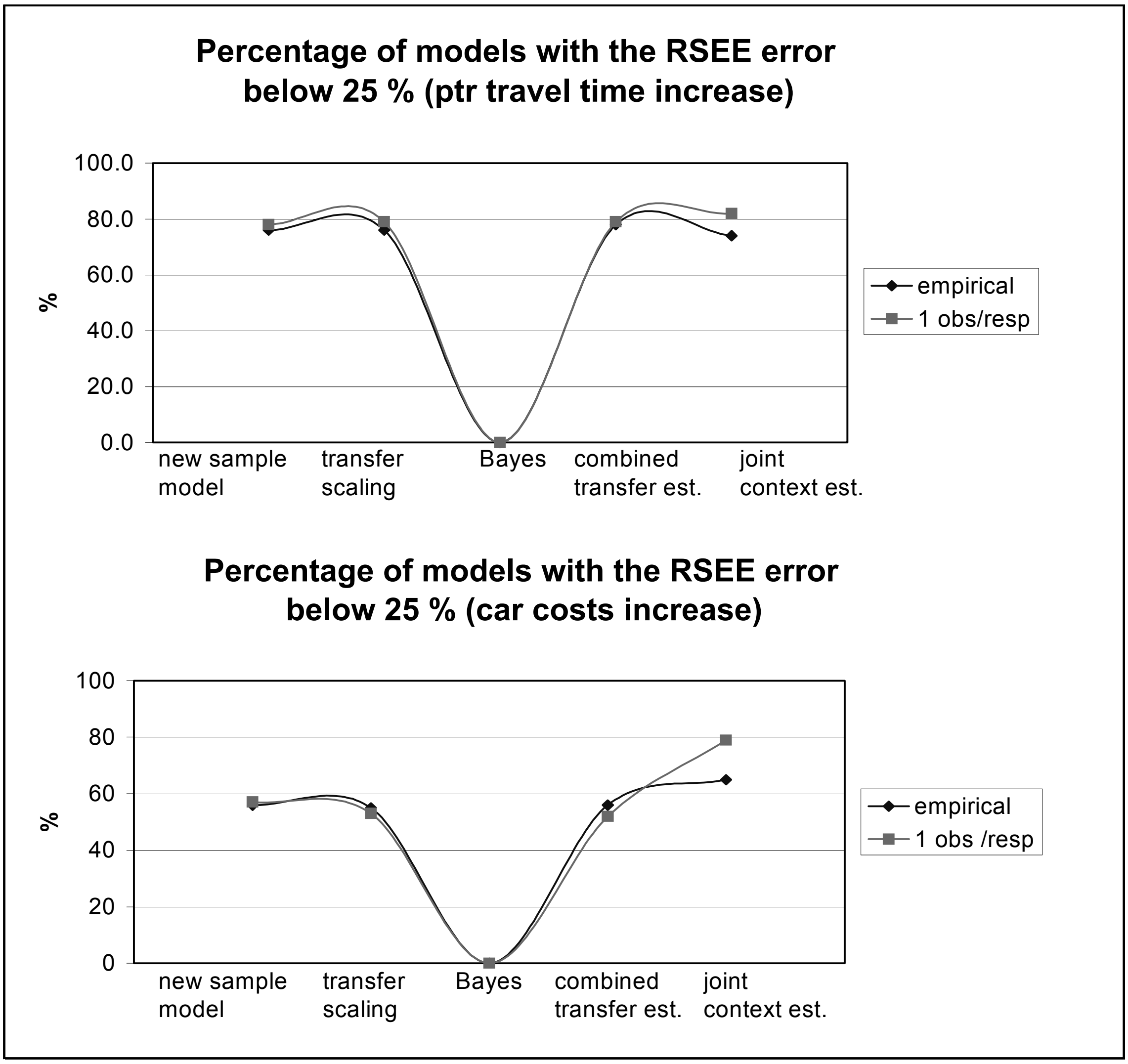


Figure H.3: *The RSEE for samples taken from the one observation per respondent data and from empirical reference data.*

Table H.3: *The average RSEE based on the 100 models estimated using empirical reference data and one observation per respondent data.*

Home-based work trips (average RSEE)				
	empirical reference data		**one observation/respondent data**	
	ptr time + 30 %	car costs +10 %	ptr time + 30 %	car costs +10 %
new model	19.6	27.6	16.1	25.1
transfer scaling	18.3	29.1	17.1	26.9
Bayes	63.1	88.5	55.2	54.9
combined transfer est	18.1	33.0	16.2	25.1
joint context est.	17.1	22.9	15.6	14.6

APPENDIX I

The average RSEE and RSEEF-values for 100 transferred models, home-based work trips in the Turku region. (The best method and results are shaded)

Home-based work trips				
	RSEE (%)		RSEEF (%)	
sample size 425	ptr time + 30 %	car costs + 10 %	ptr time + 30 %	car costs + 10 %
new model	34.1	63.5	30.0	74.2
transfer scaling	34.7	65.1	30.6	65.7
Bayes	65.3	102.4	31.8	98.5
combined transfer est.	33.7	61.6	29.3	73.8
joint context est.	32.4	53.0	17.0	64.6
sample size 850				
new model	26.8	41.3	24.1	49.4
transfer scaling	26.4	41.8	23.7	44.5
Bayes	66.1	98.8	35.9	96.3
combined transfer est.	26.5	40.7	23.5	52.2
joint context est.	19.6	35.5	13.5	52.9
sample size 1,700				
new model	17.2	28.1	15.3	34.5
transfer scaling	17.6	32.2	18.0	39.6
Bayes	65.6	90.1	38.6	95.2
combined transfer est	17.5	28.7	15.7	36.5
joint context est.	14.0	31.6	10.9	45.0
sample size 3,400				
new model	13.0	19.0	10.7	29.5
transfer scaling	12.6	20.9	13.0	30.7
Bayes	58.7	82.4	38.2	90.4
combined transfer est	13.2	19.1	10.9	29.8
joint context est.	11.2	18.0	9.2	27.3
Sample size 4,675				
new model	11.7	15.1	10.5	25.7
transfer scaling	12.1	19.1	13.6	34.8
Bayes	55.6	65.2	34.0	87.5
combined transfer est	11.9	17.0	10.4	26.0
joint context est.	10.3	15.5	9.4	25.0

APPENDIX J

The average RSEE and RSEEF-values for 100 transferred models, other home-based trips in the Turku region. (The best method and results are shaded)

Other home-based trips				
	RSEE (%)		RSEEF(%)	
Sample size 140	ptr time + 30 %	car costs +10 %	ptr time + 30 %	car costs +10 %
new model	75.7	32.7	67.6	27.1
transfer scaling	81.9	33.4	51.8	24.7
Bayes	96.9	23.7	32.5	6.6
comb. transf. est.	75.2	32.1	65.9	25.9
joint context est.	45.5	30.7	35.2	19.4
sample size 275				
new model	60.5	25.8	51.3	19.7
transfer scaling	77.3	29.4	50.7	18.1
Bayes	101.7	23.7	34.3	6.9
comb. transf. est.	59.5	25.5	49.2	17.1
joint context est.	36.6	23.2	27.3	12.9
sample size 550				
new model	41.0	20.3	35.2	14.8
transfer scaling	63.6	23.1	42.0	16.1
Bayes	105.8	23.1	42.2	5.0
comb. transf. est	40.3	20.7	34.6	14.6
joint context est.	27.1	17.9	25.3	10.4
sample size 1,100				
new model	29.9	14.2	23.6	9.7
transfer scaling	64.1	18.9	41.2	11.1
Bayes	106.7	22.6	46.5	4.8
comb. transf. est	29.5	14.3	23.3	9.7
joint context est.	22.6	13.6	16.9	7.5
sample size 2,300				
new model	23.9	10.9	19.0	5.8
transfer scaling	66.4	16.7	42.9	8.8
Bayes	96.9	20.1	48.0	3.9
comb. transf. est	23.6	10.9	19.0	5.8
joint context est.	19.4	11.5	16.0	9.0
sample size 4,675				
new model	15.0	7.1	11.9	3.8
transfer scaling	64.1	14.4	39.0	7.6
Bayes	79.0	17.8	49.2	3.2
comb. transf. est	14.8	7.1	11.9	3.7
joint context est.	15.3	7.2	10.0	5.4

APPENDIX K

The variation of RSEE for home-based work trips in the Turku region

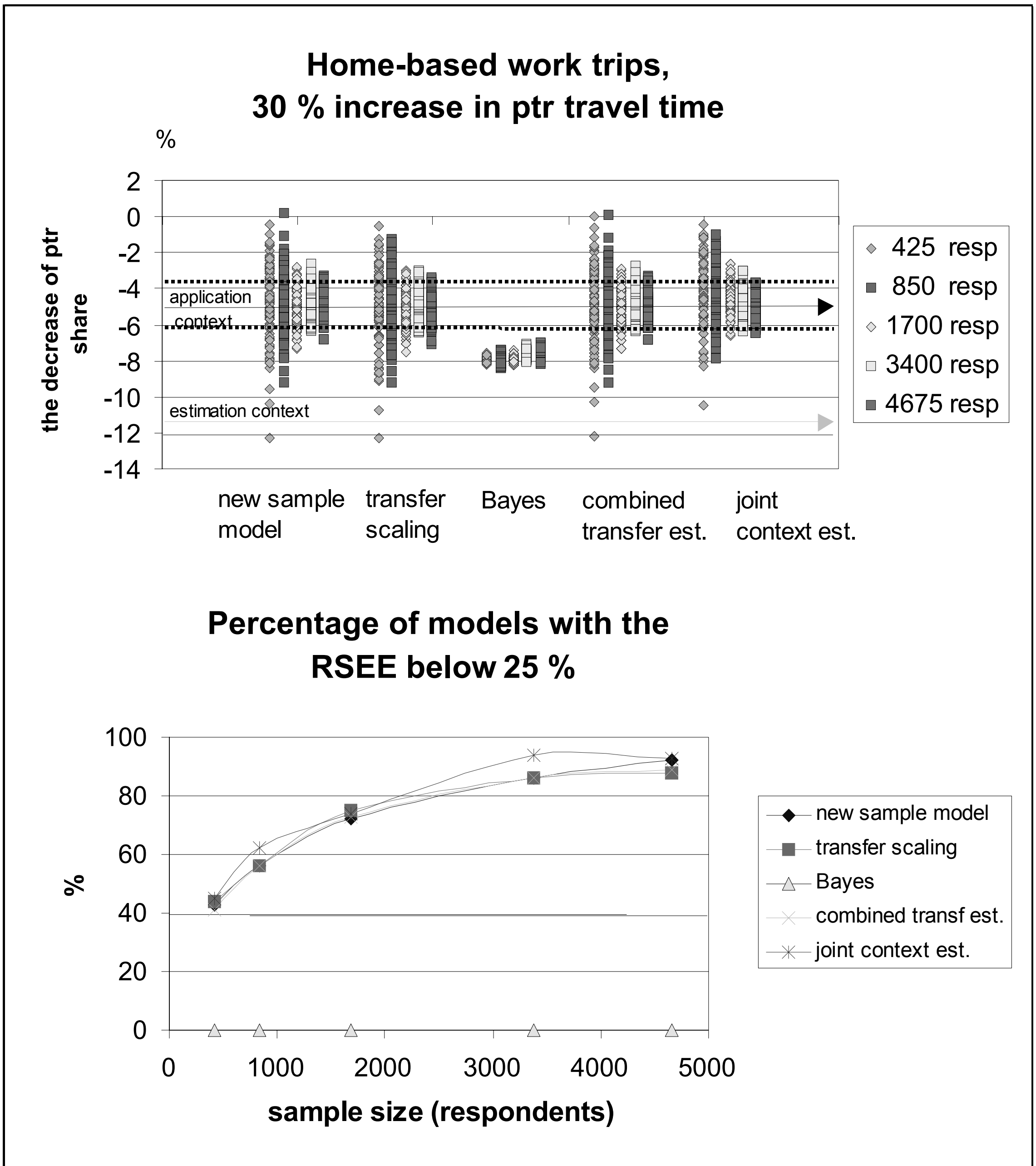


Figure K.1: *The variation of changes in public transport shares (ptr) by transfer method, when public transport travel time is increased by 30 percent in the Turku region; home-based work trips.*

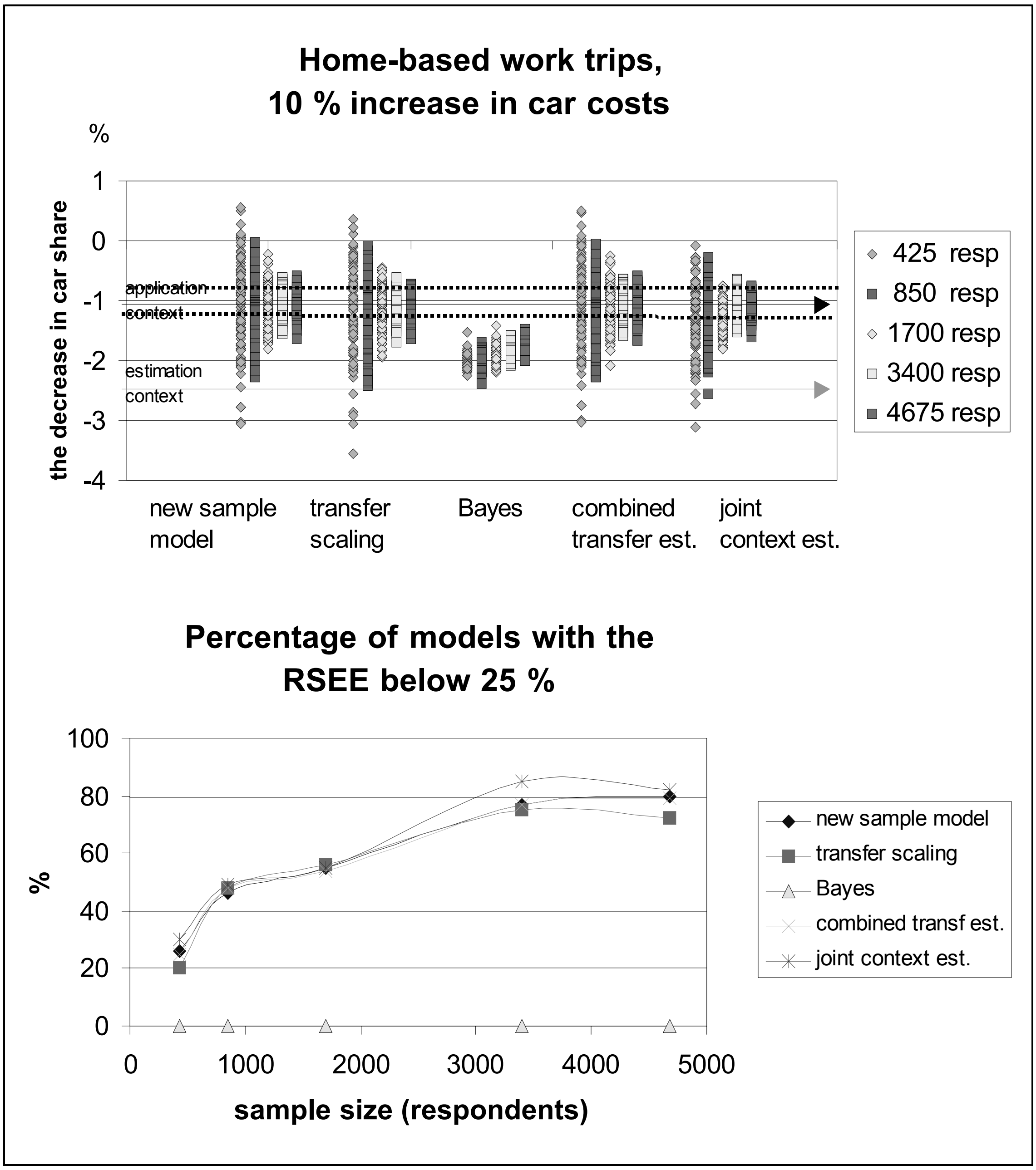


Figure K.2: *The variation of change in car shares by transfer method, when car costs are increased by 10 percent in the Turku region; home-based work trips.*

APPENDIX L

An example of MAEF-index (comparison of observed and predicted trips when using entire set of Turku data)

The Turku region is devided into the following four zones:

- Central city area of Turku (wide area 17)
- Suburb (wide areas 11, 12, 14, 16, 18 ja 20)
- Kaarina + Raisio + Naantali (wide areas 7, 9, 10, 15, 22-24)
- Lieto + Piikkiö (wide areas 2, 4-6 ja 13)

Table L.1 and L.2 present the weighted generation-attraction matrices of observed and predicted trips based on the models estimated from the entire set of 1997 Turku data.

Table L1: The comparison of observed and predicted trips for home-based work trips (predicted/observed)

	Attraction			
Generation	(1)	(2)	(3).	(4)
Central city area	18114/21096	6692/8293	1854/2531	282/654
Suburb	30554/31564	29265/25186	8236/5010	1441/1277
Kaarina+Raisio+Naant.	10611/10526	10365/7975	9542/13234	723/995
Lieto + Piikkiö	3702/3058	3616/2091	1475/1102	1459/3974
Total	62981/66244	49938/43545	21107/21877	3905/6900

Table L.2: The comparison of observed and predicted trips for other home-based trips (predicted/observed)

	Attraction			
Generation	(1)	(2)	(3)	(4)
Central city area	66958/76445	7957/16374	556/2689	20/712
Suburb	60847/68201	114878/76419	8722/9180	455/1702
Kaarina+Raisio+Naant.	10409/13720	15053/9264	56685/46039	371/546
Lieto+Piikkiö	3465/4735	6089/3113	1597/1579	17265/14561
Total	141679/163101	143977/105170	67560/59487	18111/17521

APPENDIX M

MD-values for different combinations of joint context estimation; home-based work trips.

Turku region, home-based work trips			sample size 425 respondents				
	cost	travel time	cars/hous.	dist. >5	dist. 0-5	most importan	total
A	63.4	22.5	59.3	23.6	13.7	145.2	182.5
B	41.9	27.1	45.6	29.5	16.0	**114.5**	**160.0**
C	48.6	36.1	51.7	41.3	18.2	136.4	196.0
D	58.8	36.1	56.2	44.3	18.2	151.1	213.7
E	58.8	45.1	135.3	47.2	21.7	239.2	308.1
F	54.3	36.1	59.3	44.3	19.4	149.7	213.3
G	53.2	40.6	127.7	44.3	19.4	221.4	285.1
H	60.0	45.1	103.4	50.2	19.4	208.4	278.0
I	57.7	31.6	132.3	32.5	16.0	221.5	269.9
			sample size 850 respondents				
A	55.4	18.0	51.1	18.9	8.3	124.5	151.8
B	30.6	22.5	35.0	20.7	10.3	**88.1**	**119.0**
C	33.9	27.1	38.0	28.0	11.7	99.0	138.8
D	36.2	27.1	44.1	28.6	11.9	107.3	147.8
E	39.6	31.6	94.3	27.4	10.0	165.4	202.9
F	35.1	27.1	39.5	28.0	12.3	101.6	142.0
G	37.3	31.6	95.8	28.9	12.4	164.7	206.0
H	49.8	40.6	77.5	41.6	21.7	167.9	231.2
I	37.3	27.1	97.3	26.3	10.3	161.7	198.2
			sample size 1700 respondents				
A	46.4	18.0	40.4	17.1	7.1	104.9	129.0
B	27.2	18.0	30.1	15.9	7.3	**75.3**	**98.5**
C	27.2	22.5	31.5	18.9	9.0	81.2	109.1
D	27.2	18.0	36.6	18.9	8.8	81.8	109.5
E	24.9	18.0	50.2	18.3	7.5	93.1	118.9
F	27.2	22.5	31.9	18.9	9.1	81.6	109.6
G	27.2	22.5	53.5	18.9	9.3	103.2	131.4
H	28.3	22.5	54.0	19.8	10.5	104.8	135.1
I	26.0	18.0	54.7	18.3	7.5	98.8	124.6
			sample size 3400 respondents				
A	35.1	9.0	30.3	12.4	4.7	74.3	91.4
B	17.0	13.5	22.8	10.9	4.6	53.3	68.8
C	17.0	18.0	22.0	12.1	5.2	57.0	74.4
D	17.0	13.5	24.6	12.4	5.2	55.1	72.8
E	17.0	13.5	39.2	13.0	5.5	69.7	88.2
F	15.8	13.5	21.0	12.4	5.5	**50.3**	**68.2**
G	18.1	18.0	38.6	13.0	5.4	74.7	93.1
H	18.1	13.5	41.7	13.0	5.6	73.3	91.9
I	17.0	13.5	38.8	12.4	4.6	69.3	86.2
			sample size 4675 respondents				
A	30.55	*13.5*	32.53	10.92	5.70	76.60	93.22
B	18.10	9.02	25.84	8.56	5.70	52.96	67.22
C	15.84	13.5	23.72	9.45	5.70	53.08	68.22
D	13.58	13.5	27.06	9.74	5.81	54.16	69.71
E	14.71	*13.5*	38.77	9.45	6.61	67.00	83.05
F	13.58	13.5	23.11	9.45	6.04	**50.21**	**65.69**
G	16.97	13.5	38.16	9.74	6.15	68.65	84.55
H	19.23	*13.5*	43.02	15.35	7.29	75.78	98.42
I	13.58	9.0	38.46	9.15	5.47	61.05	75.67

The models estimated by using the joint context estimation:

A) all coefficients are common
B) distance 0-5 km is data-specific
C) distance coefficients are data-specific
D) trip cost and distance coefficients are data-specific
E) trip cost, cars/household and distance coefficients are data-specific
F) travel time and distance coefficients are data-specific
G) cars/household and distance coefficients are data-specific
H) travel time, cost and distance coefficients are data-specific
I) coefficients of cars/household, cost and distance (0-5) variables are data-specific

APPENDIX N

MD-values for different combinations of joint context estimation; other home-based trips

Turku region, other home-based trips				sample size 140 respondents				
	cost	travel time	cars/hous.	dist. >5	dist. 0-5	nr.of transfers	most importan	total
A	42.6	88.9	50.5	85.8	13.6	57.9	182.0	339.3
B	32.2	51.9	41.7	52.2	13.8	69.1	**125.8**	**260.9**
C	32.8	59.3	55.4	121.5	14.9	68.6	147.5	352.5
D	30.5	74.1	76.2	122.7	15.3	64.0	180.8	382.8
E	31.6	111.1	74.7	124.7	15.6	54.9	217.4	412.6
F	33.4	111.1	39.9	123.9	16.8	67.5	184.4	392.5
G	37.4	59.3	74.1	118.2	15.8	70.0	170.8	374.8
H	31.6	111.1	74.2	120.6	16.6	63.4	216.9	417.5
I	32.8	74.1	73.9	78.9	13.7	65.0	180.8	338.4
				sample size 275 respondents				
A	38.0	74.1	40.9	76.1	9.2	51.5	153.0	289.8
B	22.4	44.4	29.0	42.5	8.7	64.7	**95.8**	**211.7**
C	23.3	44.4	28.1	48.2	10.1	63.8	**95.8**	217.9
D	21.9	59.3	39.2	47.0	10.4	59.5	120.4	237.3
E	26.5	88.9	41.8	55.1	11.8	48.9	157.2	273.0
F	22.4	81.5	28.1	50.2	11.3	62.9	132.0	256.4
G	25.9	51.9	41.8	51.8	10.8	62.0	119.6	244.2
H	24.2	81.5	47.7	53.4	11.7	58.6	153.4	277.1
I	25.3	51.9	41.8	61.1	9.2	61.1	119.0	250.4
				sample size 550 respondents				
A	30.5	59.3	32.1	64.8	6.0	45.7	121.9	238.4
B	16.7	42.4	28.4	21.2	6.1	52.6	87.5	167.4
C	16.7	29.6	21.2	27.9	6.8	59.7	**67.5**	**162.0**
D	16.7	44.4	28.0	27.5	6.9	53.8	89.1	177.3
E	19.6	59.3	34.4	32.8	7.4	42.8	113.3	196.3
F	16.1	59.3	21.2	29.1	7.6	58.8	96.6	192.1
G	19.6	37.0	34.1	30.4	6.8	57.4	90.7	185.3
H	17.8	51.9	37.0	29.1	7.4	53.3	106.7	196.5
I	17.8	37.0	34.2	41.3	5.8	56.5	89.0	192.6
				sample size 1,100 respondents				
A	23.6	44.4	20.9	47.8	4.8	36.6	88.9	178.1
B	9.8	22.2	22.2	24.3	4.8	52.6	54.2	135.9
C	12.7	22.2	15.9	23.5	5.6	52.6	**50.8**	**132.5**
D	12.7	29.6	18.8	23.5	5.8	45.1	61.1	135.5
E	12.7	44.4	23.1	25.1	5.9	31.3	80.2	142.5
F	12.7	44.4	15.9	23.9	6.5	51.5	73.0	154.9
G	13.8	29.6	25.6	24.7	5.9	49.9	69.0	149.5
H	13.2	37.0	24.0	24.3	6.6	44.1	74.2	149.2
I	13.8	29.6	26.3	31.6	4.7	51.9	69.7	157.9
				sample size 2,300 respondents				
A	17.3	37.0	16.2	32.4	3.9	27.4	70.5	134.2
B	8.6	19.3	17.9	16.2	3.9	41.2	45.8	107.1
C	9.2	22.2	11.9	16.2	3.9	38.9	**43.4**	**102.3**
D	9.8	30.4	13.6	16.2	4.8	29.7	53.8	104.5
E	9.8	33.3	15.5	17.4	4.6	22.9	58.6	103.5
F	9.2	35.6	11.9	16.2	4.8	36.6	56.7	114.3
G	10.9	24.4	15.3	16.2	4.8	36.6	50.6	108.2
H	10.4	33.3	17.9	20.2	4.8	22.9	61.6	109.5
I	9.8	24.4	15.3	24.3	3.9	36.6	49.5	114.3
				sample size 4,675 respondents				
A	11.5	22.2	12.4	20.6	2.9	19.2	46.1	88.8
B	7.5	14.8	16.5	13.4	3.9	29.0	38.8	85.1
C	7.5	14.8	11.3	12.1	3.9	27.7	33.6	77.3
D	7.5	14.8	11.0	11.3	3.9	20.4	**33.3**	**68.9**
E	7.5	22.2	13.3	13.0	3.9	16.5	43.0	76.4
F	6.9	22.2	10.5	11.3	3.9	27.2	39.6	82.0
G	8.6	14.8	13.1	12.6	3.9	27.0	36.5	80.0
H	7.5	22.2	13.5	16.6	3.9	17.8	43.2	81.5
I	7.5	14.8	13.4	16.6	3.9	26.8	35.7	83.0

The models estimated by using the joint context estimation:

A) all coefficients are common
B) distance 0-5 km is data-specific
C) distance coefficients are data-specific
D) trip cost and distance coefficients are data-specific
E) trip cost, cars/household and distance coefficients are data-specific
F) travel time and distance coefficients are data-specific
G) cars/household and distance coefficients are data-specific
H) travel time, cost and distance coefficients are data-specific
I) coefficients of cars/household, cost and distance (0-5) variables are data-specific

APPENDIX O

The variation of RSEE for other home-based trips

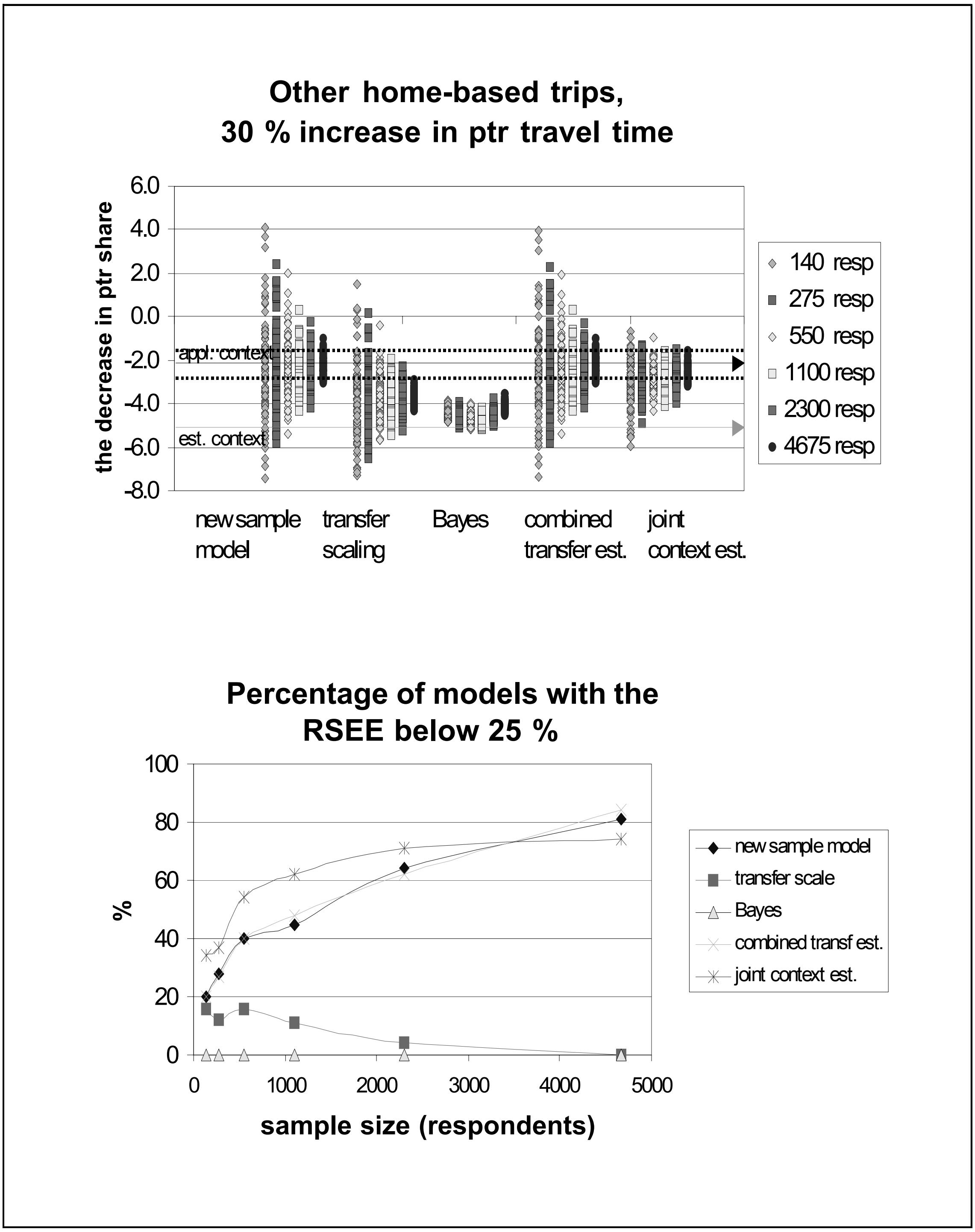


Figure O.1: *The variation of change in public transport (ptr) shares by transfer method; when public transport travel time is increased by 30 percent in the Turku region; other home-based trips.*

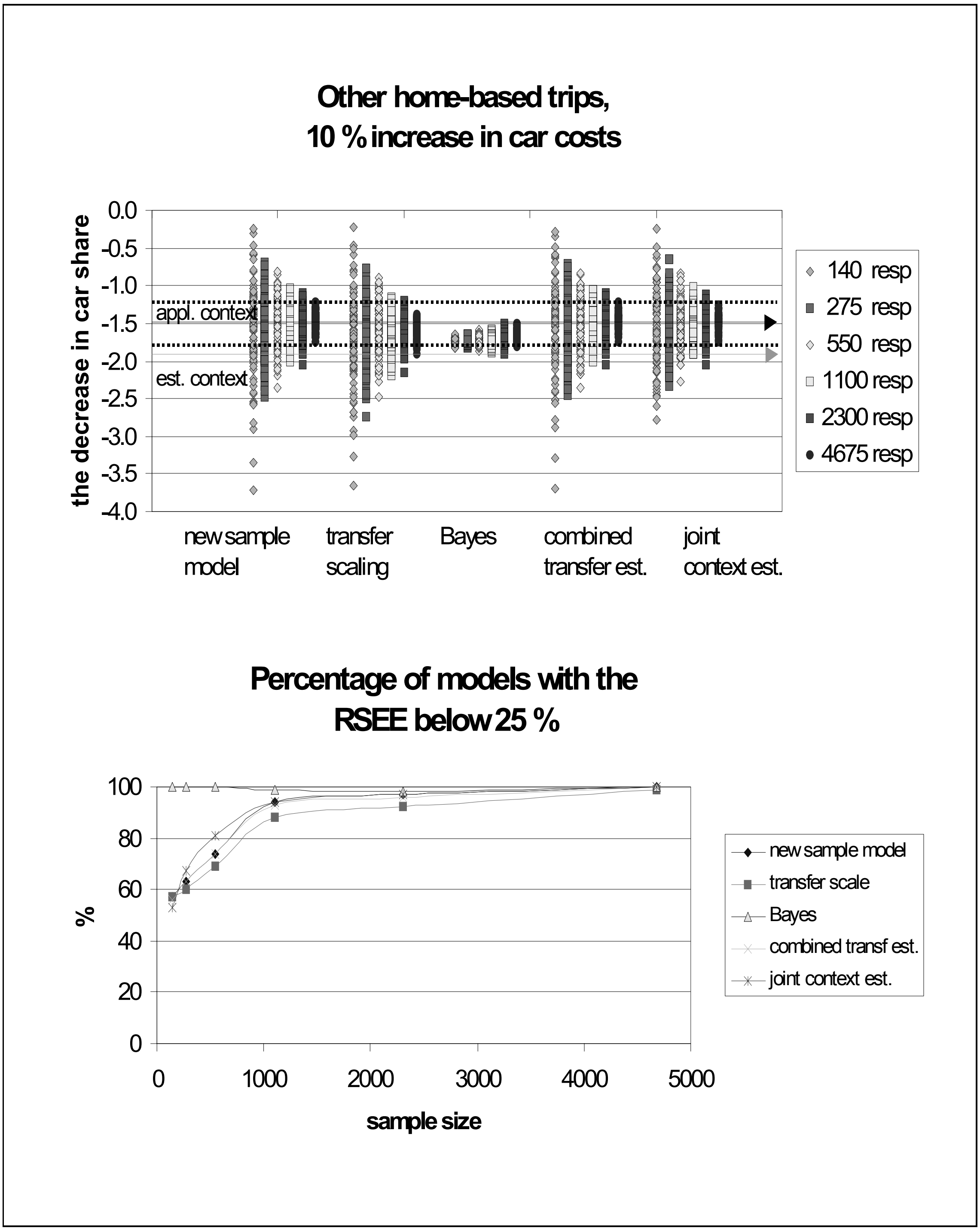
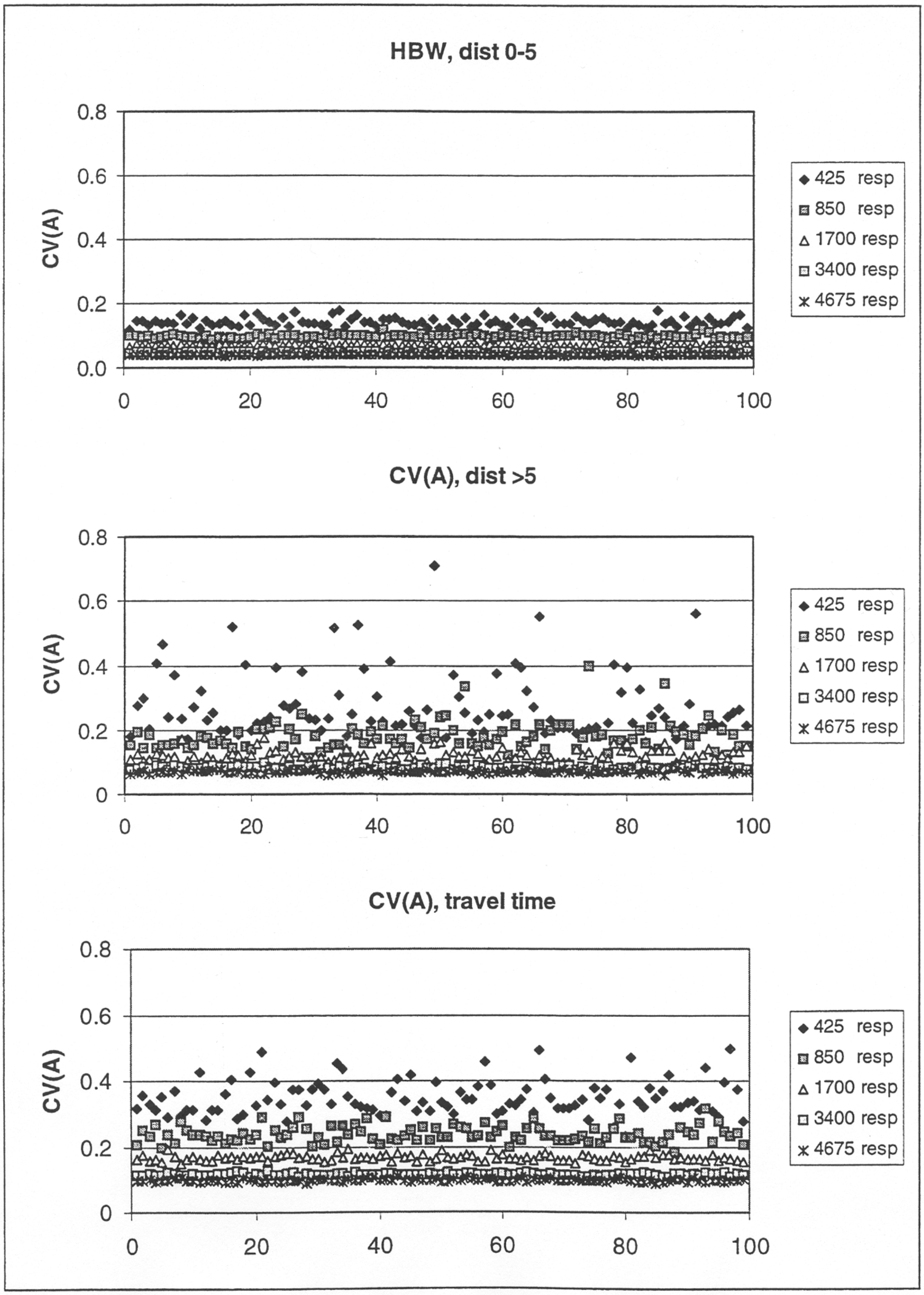
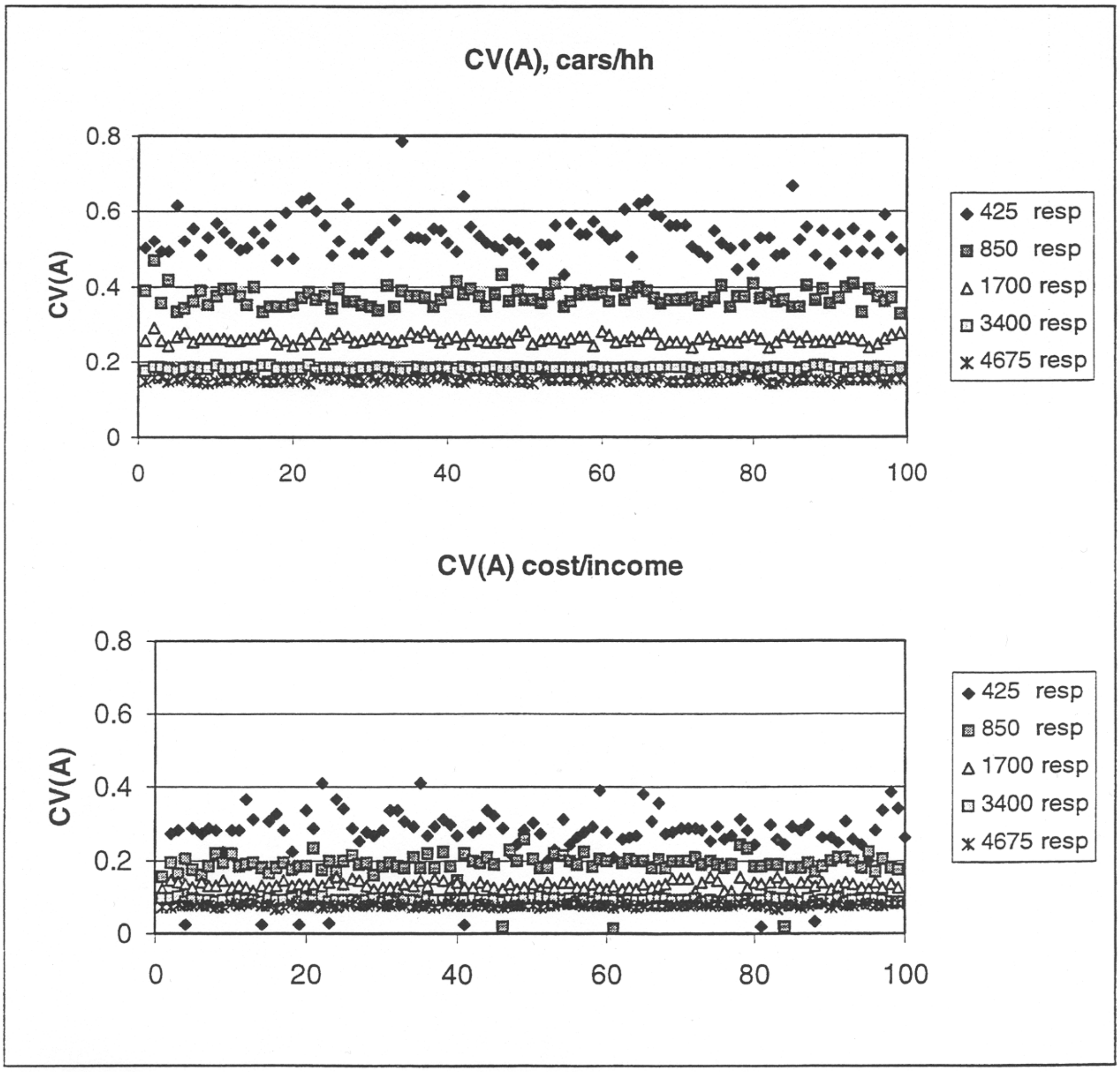


Figure O.2: *The variation of change in car shares by transfer method, when car costs are increased by 10 percent in the Turku region; other home-based trips*

APPENDIX P

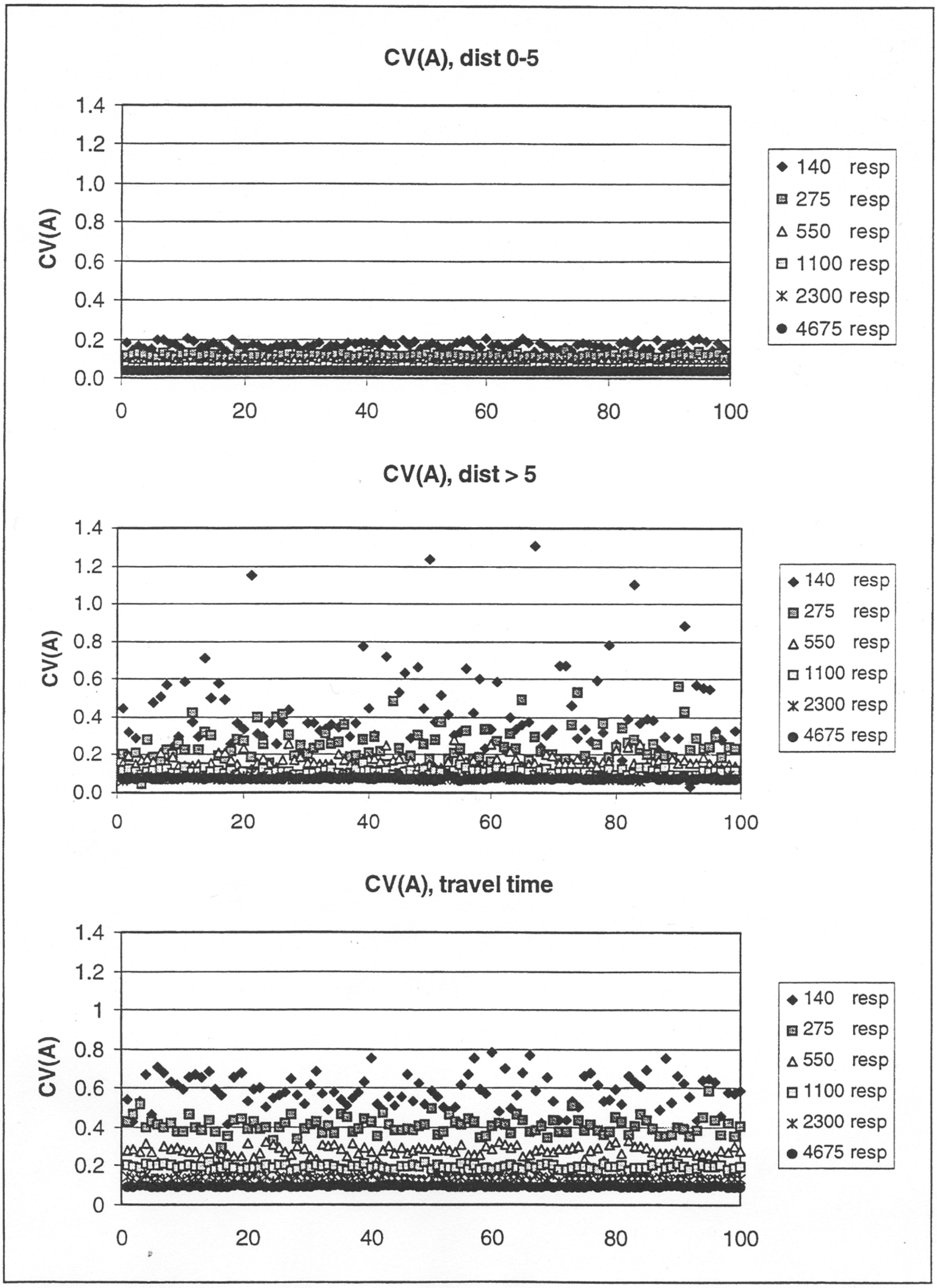
The variation of CV(A) in home-based work trips

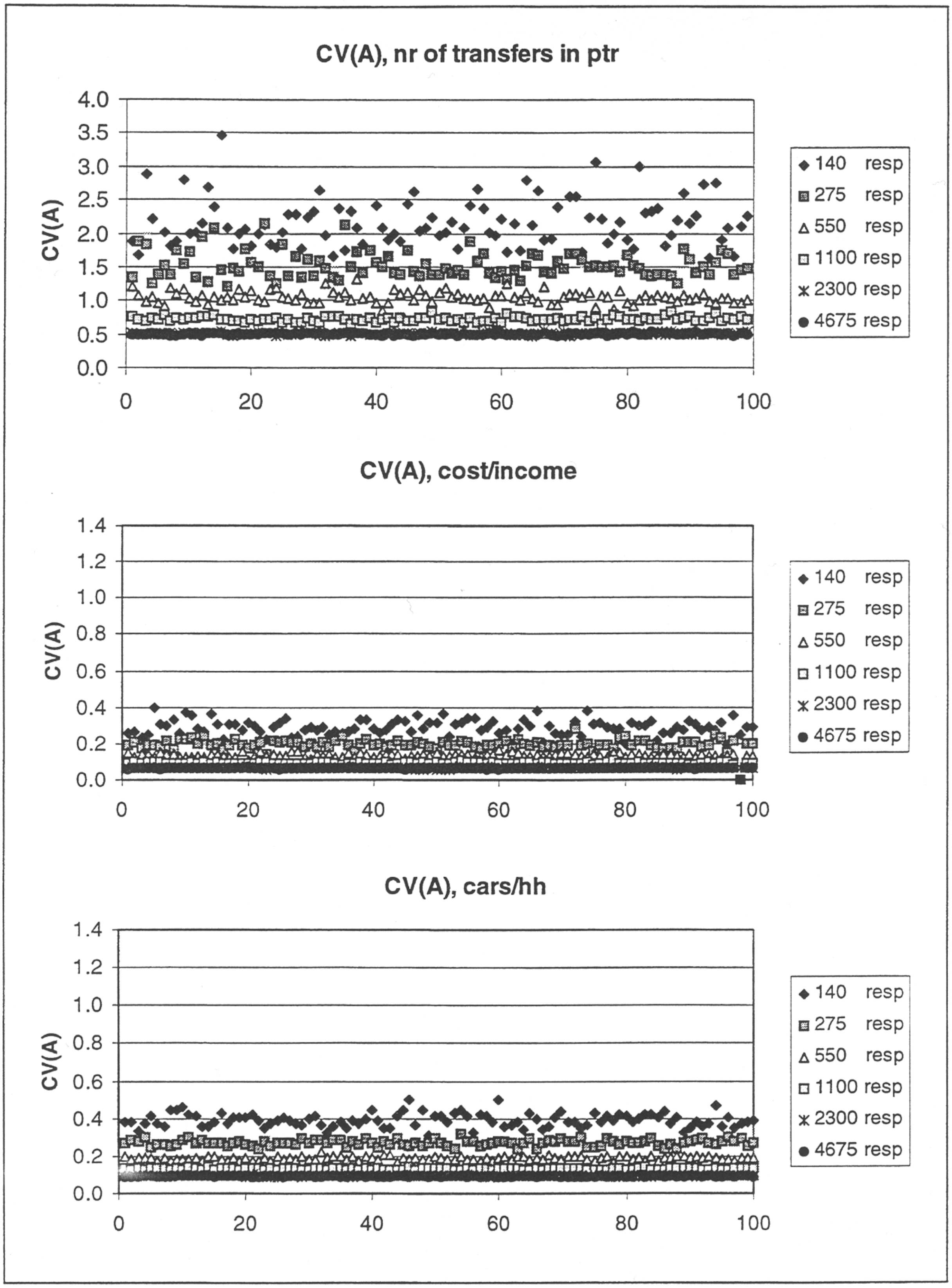



CV(A), cars/hh
CV(A)
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
425 resp
850 resp
1700 resp
3400 resp
4675 resp
CV(A) cost/income
CV(A)
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
425 resp
850 resp
1700 resp
3400 resp
4675 resp

APPENDIX Q

The variation of CV(A) in other home-based trips




CV(A), nr of transfers in ptr
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
CV(A)
0
20
40
60
80
100
140 resp
275 resp
550 resp
1100 resp
2300 resp
4675 resp
CV(A), cost/income
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
CV(A)
0
20
40
60
80
100
140 resp
275 resp
550 resp
1100 resp
2300 resp
4675 resp
CV(A), cars/hh
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
CV(A)
0
20
40
60
80
100
140 resp
275 resp
550 resp
1100 resp
2300 resp
4675 resp