2012 | Faculteit Bedrijfseconomische Wetenschappen

DOCTORAATSPROEFSCHRIFT

Bayesian Approaches for Origin-Destination Modeling and Traffic Assignment Inference

Proefschrift voorgelegd tot het behalen van de graad van doctor in de verkeerskunde,te verdedigen door:

Konstantinos Perrakis

Promotor: prof. dr. Davy Janssens
Copromotor: prof. dr. Dimitris Karlis (AUEB Athene, Department of Statistics)

D/2012/2451/35

universiteit hasselt
KNOWLEDGE IN ACTION

Acknowledgements

At this final phase of my doctoral study I want to express my gratitude to all the people who contributed to this research and also supported me along the way, each one in his or her own particular way.

First of all, I would like to kindly thank my promotor prof. dr. Davy Janssens. I thank him for providing me with the opportunity to work on this interesting research topic and for constantly making clearer to me and reminding me the overall research framework and perspective of my PhD. Next, I want to express my sincere gratitude to my co-promotor prof. dr. Dimitris Karlis. Firstly, for encouraging and supporting me to pursue this PhD. Secondly, for his precious scientific guidance, without which I would not have been able to cope successfully with what was expected from this doctoral research. I also cordially thank prof. dr. Mario Cools who was always there for me, to answer thousands of questions, to provide insightful guidance and to support me in all possible ways.

Furthermore, I am grateful for the contributions made by prof. dr. Geert Wets and prof. dr. Koen Vanhoof who kept track of this doctoral research from day one and provided constructive comments and ideas. I also want to acknowledge prof. dr. ir. Tom Bellemans and the external member of the jury, prof. dr. Harry Timmermans, for their careful reviews and comments.

Moreover, I would like to extend my appreciation to all colleagues at IMOB for creating an excellent working environment. Particularly, I want to thank Kristel Hertogs who helped me tremendously with all practical, every-day issues from the very first day I arrived in Belgium until the very last days of my PhD, Dirk Roox who provided me with the TransCAD macros and saved me from spending immense time and efforts, and Nadine Smeyers who took the kind effort of translating the summary of this dissertation in Dutch for me.

During these four years, colleagues from IMOB and Hasselt University became good and close friends. Some are still working at IMOB and Hasselt University, others are not. Regardless, all of them became, still are and always will be dear friends to me. So, I like to thank, warmly, Diana Kusumastuti, George Sammour, Ali Pirdavani, Hans Tormans, Mario Cools, Feng Liu, Panos

Gkorezis, Vivi Kolokytha, Wouter Sillen and Francois Rineau for all the good times we shared together.

My warmest gratitude goes to my parents, Matthaios and Lotte, in Greece and to my elder brother, Markos, and his family in Canada. Their presence, even from a distance, was a constant source of support. Finally, I thank Myrto. My friend and companion in life for the last 8 years. I thank her for her love, patience, understanding and joyfulness.

Konstantinos Perrakis
September 24, 2012

Summary

This dissertation is related to the topic of origin-destination (OD) analysis and to relative aspects surrounding this topic within the domain of transportation. In essence, an OD matrix summarizes the travel-demand of a given geographical area and is the basic input to the last modeling phase of any sequential travel-demand forecasting model which involves a traffic assignment procedure. The core of this dissertation is a new methodological approach to OD modeling. The approach is statistical, based on Bayesian Poisson mixture modeling, and can be viewed as a modern direct-demand/gravity modeling framework which incorporates the first two phases of a four-step model. Bayesian methods provide an appropriate working framework for quantifying uncertainties related to parameters as well as predictions of short-term traffic.

The proposed methodology is implemented on OD data derived from the 2001 Belgian census study. The study focuses on OD movements between the 308 municipalities of Flanders. A set of 25 explanatory is used to model expected OD trips under the assumption of a log-link function. Initially, a comparison is carried out between a Poisson model and a negative binomial model which indicates that the latter is noticeably a more suitable model due to extreme over-dispersion. Emphasis is placed on the hierarchical negative binomial structure, which is a Poisson-gamma (PG) model with random effects accounting for heterogeneity. The PG model is further compared with a Poisson-lognormal (PLN) and a Poisson-inverse Gaussian (PIG) model. In this first full Bayesian PIG application it is shown that the model has desired distributional properties. In addition, the PIG model provides the best marginal fit. Concerning parameter significance and interpretation, all explanatory variables have statistically significant effects which have consistent interpretations with respect to transportation modeling expectations.

Moreover, the proximity of PG and PIG predictions to the observed data is evaluated according to several measures of discrepancy. The overall fit is found in general to be satisfactory. One important finding is that both models tend to underestimate the number of zero-valued OD pairs. Although replicating the number of zero-valued cells is not one of the primary goals of the analysis, it is shown that zero-valued cells can have a strong cumulative influence on total

travel-demand. In general, one of the advantages of using Bayesian methods is that one can predict the short-term distribution of any type and/or combination of trips that are of interest. This provides predictive distributions which are particularly useful in transport planning and policy evaluation.

Subsequent research focuses on traffic-assignment and network-congestion inference. The methodology is based on utilizing the predictive output of the models as input to traffic assignment. Specifically, two methods of inputting OD predictions are discussed. In the first method an OD summary is calculated first and then assigned to the network, whereas in the second method all OD predictions are assigned to the network individually. Method 1 leads to approximate-network inference and is computationally less demanding, but not as exact as method 2. In general, method 2 is promoted and advocated as it provides a suitable tool for full-network inference regarding point and interval estimates, link flow distributions and identification of congested links by means of probability estimates.

The methods are compared on the Flemish road network for traffic concerning going-to-work/school trips by Flemish residents between the peak hour from 7 am to 8 am. Initial results, based on deterministic user equilibrium (DUE) assignment, indicate that traffic flows in Flanders are denser around the major municipal centers of Antwerp, Ghent, Leuven and Bruges, and on the highways which connect these cities with each other and also with Brussels. In addition, eleven congested links are identified through method 2 as having a non-zero probability of exceeding a volume-over-capacity (V/C) threshold value of 0.95, the majority of which belonging to Antwerp and Ghent. Contrary, when using the expected V/C ratio as an identification criterion, four of these links are not identified. In general, the comparison between the two inputting methods provides some initial evidence that method 1 might be suitable when the sole goal is to have a point estimate of the expected state of the network. Specifically, in relation to the behavior of total system travel time (TSTT) and Jensen's inequality, the estimate from method 1 is found to be indeed smaller than the estimate from method 2 in accordance to theory. Nonetheless, in practical terms the two estimates are relatively close. Regarding percentile estimates, estimates from method 1 result in interval estimates which are evidently narrower and thus fail to capture the complete variability which is induced by travel-demand uncertainty. Additional comparisons between PG and PIG predictions reveal that the choice of the statistical model also has a certain influence concerning inference for aggregated link flow distributions. Nevertheless, main inferences concerning the behavior of TSTT and V/C ratios are not affected.

Further traffic assignment experiments are implemented next by comparing results between DUE assignment and stochastic user equilibrium (SUE) assignment under both probit and logit route-choice models and for different values of perception-error variance. These comparisons are conditional on PG predictions. Results concerning TSTT are again in agreement with the theory related to Jensen's inequality. Results for aggregated link volumes are less straightforward to interpret, nevertheless, some general conclusions are in accordance with theoretical expectations. First, DUE assignment allocates more traffic to high-capacity links, while SUE assignment allocates more traffic to medium-capacity links. Second, when considering the total amount of traffic, SUE assignment produces more traffic than DUE and in addition traffic under SUE increases as error-perception variance increases. Regarding congestion analysis, results reveal that the selection of assignment model does not seriously affect the general allocation of links in relation to expected V/C ratios. On the other hand, variability is present when examining individual V/C distributions. In addition, under SUE assignment bimodal as well as multimodal V/C distributions arise. These results grant additional support to the use of probability estimates as opposed to centrality estimates as indicators of congestion.

Finally, the recently developed radiation model is discussed in certain detail due to its strong relevance with this current research. Initial attempts of assimilating the radiation model within the modeling framework, considered here, are illustrated. Initially, the variable of circular area population – introduced in the radiation model – is used as an explanatory variable with negative binomial likelihood assumptions. In addition, a first possible Bayesian extension of the radiation model is discussed.

Samenvatting

Dit proefschrift gaat over de schatting van herkomst-bestemmingsmatrices (origin-destination (OD) matrices) en de hieraan gerelateerde aspecten in het domein van vervoer. In essentie bevat een OD-matrix een overzicht van de vervoersvraag van een bepaald geografisch gebied en behoort het tot de basisinput voor de laatste modelleringsfase van elke sequentieel vervoersvraagprognosemodel dat gebruik maakt van verkeerstoedeling. De kern van dit proefschrift is een nieuwe methodologische benadering voor het schatten van OD-matrices. Het is een statistische aanpak, gebaseerd op het Bayesiaanse Poisson mixed model, dat beschouwd kan worden als een modern directe vraagmodel dat de eerste twee fasen van een vier-stapsmodel bevat. Bayesiaanse methoden bieden een passend kader voor het kwantificeren van onzekerheden met betrekking tot parameters alsook voor korte-termijn verkeersvoorspellingen.

De voorgestelde methodologie werd toegepast op OD-gegevens, afkomstig van de Belgische census gehouden in 2001. In dit doctoraat is de analyse gericht op OD-relaties tussen de 308 Vlaamse gemeenten. Er wordt een set van 25 verklarende variabelen gebruikt om verwachte OD-trips te modelleren, onder de veronderstelling van een log-linkfunctie. Eerst wordt een Poisson model vergeleken met een negatief binomiaal model. Hieruit blijkt duidelijk dat het laatste model geschikter is omwille van extreme overdispersie. De nadruk wordt gelegd op de hiërarchisch negatieve binomiale structuur. Dit is een Poisson-gamma (PG) model met random effecten die heterogeniteit vertegenwoordigen. Het PG-model wordt verder vergeleken met een Poisson-logaritmisch (PLN) en een Poisson-Invers Gaussiaans (PIG) model. In deze eerste volledige toepassing van het Bayesiaanse PIG model wordt aangetoond dat dit model gewenste distributie-eigenschappen heeft. Bovendien biedt het PIG-model de beste marginale fit. Wat het belang en de interpretatie van de parameters betreft, hebben alle verklarende variabelen statistisch significante effecten waarvan de interpretatie overeenkomt met de verwachte vervoersmodelleringen.

Bovendien wordt de benadering van PG- en PIG-voorspellingen op de waargenomen gegevens geëvalueerd volgens verschillende afwijkingsgradaties. De algemene fit is meestal bevredigend. Een belangrijke vaststelling is dat beide

modellen de neiging hebben om het aantal OD-paren met nulwaarden te onderschatten. Hoewel het niet de belangrijkste doelstelling van de analyse is om het aantal cellen met nulwaarde te herhalen, werd aangetoond dat cellen met een nulwaarde een sterk cumulatieve invloed kunnen hebben op de totale vervoersvraag. Over het algemeen is één van de voordelen van de Bayesiaanse methoden dat men het korte-termijn type en/of de combinatie van trips waarin men geïnteresseerd is kan voorspellen. Dit levert voorspellingen op met betrekking tot distributies die vooral handig zijn bij vervoersplanning en beleidsevaluatie.

Verder onderzoek is gericht op verkeerstoedeling en de deductie van netwerkcongestie. In deze methode wordt de voorspellingen van de bovenvermelde modellen gebruikt als input voor de verkeerstoedeling. Meer specifiek worden er twee methoden besproken, die gebruik maken van de OD-voorspellingen. In de eerste methode wordt er eerst een samenvattende OD-matrix berekend en vervolgens wordt dit toegewezen aan het netwerk. In de tweede methode daarentegen, worden alle OD-voorspellingen afzonderlijk aan het netwerk toegewezen. Methode 1 leidt tot een deductie van het geschatte netwerk en is makkelijker te berekenen, maar is niet zo accuraat als methode 2. Over het algemeen wordt methode 2 aanbevolen aangezien het geschikt is voor de deductie van een volledig netwerk met betrekking tot punt- en intervalschattingen, linkflowdistributies en de identificatie van overbelaste links door middel van waarschijnlijkheidsschattingen.

De methoden worden vergeleken op het Vlaamse wegennet voor het woon-werkverkeer en woon-schoolverkeer, gedaan door Vlamingen tussen het piekuur van 7u tot 8u. Uit de eerste resultaten, bekomen door een deterministische gebruikersevenwichttoewijzing (DUE), blijkt dat er in Vlaanderen grotere verkeersstromen zijn rond de belangrijkste stedelijke centra Antwerpen, Gent, Leuven en Brugge, en op de snelwegen die deze steden met elkaar en met Brussel verbinden. Bovendien worden er door middel van methode 2 elf overbelaste links geïdentificeerd. Deze links hebben een kans groter dan nul om de drempelwaarde van 0,95 van volume over capaciteit (V/C) te overschrijden. De meerderheid hiervan behoort tot Antwerpen en Gent. Wanneer men daarentegen de verwachte V/C-verhouding gebruikt als een identificatiecriterium worden vier van deze links niet geïdentificeerd. Over het algemeen wijst een vergelijking tussen de twee inputmethoden een eerste bewijs dat methode 1 geschikt zou kunnen zijn wanneer de enige doelstelling erin bestaat om een puntschatting te verkrijgen van de verwachte staat van het netwerk. In het bijzonder met betrekking tot het gedrag van de totale reistijd (TSTT) en Jensens ongelijkheid, blijkt de schatting van methode 1 inderdaad kleiner te zijn dan de schatting van methode 2, wat in overeenstemming is met de theorie. In de praktijk liggen de twee schattingen echter relatief dicht bij elkaar. Met betrekking tot percentuele schattingen, leveren de schattingen van methode 1 intervalschattingen op die blijkbaar kleiner zijn en daarom niet de volledige breedte aan variatie bevatten die eigen is aan de vervoersvraagonzekerheid. Bijkomende vergelijkingen tussen PG- en PIG-predicties tonen aan dat de keuze van het statistische model ook een zekere invloed heeft op de deductie van geaggregeerde linkflowdistributies. Niettemin worden de belangrijkste deducties met betrekking tot het gedrag van TSTT en V/C-ratio niet beïnvloed.

Vervolgens worden bijkomende verkeerstoedelingsexperimenten uitgevoerd door de resultaten van de DUE-toedeling en de stochastische gebruikersequilibrium (SUE)-toedeling te vergelijken onder zowel probit als logit routekeuzemodellen en voor verschillende waarden van variatie in perceptiefouten. Deze resultaten zijn gebaseerd op de PG-voorspellingen. De TSTT-resultaten komen opnieuw overeen met Jensens theorie van ongelijkheid. De resultaten voor geaggregeerde linkvolumes zijn minder gemakkelijk te interpreteren. Niettemin stemmen enkele algemene conclusies overeen met de theoretische verwachtingen. Ten eerste wordt bij DUE-toewijzing meer verkeer toegewezen naar links met een hoge capaciteit, terwijl bij SUE-toewijzing meer verkeer toegewezen wordt naar links met een gemiddelde capaciteit. Wanneer men ten tweede de totale hoeveelheid verkeer in acht neemt, produceert een SUE-toewijzing meer verkeer dan een DUE-toewijzing en bovendien neemt het volume van verkeer onder SUE-toewijzing toe naarmate de foutperceptievariantie toeneemt. Wat congestieanalyse betreft wijzen resultaten uit dat de keuze van het toewijzingsmodel geen grote impact heeft op de algemene toewijzing van links met betrekking tot verwachte V/C ratio's. Anderzijds is variabiliteit aanwezig bij individuele V/C distributies. Bij SUE-toewijzing ontstaan bovendien zowel bimodale als multimodale V/C distributies. Deze resultaten geven een extra stimulans voor het gebruik van waarschijnlijkheidsschattingen in plaats van gemiddeldeschattingen als congestie-indicator.

Tot slot wordt het onlangs ontwikkelde radiatiemodel omwille van het sterke verband met dit doctoraatsonderzoek tot op een bepaald niveau van detail bediscussieerd. Eerste pogingen om het radiatiemodel te assimileren binnen het modelleringskader van dit proefschrift worden geïllustreerd. In eerste instantie wordt de "cirkelvormige zonebevolking"-variabele – geïntroduceerd in het radiatiemodel – gebruikt als een verklarende variabele met negatief binomiale waarschijnlijkheidsveronderstellingen. Bovendien wordt een eerste mogelijke Bayesiaanse uitbreiding van het radiatiemodel besproken.

List of Abbreviations

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BPR	Bureau of Public Roads
BRUE	Bounded Rational User Equilibrium
BUE	Behavioral User Equilibrium
CA	Cumulative Average
DIC	Deviance Information Criterion
DUE	Deterministic User Equilibrium
EM	Expectation Maximization
GIG	Generalized Inverse Gaussian
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
HGLM	Hierarchical Generalized Linear Model
HPD	Highest Posterior Density
HW	Highways
LR	Local Roads
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MH	Metropolis-Hastings
ML	Maximum Likelihood
MRR	Main Regional Roads
OD	Origin-Destination
PG	Poisson Gamma
PIG	Poisson Inverse Gaussian
PL	Pseudo Likelihood
PLN	Poisson Log-Normal
QL	Quasi Likelihood
REML	Restricted Maximum Likelihood
RMSE	Relative Mean Square Error
SRR	Small Regional Roads
SUE	Stochastic User Equilibrium
TSTT	Total System Travel Time

UE	User Equilibrium
V/C	Volume-over-Capacity
VIF	Variance Inflation Factor

Contents

List of Figures

List of Tables

1 Introduction

The ultimate aim of transportation analysis and planning is to monitor traffic and to ensure that travel needs of individuals are being satisfied, unimpeded, within a given transport network. In terms of economics this translates to a state of balanced equilibrium between *travel-demand* and *travel-supply*. From this perspective, travel-demand is perceived as something that is being *derived* by the existence of certain needs and not as something with an end in itself. Movement of goods is clearly subjected to this line of reasoning and one can additionally reason that for most cases individuals also travel to a certain destination in order to do a certain activity or fulfill a certain need, for instance work, school, leisure and so forth[1]. Travel-supply on the other hand refers to the operating capacity of a transport system which is defined by a given infrastructure (e.g. road network), a management system (e.g. driving rules, traffic signs and signals) and a set of transport modes and their corresponding operators (e.g. bus, rail services).

Although demand and supply are in principle not independent and as in any economic environment there exist intricate two-way interactions and dynamics in between, one can note that demand is subjected to extreme spatial and temporal short-term variations while supply is on short-term relatively constant and is characterized by changes that occur during long-term periods. Thus, transportation planners focus, primarily, on obtaining a best possible estimate of travel demand and address, subsequently, the appropriateness of the supply-side given the demand estimate[2]. When the system is in balance, the supply satisfies and potentially widens the opportunities for fulfilling the needs of individuals and other agents. Contrary, a system which is in imbalance (e.g. a congested or poorly connected system) restricts options and delays or limits economic and social activities, and thus overall development (Ortúzar and Willumsen, 2001).

[1] The "derived demand" concept based on economic principles implies that the act of travelling itself does not offer any positive utility and that minimizing travel time is the objective aim of any traveler. A discussion over the overall validity of this notion would be beyond the scope of this dissertation. Interested readers are referred to Ory (2007) and the references therein for recent, alternative approaches which partially challenge the "derived demand" point of view.

[2] It should be noted that supply-uncertainty modeling is also gradually gaining attention; see for instance the studies of Siu and Lo (2008), Boyles et al. (2010) and Gardner et al. (2011).

This dissertation, in its essence, explores ways of modeling and quantifying travel demand uncertainties, and incorporating these uncertainties in the assessment of travel supply performance. This introductory chapter starts with a brief description of the overall travel demand framework and with an overview of existing travel demand modeling approaches. In the following sections, the concept of uncertainty in transport modeling is discussed and the issue of availability of OD data is commented. The overall scope, aims and setting of this dissertation are presented next. The chapter concludes with an outline of the dissertation.

1.1 Problem statement

In transportation analysis the travel demand within any geographical area, dividable into a given number of non-overlapping zones, can be summarized by an OD matrix which – in a very general context – contains the number of *trips* or *flows*[3] that have occurred from each zone of that area to every other zone. Consider an area which can be divided into m zones, and let T_{od} denote the number of flows from zone of *origin* o to zone of *destination* d where $o, d = 1, 2, \ldots, m$. The OD matrix **T**, is then

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1m} \\ T_{21} & T_{22} & \cdots & T_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ T_{m1} & T_{m2} & \ldots & T_{mm} \end{bmatrix}.$$

The elements T_{od} for $o \neq d$, i.e. when the zones of origin and destination are distinct, correspond to *inter-zonal* flows, whereas the elements across the main diagonal T_{oo} for $o = d$ correspond to *intra-zonal* flows as in this case there is no distinction between zone of origin and zone of destination. In vector notation the matrix **T** can be represented by a n-dimensional vector **y** with elements y_i for $i = 1, 2, \ldots, n$ and $n = m^2$, namely $\mathbf{y} = (y_1, y_2, y_3, \ldots, y_n)^T = (T_{11}, T_{12}, T_{13}, \ldots, T_{mm})^T$.

[3] The terms "flows" and "trips" are used interchangeably throughout the dissertation, depending on context.

The intra-zonal flows under this notation are the elements $y_1, y_{m+2}, y_{2m+3}, y_{3m+4}, \dots, y_{(m-2)m+(m-1)}, y_n$.

The inferential scope in OD estimation is depended on several defining aspects. A first aspect is related to the desired level of spatial resolution for a given geographical area which in turn depends on the definition of the zonal unit-scale. This means that when the geographical area of interest is determined, the next step is to define on which zone-level to implement the analysis, since it is usually possible to define several nested spatial zone layers according to natural or administrative divisions, e.g. a city can be divided in districts, municipalities, postal-code areas and so forth. Therefore, the problem can associate to micro, meso or macro-scopic levels of spatial resolution[4]. A second defining aspect is related to temporal resolution, which from a transportation planning perspective can be classified in short-term and long-term travel-demand cases. Short-term travel-demand analysis is focused on OD flows which occur on relatively small time intervals, for instance hourly or daily flows, whereas long-term travel-demand analysis is related to coarse time frames, e.g. monthly or even yearly intervals. For the latter case usually static estimation approaches are used, while in short-term travel-demand analysis static as well as dynamic modeling procedures are employed. A third inferential aspect relates to the classification of OD flows by trip-purpose. The OD flows may refer to total aggregated trips, trips related to working, educational, leisure, freight and other activities or to combinations of different types of activities. Model formulation and selection of explanatory variables will depend to a large degree on the type of flows which are under consideration. Within the field of transportation, modeling applications include all of the aforementioned cases depending on context, data availability and inferential scope. Interested readers are referred to Ortúzar and Willumsen (2001) for more information.

It is essential to note that transport modeling is typically sequential and within the general demand modeling framework OD estimation is itself part of a larger inferential problem. The final phase of the sequential procedure relates to

[4] Although this is a common complexity arising in problems related to spatial analysis, OD estimation differs from the majority of problems which arise within this field of research. The common case in spatial analysis applications is that counts of occurrences are observed within each spatial zone. In these cases analysis takes place on the m-dimensional space of zones, whereas in OD estimation analysis is required on the n-dimensional space of OD pairs.

a problem of major interest which is known as *traffic assignment*. Traffic assignment involves allocating the $n-m$ inter-zonal OD flows on a corresponding transport network consisting of all the available *links* which define the possible *routes* from zone of origin o to zone of destination d for $o, d = 1, 2, \ldots, m$ when $o \neq d$. In vector notation the traffic assignment problem can be broadly described as follows; if we denote by l the total number of links, then traffic assignment seeks an operator **A** which will map the inter-zonal OD flows from the $(n-m)$-dimensional vector space of OD pairs to the l-dimensional vector space of links and result to a vector of link flows or volumes $\mathbf{v} = (v_1, v_2, \ldots, v_l)^T$, i.e. $\mathbf{Ay} = \mathbf{v}$, by minimizing an objective function of travel time with respect to link flows. The operator **A** is commonly non-linear. Put into practice this is a complicated problem and several traffic assignment models have been developed over the years each imposing specific assumptions on the objective function to be minimized. In summary, traffic assignment models take as input the structure of a transportation network (travel-supply) and a corresponding OD matrix (travel-demand). The output itself depends on the complexity of the assignment model; nevertheless all assignment models produce a minimal output which includes traffic volumes and corresponding travel times or costs for each network link (Patriksson, 1994). The major aims of assignment models as summarized in Mathew and Rao (2007) are the following:

1. To estimate the volume of traffic on the links of the network and obtain aggregate network measures.
2. To estimate inter-zonal travel cost.
3. To analyze the travel pattern of each OD pair.
4. To identify congested links and to collect traffic data useful for the design of future junctions.

Traffic assignment is in general the final goal of any transportation planning procedure as it provides a basis for planning and investment decisions concerning infrastructure and transport policy measures, e.g. road expansions or closures, construction of new roads, re-routing schemes, toll pricing policies and many other issues.

1.2 OD modeling approaches

Literature related to OD modeling either exclusively or as part of a general travel demand framework is vast and immensely diverse. The bridging of scientific disciplines and the rapid evolution of computers facilitated a gradual and quick mixing of transportation engineering and economic concepts with elements from diverse scientific fields such as behavioral psychology, statistics and computer science. It would be difficult – if not impossible – to attempt to define a simple, non-overlapping classification of OD modeling approaches based on purely methodological terms. The classification which is commonly adopted (e.g. Rasouli and Timmermans, 2012) relates more to the historical evolution and philosophical perspective of travel demand forecasting and less to specific methodological tools which are utilized under each approach.

From a historical point of view three generations of travel demand models are distinguishable. The first generation of models has its origins in the late 1950's. Under this approach travel demand is modeled according to a *four-step* modeling sequence. Four-step models are aggregated, *trip-based* models, i.e. they refer to total number of trips and do not take into account the sequencing or chaining of trips made by specific individuals. The four-step model is discussed in more detail in comparison to the following modeling approaches as this model relates more to the line of research pursued here. Nevertheless, an analytic description would exceed the scope of this dissertation since each modeling step is a separate research category by itself. Therefore, attention is restricted to a brief description of the 4-step modeling procedure and to some basic references related to OD estimation. Interested readers are referred to the books of Ortúzar and Willumsen (2001), and Hensher and Button (2000) for more information.

The four modeling steps are (a) trip-generation, (b) trip-distribution, (c) modal-split and (d) traffic-assignment. OD estimation is treated in steps (a) and (b), the subsequent step (c) involves disaggregating the OD matrix with respect to mode choice and step (d), as described previously, is essentially the final goal of the sequential procedure. The first step of trip-generation consists of estimating the marginal totals $O_o = \sum_d T_{od}$ and $D_d = \sum_o T_{od}$, for $o, d = 1, 2, \ldots, m$, which are referred to as *trip-productions* and *trip-attractions*, respectively. In this phase regression models or cross-classification models are typically used to estimate trip-productions and trip-attractions as functions of socio-economic, location and land-use characteristics. The second step, trip-distribution is the step in which an OD estimate is obtained by distributing the marginal totals to the cells of the matrix. Models used within this step date back decades ago and include simple *growth-factor* models for short-term OD estimation, *gravity* models which were the first models to relate trip-distribution with external socio-economic factors and also distance, *intervening-opportunities* models which related trip-distribution not so much to distance as to the relative accessibility of opportunities for satisfying a trip objective, and finally *direct-demand* models which were developed in order to incorporate steps (a), (b) and (c). The first application of a gravity model was in Casey (1955). Furness (1965) provided a significant contribution by introducing “balancing” factors in growth-factor and gravity modeling. Stouffer (1940) and Schneider (1959) developed the intervening-opportunities model, while Kraft (1968), Domencich et al. (1968) and Manheim (1979) were among the first to use direct-demand modeling. Finally, Wilson (1970, 1974) provided one of the most influential contributions in transportation modeling by introducing an entropy-maximizing perspective which included growth-factor, gravity and intervening-opportunities models as special cases. The approach of Wilson linked OD estimation to information theory, error measures and maximum-likelihood. In general, the sequential four-step procedure remains widely accepted by transportation planners so that many applications up to the present are still very much based on the principles of the aforementioned trip-distribution models[5].

The second generation of models appeared around the mid 1970’s and marked a transition from aggregated to disaggregated modeling as the focus of attention shifted to the needs of individual travelers. This approach is based on utility-maximization and individual-choice behavior theories and is known as *discrete-choice* modeling. It is characterized also as a *tour-based* modeling

[5] On the other hand, four-step modeling has also been subjected to significant criticism, mainly due to the assumption of independent modeling at each step and also due to the vagueness concerning implementation of feedback loops between steps. In a very interesting paper by Boyce (2002), which also provides an overview of the historical evolution of four-step modeling, it is argued that an integrated approach combining steps (b), (c) and (d) is far more realistic (see also the references therein). Another criticism relates to propagation of error due to spatial, demographic and temporal aggregations (Walker, 2005; Davidson et al., 2007).

approach as individual trips are explicitly connected in tours, i.e. chains of trips that start and end at a given location. Multinomial regression models dominate this field with the first transport applications of logit, hierarchical logit, probit and nested logit regression models being by Domenchich and McFadden (1975), Williams (1977), Daganzo (1979) and Ben Akiva (1974), respectively. An analytic overview of discrete-choice models including recent and more complex modeling approaches can be found in Hess (2005). Information can also be found in Ben Akiva and Bierlaire (1999) and Ortúzar and Willumsen (2001).

The third generation of travel demand forecasting models known as *activity-based* models appeared more or less in parallel with discrete-choice models[6]. However, the development of activity-based models was boosted during the 1990's. The underlying premise of activity-based modeling is that traveling is derived from activities and thus it should be understood within the context of activity participation. Theoretically, a full activity-based model delivers significant additional information in comparison to a trip-based and a tour-based model, namely activity participation, destination choice, time, duration, mode choice and route choice. The term "framework" might be more suitable to describe an activity-based approach, as these models typically embrace a series of submodels for activity type, activity duration, destination and mode choice modeling. The common characteristic of all activity-based models is that they employ *agent-based* or *micro-simulation* either in principle or in order to link independent or weakly-joined submodels. Examples of comprehensive activity-based models include CEMDAP (Bhat et al., 2004) which utilizes a series of independent econometric submodels, FAMOS (Pendyala et al., 2005) with nested logit submodels, ALBATROSS (Arentze and Timmermans, 2004) a strongly linked rule-based model which is also the core of FEATHERS (Bellemans et al., 2010), and TASHA (Roorda et al., 20078) which combines ad hoc rules, sampling approaches and simple discrete-choice models for the various lower level models. A literature review on activity-based models is

[6] Although activity-based modeling is in general regarded as a newer trend in comparison to discrete-choice modeling, the first actual study in which activities and travel behavior were integrated is in Jones et al. (1983), very close to the time that the first discrete-choice modeling approaches appeared.

provided by Henson et al. (2009), a general discussion over microsimulation models can be found in Vosha et al. (2002)[7].

Certain remarks and general comments concerning the aforementioned classification are noteworthy. First, the classification is still overlapping to some extent. Although the distinction between four-step and activity-based modeling is quite clear, the positioning of discrete-choice modeling is rather ambiguous. That is due to the fact that discrete-choice models are in general difficult to classify. In many occasions, discrete-choice models gradually replaced the trip-generation and trip-distribution phases of four-step models marking a shift from trip to tour-based modeling which was in some occasions, perhaps prematurely, named “activity-based” modeling (Rasouli and Timmermans, 2012). In addition, discrete-choice modeling is the main methodological tool for many modern activity-based models. Information on discrete-choice modeling within activity-based frameworks can be found in Bhat and Koppelman (1999). Second, it should be noted that all travel demand forecasting paradigms are sequential and therefore concerns related to properly defined feedback loops between steps or models should not constrict to the four-step model. Activity-based models often utilize numerous submodels in order to predict travel demand and in addition still depend on conventional assignment models for traffic allocation. Although, the ultimate vision concerns fully integrated approaches which will include traffic assignment procedures, up to the present such approaches are still on a conceptual or experimental stage (see e.g. Lin et al., 2008). Third, it is undoubtedly true that the gradual shift from trip to tour to activity-based modeling marked a significance increase of behavioral realism in travel demand forecasting. Nevertheless, with increasing behavioral realism the computational complexity and data-input requirements are also increasing and while the increase in model complexity follows an exponential rate the gain in terms of behavioral realism is at a lower rate (Cools, 2009).

Finally, there exists another approach which focuses exclusively on OD estimation and which is completely independent and outside the framework of the three travel demand forecasting paradigms. Interestingly, this approach can potentially provide a solution to the feedback problem between traffic

[7] Interestingly, the recent emergence of *dynamic activity-based* models (e.g. Habib and Miller, 2009; Arentze and Timmermans, 2011) is foreseen by some authors as a fourth generation of travel demand models.

assignment and demand modeling. OD estimation from link counts or link traffic, as it is commonly described, relies on information from link traffic data[8]. Under this approach, the traffic assignment problem, discussed in section 1.1, is actually inverted and the goal is to estimate or update a vector of OD flows from an observed vector of link flows. The main problem is that traffic count data are typically observed only in a small subset of the link flow vector and therefore the number of links which contain information is smaller than the number of OD pairs which results in an underspecified system of equations. Thus, external information in the form of a "prior" outdated OD matrix is needed in order to impose constraints. The goal then is to find the most plausible OD estimate given the observed link flows and the "prior" OD matrix.

According to the categorization of Abrahamsson (1998) there are three main groups of methods; traffic-modeling based methods, gradient-based solution techniques and statistical-inference methods. Traffic-modeling methods seek to minimize the information or maximize the entropy with respect to the "prior" OD matrix while retaining the constraints from the link flows. Contributions in this area of research can be found in Van Zuylen and Willumsen (1980), Jörnsten and Nguyen (1983), Fisk (1989) and Kawakami et al. (1992). Gradient techniques consider the "prior" OD matrix as an initial solution and the matrix is then adjusted to reproduce traffic counts by iteratively calculating directions based on the gradient of a specific objective function. Gradient-based studies include those of Spiess (1990), Drissi-Kaïtouni and Lundgren (1992) and Chen (1994). Statistical-inference methods assume that the OD and link flows are realizations of random processes and thus by adopting an appropriate distributional assumption, usually a Poisson or an approximate normal distribution, the problem reduces to an estimation problem for the corresponding parameters of interest. Statistical-inference methods can be further classified into three groups; maximum likelihood methods (Spiess, 1987; Cascetta and Nguyen, 1988, Lo et al., 1996; Vardi, 1996 and Hazelton, 2000), least squares and generalized least squares methods (Cascetta, 1984; Bell, 1991; Yang et al., 1994 and Bierlaire and Toint, 1995) and Bayesian methods (Maher, 1983; Tebaldi and West, 1996; Li, 2005, 2009; Castillo et al., 2008 and

[8] Traffic data on links are in general easier and less expensive to collect than traffic data between OD pairs, e.g, through automatic sensor detectors.

Hazelton, 2008, 2010). For a more philosophical classification and discussion of methods interested readers are referred to Timms (2001). A drawback with OD estimation from link traffic is that the majority of applications are on small theoretical networks[9], except of some gradient-based solution methods (e.g. Spiess, 1990; Chen, 1994). Therefore, despite the mathematical elegance of many of these methods, the transition from theory to practice in real-world problems remains a challenge. The inherent problem of having fewer equations than the number of unknowns poses serious limitations for applications on large-scale networks. Limitations and prospects of OD estimation from link traffic are discussed in Marzano et al. (2009).

1.3 Uncertainty in transport modeling

The topic of uncertainty analysis has, in general, received limited attention within the domain of transportation. Until recent years few studies dealt with this matter, studies which were sporadic and mainly on an ad hoc basis (e.g. Bonsall et al., 1977; Ashley, 1980). However, the number of papers inquiring uncertainty issues is gradually increasing over the last years. This might be partially explained as a side effect from an overall increase of uncertainty-related studies across diverse scientific fields, for instance ecology (Li and Wu, 2006), forensics (Brach and Dunn, 2004), engineering (Ayyub and Gupta, 1998) and so forth. Another, perhaps more influential reason, relates to the increasing and justifiable interest in the accuracy of long term travel demand forecasts which is reflected in many recent comprehensive studies, namely Parthasarathi and Levinson (2010), Flyvbjerk (2005), Flyvbjerk et al. (2005, 2006), Bain and Polakovic (2005) and Richmond (2001). From a policy-making perspective these studies provide alarming results, to say the least, concerning under or overestimation of traffic forecasts and also constitute the need of uncertainty analysis more relevant than ever.

A first detailed literature review of studies on transport uncertainty analysis is provided by de Jong et al. (2006). The authors distinguish two main sources of uncertainty, *input-uncertainty* and *model-uncertainty*, and investigate the related literature under the prism of this distinction. In this study input-

[9] A striking example can be found in Tebaldi and West (1996) and the corresponding comment by Vardi.

uncertainty is strictly defined in terms of forecasting for future years and therefore the authors consider as input-uncertainty the unknown values of future exogenous variables used as input in a given model (e.g. future incomes). Model-uncertainty on the other hand relates to two types of errors; model specification error and estimation error. Model specification error can be due to omitted variables, inappropriate functional forms and distributions for random components while estimation error is due to the fact that the true parameters are unknown and need to be estimated. De Jong et al. (2006) conclude with a small number of 21 studies which take into account one of the two forms of uncertainty and quantify the impact in terms of variance, standard deviation, 95% confidence intervals and percentile estimates. Out of these 21 studies only six quantify the impact of uncertainty on link flows, the key output concerning project evaluation as the authors comment. Furthermore, only seven studies take into account both types of uncertainty, nine solely focus on model-uncertainty and four on input uncertainty, while in one study the distinction between input and model uncertainties is not recognized.

A second, more recent and up-to-date literature review is that of Rasouli and Timmermans (2012). This review focuses exclusively on travel demand uncertainty studies within the four-step, discrete-choice and activity-based frameworks and also provides a very good overview of these frameworks. Input-uncertainty is more broadly defined by Rasouli and Timmermans, including not only the uncertainty due to future unobserved input data but also the uncertainty originating from sampling bias, survey design, reporting and/or coding mistakes, and missing/incomplete information. The definition of model uncertainty is essentially the same as in de Jong et al. (2006). The authors present a list of 14 papers which take into account at least one of the two types of uncertainty, three studies within four-step modeling, seven studies within discrete-choice modeling and four studies within activity-based modeling. In general, both literature reviews highlight and criticize the fact that despite the thousands of papers related to transport model forecasts found in journals, reports and conference proceedings every year, the literature on uncertainty analysis has been fairly limited. As commented by Rasouli and Timmermans (2012), including uncertainty analysis rather complicates things and often it seems more effective to ignore uncertainty considerations during policy development processes. The authors insightfully further comment that uncertainty analysis becomes relevant only when the following circumstances arise either individually or in conjunction; i) there are divergent political views and model results favor one particular position, ii) there are high financial, political, societal risks, iii) the policies to be implemented are controversial in terms of cost-benefit and iv) there are concerns about the limitations of a model.

Nonetheless, there seems to exist a general upward trend in the number of transport studies dealing with uncertainty analysis either because of research which points out obvious misjudgments in long-term forecasting or because of rising awareness concerning environmental (e.g. CO_2 emissions), economical (e.g. non-renewable energy sources) and societal (e.g. traffic casualties) issues. Additional evidence for this upward trend can be found in studies related to modeling of travel times (Ettema and Timmermans, 2006; Li and Rose, 2011), infrastructure management (Kuhn and Madanat, 2005; Ng et al., 2011) and specific travel-demand approaches (Matas et al., 2012). Furthermore, there exists a rapidly growing body of research related to uncertainty in traffic assignment models starting from Waller et al. (2001)[10]. Finally, Poole and Raftery (2000) and Ševčíková et al. (2007) provide the first formal probabilistic-based approaches for assessing uncertainties in sequential deterministic and agent-based simulation frameworks within the context of urban-simulation models.

1.4 Availability of OD data

Availability of OD data commonly originates from travel surveys and in some cases from census studies. Travel survey data are of course less expensive and easier to collect, manage and follow-up in relatively small time intervals by private agents or by organizations such as transportation centers and institutes, whereas census studies are expensive, large-scale projects, managed by governmental agencies and implemented over large periods of time. Nevertheless, despite the economical and practical advantages of travel survey

[10] This topic is discussed in more detail in chapter 4, section 4.1, as it is very relevant with some of the material presented in this dissertation.

data[11], OD estimates from travel surveys are rarely used for direct statistical inference. In relation to the previous discussion, sampling estimates propagate input-uncertainty, as defined by Rasouli and Timmermans (2012), significantly. According to Stopher and Greaves (2007) travel surveys are bounded by considerable error-producing characteristics, namely inadequate sample sizes, non-representative samples, large non-response rates and under-reporting of trips. Awareness regarding these problems is not new, for instance discussions concerning sample size and non-representation date back to Wermuth (1981) and Brög and Erl (1982), while the notion that sample-expansion is not an appropriate solution is first pointed out in Brög and Ampt (1982). The latter is nowadays widely accepted within the scientific community, a notable recent empirical study by Cools et al. (2010) verifies that sample estimates are considerably biased even under large sampling rates. The compounding problems of non-response and under-reporting captured attention relatively later (e.g. Richardson et al. 1995), nevertheless, are equally prevalent.

Given the unreliability of travel surveys for OD estimation, the trip-distribution phase in four-step models can be viewed essentially as a “correction” procedure. That is, trip-distribution is actually delivering updated or improved OD estimates given initial OD estimates which are either outdated or approximate. For instance, the entropy-maximization approach of Wilson is based on knowledge of a macro-state (aggregated OD flows) and a base OD matrix (initial OD estimate) and provides the most likely to be observed meso-state (improved OD estimate). Similarly, methods based on link traffic deliver the most likely OD matrix given the information from link counts and an initial OD estimate. In other words, the methodological framework of these modeling approaches associates with the type of information which is available and most commonly this information is weak, partial or outdated.

On the other hand, research with OD matrices derived from census studies has been to a large degree overlooked. The main reason of course is that travel census data are usually not available due to the increased costs of collecting such data. Another possible explanation is that traditional transportation

[11] That is not to say that travel surveys are easy per se in development. Travel surveys in general require extremely careful and detailed planning concerning questionnaire format and design, definition of sampling framework and sample size among other issues. Details on travel survey data-collection methods (not restricted to OD surveys) can be found in TRB Travel Survey Manual (2012).

planning approaches are related more to traffic engineering and less to statistical modeling, and therefore even when travel census data are available the required data-analytic tools for implementing analysis on large-scale are not available. Regardless of reasons, OD information from census studies provides opportunities for implementing direct statistical methodology as in this case the main problems of inadequate sample size and non-representation are not of concern.

1.5 Scope of research

The purpose of this dissertation is to investigate new ways of OD modeling for cases where reliable historical travel-demand information is available. To this end, an OD matrix derived from the 2001 Belgian travel census is under consideration. The main objectives are the following. First, to demonstrate that statistical OD modeling can potentially replace the first phases of traditional travel-demand modeling. Second, to investigate traffic assignment inference given the predictions of well-validated statistical models.

Given these objectives, the first problem is to find suitable statistical models which will capture the underlying mechanisms of a large and complex dataset as is the census OD for the Belgian region of Flanders. The second problem is to find a formal way of generating OD predictions which can be subsequently utilized as input into traffic assignment models. The first problem is addressed within the modeling framework of Poisson-mixture regression models. This distributional family is suitable for handling complex count datasets for which the basic Poisson assumption is not adequate mainly due to the problem of over-dispersion. The second problem is addressed by adopting a Bayesian approach. Bayesian methods, in general, provide a natural framework for fitting hierarchical models which require estimation of parameters as well as random effects and also for generating realizations or predictions of new data.

As shown in the following chapters of this dissertation, the proposed approach provides a range of advantages for transport modeling. These advantages can be summarized as follows:

- Incorporating the trip-generation and trip-distribution steps of the four-step model into statistical models that provide a wider inferential scope for parameters of scientific interest as well as for predictions of OD flows.
- Parameters can still be interpreted in terms of trip-production and trip-attraction concepts while OD predictions on various hierarchical levels may serve as predictive scenarios in transportation planning.
- The statistical models under consideration and the experimental design adopted lead to a new and alternative perspective of direct-demand gravity modeling, thus retaining a strong relationship with traditional transportation modeling.
- Delivering robust probabilistic estimates for link traffic and network congestion in the form of predictive link volume and volume-over-capacity distributions conditional on the assignment model.
- Providing an appropriate working framework for quantifying model-uncertainty during OD modeling and input/model-uncertainty during traffic assignment modeling.

1.6 Structure of dissertation

Chapter 2 is an introductory chapter dealing with some basic aspects of Bayesian theory and methodology. The chapter is by no means exhaustive in its content. Its main purpose is to provide information for readers which are unfamiliar to Bayesian approaches so that the methods used throughout this dissertation, especially in chapter 3, can be easily comprehended. As such, the chapter starts with the fundamental concepts of Bayesian theory related to the Bayes theorem and the prior and posterior distributions. Following that, inferential tools for the main aspects of Bayesian inference are presented, particularly concerning posterior inference, predictive inference and model comparison issues. The next section deals in brief with Markov Chain Monte Carlo simulation with emphasis on the Metropolis-Hastings algorithm and related convergence diagnostics which are utilized for the purposes of this dissertation. Finally, guidelines and examples are provided concerning the practical issue of calculating posterior quantities of interest given the output from Markov Chain Monte Carlo.

The third chapter is dealing with the statistical modeling part of this dissertation. Initially, the 2001 Belgian census study, the administrative structure of the Flemish region, the OD data and the explanatory data are described. In the following section the first methodological steps and results are presented; namely the assumptions for the Poisson and negative binomial models, the correspondence between negative binomial and Poisson-gamma modeling, the details of Metropolis-Hastings simulation and the results of predictive inference and model comparison. The section concludes with predictive goodness-of-fit of the Poisson-gamma model and with a note explaining why prediction is based on the Poisson-gamma and not on the negative binomial representation. Next, the modeling framework is extended and the Poisson-mixture perspective is explained in more detail. In this section the Poisson-gamma, Poisson-lognormal and Poisson-inverse Gaussian models are described, details of Metropolis-Hastings simulation are provided and results related to posterior inference and model comparison are discussed. The section ends with predictive inference, this time under both Poisson-gamma and Poisson-inverse Gaussian predictions. The chapter concludes with a note on an alternative direct-demand, gravity modeling representation of the models and with a summary of results.

Chapter 4 investigates the benefits and the questions which arise when Bayesian OD predictions are used as input into traffic assignment. The first two sections are short introductions providing basic literature reviews for two important contemporary research directions highly related to the material of the chapter, namely traffic assignment under demand-uncertainty and congested/critical link identification. The following section is a brief description of the road network of Flanders. Next, the methodology of the chapter is discussed where two inputting methods are proposed; one suitable for approximate network inference and one leading to full network inference given demand uncertainty. Results from the two methods based on Poisson-gamma predictions and deterministic user equilibrium assignment are compared in the following section, covering issues such as the average state of the network, total system travel time behavior, performance of point and interval estimates, and congested link identification under full network inference. The following section deals more or less with similar issues but under the perspective of comparing results based on Poisson-gamma predictions to results based on Poisson-inverse Gaussian predictions and with more weight given to full network inference under demand-uncertainty. Results are summarized in the final section of the chapter.

Chapter 5 contains relevant research material from recent experiments and modeling approaches. The chapter is divided in two main sections which deal with different topics. The first section is essentially a continuation of experiments related to those of chapter four. In this part, results from deterministic user equilibrium are compared to results from stochastic user equilibrium which is the commonly used alternative traffic assignment model. Stochastic user equilibrium experiments are implemented for different choices of route-choice models, namely probit and logit, and for 3 different levels of perception-error variance. The comparisons are conditional on Poisson-gamma predictions and are along the lines of chapter four focusing on potential differences in total system travel time, point estimates, interval estimates and congestion analysis under full network inference. The second section is related to a recent and novel methodological advance on OD modeling which has been developed independently of the research presented here. The approach is simply known as the radiation model. The potential usefulness of the radiation model within our modeling framework is investigated. A first attempt to incorporate an important explanatory variable from the radiation model in the negative binomial model is presented. The section and chapter conclude with a conceptual proposition for a Bayesian extension of the radiation model.

The dissertation ends with a general discussion in chapter 6. Initially, a summary of the main results and contributions of this research are presented. The results are then also interpreted from the perspective of input and model uncertainties. In the following two sections, considerations regarding the overall applicability of the proposed research and certain econometrical issues are discussed. Finally, possible future research directions are summarized in the last section.

2 The Bayesian approach

The foundation of Bayesian theory was set in a study published in 1763 by reverend Thomas Bayes and was formalized quickly later in 1774 by Laplace who presented the general form of Bayes theorem. In the early 20th century Bayesian theory was evoked by physicist Harold Jeffreys and econometrician Arthur Bowley who argued on behalf of Bayesian methods. The term “Bayesian” in its contemporary sense was established during a period after the 1950’s as by that time many statisticians advocated this practice of statistics as a remedy for certain deficiencies of the *classical* or *frequentist* approach, mainly the interpretation of confidence intervals and the violation of the likelihood principle under specific conditions. In conjunction with the advancements in computing technology and the advents of simulation methods, Bayesian inference is becoming more popular ever since.

The difference between frequentist and Bayesian statistics is philosophical in nature. Frequentist theory assumes a probability distribution (likelihood) for the data with unknown and *fixed* parameters. Frequentist uncertainty originates from repetition of samples and therefore estimation procedures are based on the assumption of an infinite replication of the same inferential problem for fixed values of unknown parameters. Bayesian theory assumes a probability distribution (likelihood) for the data with unknown and *random* parameters which are assigned a *prior* distribution. Bayesian uncertainty originates from the parameters, so the evaluation procedure is based on infinite sampling of parameters drawn from a distribution which is conditional on the data, i.e. the *posterior* distribution (Carlin and Louis, 1996).

Bayesian inference is described in brief in this chapter. Coverage of the topic is limited to some main concepts and to certain modeling and simulation aspects which relate to the methods used throughout this dissertation. Nevertheless, references are provided for further issues which are not discussed in detail. In addition, interested but unfamiliar readers are further referred to the books by Gelman et al. (2003) for a good introductory reading which combines theory with practice, Ntzoufras (2009) for a comprehensive guide on modeling approaches and practical implementation in WinBUGS software and

Carlin and Louis (1996) for a slightly deeper perspective on mathematical aspects.

2.1 The Bayes theorem

For observed data $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$, the Bayesian approach specifies a sampling distribution $p(\mathbf{x} \mid \boldsymbol{\theta})$ which is the likelihood of the data given a parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$ [12]. The parameter vector is considered random and is assigned with a prior distribution $p(\boldsymbol{\theta})$. The latter distribution expresses prior beliefs, i.e. it is the distribution of $\boldsymbol{\theta}$ *before* any data are observed.

From the basic properties of conditional probabilities the posterior distribution of $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{x})},$$

where $p(\mathbf{x})$ is the *marginal likelihood*, also referred to as the *prior predictive likelihood* of the data given by $p(\mathbf{x}) = \sum_{\boldsymbol{\theta}} p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$ or by $p(\mathbf{x}) = \int_{\boldsymbol{\theta}} p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$ depending whether $\boldsymbol{\theta}$ is discrete or continuous, respectively. The formula for obtaining the posterior is known as *Bayes theorem*. In general, the posterior distribution encapsulates all necessary information about the parameter vector given prior beliefs and observed data as it is the distribution of $\boldsymbol{\theta}$ *after* having observed the data. The marginal likelihood $p(\mathbf{x})$ is of major importance for model selection procedures, but this issue will not be discussed in detail here. More details can be found in Kass and Raftery (1995) who interpret the marginal likelihood as the probability of witnessing the data that were actually manifested, calculated before any data became available. Calculating the marginal likelihood is cumbersome, but – conveniently – this is not essential for evaluating the posterior. Since $p(\mathbf{x})$ does not depend on $\boldsymbol{\theta}$, the un-normalized posterior is

$$p(\boldsymbol{\theta} \mid \mathbf{x}) \propto p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}).$$

[12] For simple problems $\boldsymbol{\theta}$ might contain just one parameter, i.e. $\boldsymbol{\theta} = \theta$.

Knowledge of the un-normalized posterior is sufficient for calculating the normalized density directly (see section 2.1 for conjugate priors), approximately (e.g. see chapter 4 in Gelman et al., 2003) or through simulation (see section 2.6), therefore the above expression is particularly useful as it significantly simplifies calculations.

In general, the Bayesian framework provides a natural updating structure when new data $\mathbf{x}^{new}$ become available. In this case, the posterior distribution given $\mathbf{x}$ can be regarded as a prior for $\mathbf{x}^{new}$ and by a new application of Bayes theorem obtain an updated posterior, i.e. $p(\boldsymbol{\theta} \mid \mathbf{x}^{new}) \propto p(\mathbf{x}^{new} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{x})$, without necessarily being restricted to the same likelihood and prior distributional assumptions.

2.2 Prior distribution

Prior distributions can be classified to *informative* or *non-informative*, *parametric* or *non-parametric* and *conjugate* or *non-conjugate*, Cross-classification according to these attributes is also possible.

Informative priors are used when accumulated information about the parameter vector exists either from past studies or from knowledge of subject-area experts. When that is the case, this information can be incorporated in the prior distribution, e.g. by specifying an appropriate value for the mean and a small variance. On the other hand, when information about the parameter vector is not available or when the objective is to adopt a prior with a minimal impact so that inference will remain unaffected by external information, non-informative priors are adopted. Non-informative priors are usually constructed by specifying a large value for the variance which makes the prior almost uniform over the parameter space and therefore non-informative prior densities are also referred to as *vague, flat* or *diffuse*. The use of non-informative priors is an important topic in Bayesian inference, additional information can be found in Kass and Wasserman (1996) and the references therein.

Non-parametric priors make no assumptions concerning the probability distribution function. Such priors are usually also called *elicited* as they are customarily employed under informative frameworks. For instance, when $\boldsymbol{\theta}$ is discrete and prior information is available, then one can initially distinguish the values which are most likely to occur and assign point masses which will sum to one in a way that reflects prior beliefs. By analogy, if $\boldsymbol{\theta}$ is continuous then one can construct a prior histogram of $\boldsymbol{\theta}$. Non-parametric priors are used in specific circumstances (see e.g. Oakley and O'Hagan, 2007) and are not so common. In most cases the common assumption is that the prior belongs to a known parametric distributional family $p(\boldsymbol{\theta} \mid \boldsymbol{\eta})$ where the hyper-parameter vector $\boldsymbol{\eta}$ can be easily specified in order to construct an informative or a non-informative prior[13].

The distinction between conjugate and non-conjugate priors arises within the context of parametric prior designs. A prior distribution is called conjugate when the combination with the likelihood results to a posterior distribution which belongs to the same distributional family as the prior. It is worth noting that that all of the distributions belonging to the exponential family have conjugate priors (Morris, 1983) and that most of the likelihoods used in common statistical inference belong to the exponential family of distributions. Extensive information on conjugate priors can be found in Bernardo and Smith (1993). Non-conjugate priors usually result in intractable posteriors which do not correspond to known distributional forms. In the past Bayesian inference was mainly limited on conjugate prior/likelihood models. Nevertheless, this is not the case anymore as the wide range of available simulation methods in conjunction with computing power facilitates model-fitting under non-conjugate prior assumptions.

2.3 Posterior inference

Posterior inference involves summing-up the information contained in the posterior distribution. Commonly, this is done by calculating posterior point estimates of location and dispersion, and posterior interval estimates which are known as *credible sets*, *credible* or *credibility intervals*.

[13] It should be mentioned that when $\boldsymbol{\eta}$ is fixed then the conditional dependence notation is usually dropped, i.e. $p(\boldsymbol{\theta} \mid \boldsymbol{\eta}) \equiv p(\boldsymbol{\theta})$, which refers to the framework described in section 2.1. Nevertheless, $\boldsymbol{\eta}$ can be further assumed to be random. In that case $\boldsymbol{\eta}$ is assigned with a *hyper-prior* distribution $p(\boldsymbol{\eta})$ and then the posterior distribution under consideration is $p(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \mathbf{x}) \propto p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{\eta}) p(\boldsymbol{\eta})$. This is the framework of *hierarchical* or *multilevel* Bayesian inference.

2.3.1 Point estimates of location and dispersion

Location measures give the most likely posterior central states of the parameter vector. The most common options are the posterior mean, median and mode. The mean corresponds to the expected value of $\boldsymbol{\theta}$, the median to the value of $\boldsymbol{\theta}$ which divides the parametric space into two equal posterior probabilities and the mode to the most likely value of $\boldsymbol{\theta}$. When the posterior is symmetrical then the mean and the median will coincide and as long as the posterior is unimodal the mode will also coincide. For cases where the posterior distribution exhibits either positive or negative skewness, the median is often preferred as a location measure as it is in between the mean and the mode (Carlin and Louis, 1996). The mode in general is the most difficult measure to calculate especially for parameter vectors of large cardinality and usually mode estimates are approximated either by maximization algorithms (e.g. Bickel and Früwirth, 2006) or by adopting appropriate priors which facilitate mode finding (e.g. Eaves and Chang, 1992).

Dispersion measures include the posterior variance, standard deviation, precision (the inverse of variance), inter-quartile range and the curvature at the mode. Additional options for problems with more than one parameter include the posterior covariance matrix, the precision matrix which is the inverse of the covariance matrix and also the curvature-at-the-mode matrix which can be estimated by the matrix of the second derivates of $-\log p(\boldsymbol{\theta} \mid \mathbf{x})$ evaluated at the posterior mode (Gamerman and Lopes, 2006).

2.3.2 Credible intervals

The Bayesian analogues of confidence intervals are known as credible sets or credible intervals. As defined by Carlin and Louis (1996), a $100 \times (1-\alpha)\%$ credible set for $\alpha \in (0,1)$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is the posterior parameter space, is a subset $C \subseteq \boldsymbol{\Theta}$ such that $(1-\alpha) \leq \Pr(C \mid \mathbf{x}) = \int_C p(\boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta}$. The inequality (instead of equality) is used in order to include discrete cases for which in addition the integration is replaced by summation. In contrast to the interpretation of confidence intervals, credible intervals enable direct probability statements for the likelihood of $\boldsymbol{\theta}$ falling into C. That is, the interpretation of the above interval is

"The probability that $\boldsymbol{\theta}$ *lies in C given the observed data is at least* $1-\alpha$*".*

For cases of asymmetric or multimodal posterior distributions it is preferable to calculate the *highest posterior density* (HPD) credible set, which groups together the most likely values of $\boldsymbol{\theta}$ and hence is narrower than the equal tail credible set. Nevertheless, calculating HPD credible sets is as not straightforward. Wright (1986) presented an iterative method for univariate cases; Ghosh and Mukerjee (1995) and Hyndman (1996) introduced iterative solutions for multivariate cases.

2.4 Predictive inference

Inference about short-term *predictions* or *replications* of data is an important aspect of statistical analysis. The Bayesian approach provides a natural and convenient framework for predictive inference since a *predictive distribution* can be formally defined based on simple rules of conditional probabilities. Predictions from an assumed model can be subsequently compared to the observed data leading to formal probabilistic inference based on *Bayesian p-values*.

2.4.1 Predictive distribution

Consider predicting an *unobserved* vector $\mathbf{x}^{pred}$ which is independent of $\mathbf{x}$ conditionally on $\boldsymbol{\theta}$[14]. The distribution of $\mathbf{x}^{pred}$ given $\mathbf{x}$ is

$$\begin{aligned} p(\mathbf{x}^{pred} \mid \mathbf{x}) &= \int p(\mathbf{x}^{pred}, \boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta} \\ &= \int p(\mathbf{x}^{pred} \mid \boldsymbol{\theta}, \mathbf{x}) p(\boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta} \\ &= \int p(\mathbf{x}^{pred} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta}. \end{aligned}$$

The transition from the second to the third line is due to conditional independence. The predictive distribution provides information given the likelihood, the prior and the posterior. It is also known as the posterior predictive distribution, since it is essentially the expected distribution of future observations with respect to the posterior.

---

[14] The predicted vector $\mathbf{x}^{pred}$ should not be confused with the observed vector of new data $\mathbf{x}^{new}$ discussed in section 2.1.

The predictive distribution is the basis for predictive inference within the Bayesian paradigm. The rationale is that parameter-based inference for issues such as model adequacy is not well defined since parameters are never observed. Contrary, the predictive distribution is defined in terms of observable values and therefore it is the natural instrument for decisions concerning model adequacy. Some authors go even further and argue that the predictive distribution is also the most appropriate tool for model selection. Interested readers are referred to Geisser and Eddy (1979) and Laud and Irbahim (1995), for instance.

2.4.2 Bayesian p-values

Bayesian p-values or posterior predictive p-values, as they are also known, are derived by posterior predictive checks. Predictive checks can be used in order to assess the adequacy or the *fit* of a given model. According to Gelman et al. (1993) Bayesian p-values are generalization of frequentist p-values from the perspective that they average over the posterior distribution rather than fixing the unknown parameter at a certain point estimate $\hat{\boldsymbol{\theta}}$.

Within the Bayesian framework goodness-of-fit is evaluated by comparing observed data $\mathbf{x}$ with predicted data $\mathbf{x}^{pred}$ drawn from the predictive distribution. The discrepancy between observed and predicted data is measured through *test quantities* $T(\mathbf{x},\boldsymbol{\theta})$ which can be functions of the parameters and the data or functions of the data only. In the latter case test quantities are free of parameters, i.e. $T(\mathbf{x},\boldsymbol{\theta}) \equiv T(\mathbf{x})$, and the Bayesian p-values concur with asymptotic frequentist p-values (Gelman et al., 1993).

The Bayesian p-value is first defined in Rubin (1984) as the probability that the predicted data are more extreme than the observed data for a certain test quantity. That is,

$$\text{Bayesian p-value} = \Pr(T(\mathbf{x}^{pred},\boldsymbol{\theta}) \geq T(\mathbf{x},\boldsymbol{\theta}) \mid \mathbf{x}),$$

where the probability is defined over the posterior distribution of $\boldsymbol{\theta}$ and the predictive distribution of $\mathbf{x}^{pred}$, therefore

$$\text{Bayesian p-value} = \int\int I\left(T(\mathbf{x}^{pred},\boldsymbol{\theta}) \geq T(\mathbf{x},\boldsymbol{\theta})\right) p(\mathbf{x}^{pred} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta} d\mathbf{x}^{pred},$$

where $I\left(T(\mathbf{x}^{pred},\boldsymbol{\theta}) \geq T(\mathbf{x},\boldsymbol{\theta})\right)$ is the indicator function given by

$$I(A) = \begin{cases} 1, & \text{if } A \text{ is true,} \\ 0, & \text{if } A \text{ is false.} \end{cases}$$

Bayesian p-values close to 0.5 are suggestive of an almost exact fit of predictive data to observed data with respect to the test quantity. On the other hand values close to 0 or 1 imply a failure to capture the observed pattern.

Three clarifications are of importance concerning interpretation of p-values. First, it must be noted that Bayesian p-values can serve *only* as measures of discrepancy between an assumed model and the observed data and are not formally comparable *across* different models (Carlin and Louis, 1996). Second, Bayesian p-values should not be interpreted as strict “numerical evidence”. For instance, a p-value of 0.35 is not strictly “worse” than a p-value of 0.45; both are good, indicating that the predictions of a model are supported by the observed data. Similarly, a p-value of 0.001 is not “better” than a p-value of 0.000001 since both p-values reveal a significant discrepancy between predicted and observed data[15]. The only evidence provided by Bayesian p-values concerns the “extremeness” hypothesis. Not rejecting this hypothesis for a meaningful test quantity which describes an important aspect of the analysis, implies that a model should be altered or appropriately expanded in order to fit the observed data more consistently. Third, p-values measure “statistical significance” and not “practical significance”. The latter relates to the main objectives of an analysis. The Bayesian paradigm provides flexibility in terms of defining $T(\mathbf{x},\boldsymbol{\theta})$ in any desirable way. Therefore, model-fit can be investigated from many aspects and usually one evaluates the fit with respect to several test quantities. From this point of view extreme p-values may be ignored if they do not affect main inferences. Additionally, a model may still be deemed as “good enough” in the sense that it will provide satisfactory predictions for certain aspects of interest but not for some others. An interesting discussion concerning the last two remarks can be found in Gelman et al. (2003).

[15] By analogy, frequentists define a level of statistical significance α and investigate only whether a test statistic falls within the interval $(\alpha/2, 1-\alpha/2)$ or not.

2.5 Model selection information criteria

Model selection and model comparison relates to distinguishing the best model among a class of models, either due to uncertainty concerning the most parsimonious set of explanatory variables or due to uncertainty regarding distributional or other functional assumptions. The simplest and also most commonly-used approach is through information criteria. Within the Bayesian framework, the most popular criteria are the Bayesian version of Akaike's information criterion (AIC) (Akaike, 1974), the Bayes information criterion (BIC) also known as Schwartz criterion (Schwarz, 1978) and the Deviance information criterion (DIC) (Spiegelhalter et al., 2002).

All of the aforementioned criteria are based on the evaluation of the *deviance* which is defined as minus twice the log-likelihood, i.e. $D(\boldsymbol{\theta}) = -2\log p(\mathbf{x} \mid \boldsymbol{\theta})$. In general, the best fitting model within a group of models is the one that has the lowest value for a specific criterion. AIC is defined as

$$AIC = D(\hat{\boldsymbol{\theta}}) + 2p ,$$

where $\hat{\boldsymbol{\theta}}$ is the vector which minimizes the deviance and p is the number of parameters. BIC on the other hand is defined as

$$BIC = D(\hat{\boldsymbol{\theta}}) + p\log(n) ,$$

where n denotes sample size. Bayesian variations of AIC and BIC based on posterior summaries (Brooks, 2002) are as follows

$$AIC_{\overline{D(\boldsymbol{\theta})}} = \overline{D(\boldsymbol{\theta})} + 2p \text{ and } \mathrm{B}IC_{\overline{D(\boldsymbol{\theta})}} = \overline{D(\boldsymbol{\theta})} + p\log(n),$$
$$AIC_{D(\bar{\boldsymbol{\theta}})} = D(\bar{\boldsymbol{\theta}}) + 2p \text{ and } \mathrm{B}IC_{D(\bar{\boldsymbol{\theta}})} = D(\bar{\boldsymbol{\theta}}) + p\log(n),$$

where $\overline{D(\boldsymbol{\theta})}$ is the posterior mean of deviance and $D(\bar{\boldsymbol{\theta}})$ is the deviance based on the posterior mean of the parameter vector. DIC is a purely Bayesian criterion which according to theory is given by $DIC = \overline{D(\boldsymbol{\theta})} - p_e$ with p_e being the number of "effective" parameters. In practice this criterion is calculated as

$$DIC = 2\overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}) .$$

DIC is useful for determining the best model within a group of models, it does not indicate whether a model is "true" or not. Given that DIC is in fact a

Bayesian generalization of AIC, the same also applies for the latter. BIC on the other hand does provide an indication for the “true” underlying model and therefore this criterion is more suitable when interest lies in prediction outside the experimental sampling range, i.e. for transferable inference where the “true” underlying model is of importance.

In general, for non-hierarchical models and large sample sizes AIC and DIC will coincide, more or less. For hierarchical models matters become more complicated and the related research is ongoing as there are various likelihood levels for possible inference. AIC and BIC require number of counting parameters so these measures are limited to marginal inference, although there exist some recent alternative approaches such as the conditional AIC for hierarchical inference (Vaida and Blanchard, 2005). Concerning DIC, Celeux et al. (2006) present a study – which is followed by an extended discussion – where 8 variations of this criterion are proposed depending on whether the unobserved components in missing data/random effects/mixture models are treated as variables or parameters.

Information criteria provide a relatively simple way, in terms of computing requirements, for Bayesian model comparison. Another more formal approach is through *Bayes factors*, but this approach requires estimation of the marginal likelihood $p(\mathbf{x})$. Bayes factors will not be discussed here as they are not included in the methodological tools utilized in this dissertation. Interested readers are referred to Kass and Raftery (1995). Finally, it is worth noting that BIC provides a rough approximation to the logarithm of a Bayes factor.

2.6 Markov Chain Monte Carlo

The use of Markov Chain Monte Carlo (MCMC) methods has become extremely popular in Bayesian statistics. The methods owe their popularity due to the ability of approximating accurately high dimensional integrals through simulation. Thus, in Bayesian statistics MCMC is used extensively for non-conjugate analyses where the posterior $p(\boldsymbol{\theta} \mid \mathbf{x})$ is not of known distributional form. In what follows we merely illustrate the basic idea of MCMC and provide a practical description of the particular algorithm which is utilized for the purposes of this dissertation. References for a more in-depth reading are provided.

2.6.1 Markov chains

The basic idea of MCMC is to formulate a Markov chain from a certain starting point and iterate the chain until it converges to a stationary distribution. A Markov chain is a stochastic process $\left\{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^t\right\}$[16] characterized by two properties:

1. The distribution of $\boldsymbol{\theta}$ in period $t+1$ given the $\boldsymbol{\theta}$'s in all preceding periods depends only on the $\boldsymbol{\theta}$ in the latest time period t, i.e.

$$f(\boldsymbol{\theta}^{t+1} \mid \boldsymbol{\theta}^t, \boldsymbol{\theta}^{t-1}, \ldots, \boldsymbol{\theta}^1) = f(\boldsymbol{\theta}^{t+1} \mid \boldsymbol{\theta}^t).$$

   Alternatively, one can say that given a present state of $\boldsymbol{\theta}$, past and future states are independent.
2. If a Markov chain is irreducible, aperiodic and positive recurrent then for $t \to \infty$ the distribution of $\boldsymbol{\theta}^t$ converges to a stationary distribution.

Under this temporal context a stationary distribution does not change when shifted in time. The stationary distribution is often called the *equilibrium* distribution. Irreducible means that there is a positive probability of moving from any given state to any other state, aperiodic means that there are no absorbing states from which the chain cannot escape and positive recurrent means that the probability of returning to the initial state equals one with the expected time of return being finite.

Within the Bayesian framework the goal is to generate samples $\left\{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^t\right\}$ from a Markov chain which will be dependent samples from the posterior $p(\boldsymbol{\theta} \mid \mathbf{x})$, the *target* distribution as it is often called within this context. In other words, the equilibrium distribution should concur to the posterior distribution. When this is accomplished all that is need is to iterate long enough and then to discard an initial part of the samples, e.g. from $\boldsymbol{\theta}^1$ to $\boldsymbol{\theta}^{t_0}$ and keep the samples $\left\{\boldsymbol{\theta}^{t_0+1}, \boldsymbol{\theta}^{t_0+2}, \ldots, \boldsymbol{\theta}^t\right\}$. The discarded part is referred to as the *burn-in* period of the chain which is the period before reaching the stationary or

[16] The process is described in terms of $\boldsymbol{\theta}$ in order to keep consistency with the problem at hand of section 2.1.

equilibrium state. Monitoring the convergence to the equilibrium state is in practice implemented through the use of *convergence diagnostics* which are discussed in brief in section 2.6.3.

The two basic MCMC approaches are the *Metropolis-Hastings* algorithm and the *Gibbs sampler*. The latter first introduced by Geman and Geman (1984) is actually a special case of Metropolis-Hastings. Gibbs sampling has become quite popular due to its ease of use[17], but on the other hand it is only applicable under certain conditions. We proceed with a brief description of the Metropolis-Hastings algorithm which is the sampling method that is utilized for the proposed models of this dissertation. Information for the Gibbs sampler can be found in Casella and George (1992) and Gelfand (2000).

2.6.2 The Metropolis-Hastings algorithm

The initial form of the Metropolis-Hastings (MH) algorithm was introduced by Metropolis et al. (1953) and was later generalized by Hastings (1970). The algorithm is based initially on a *proposal distribution* $q_t(\boldsymbol{\theta}^{t-1}, \boldsymbol{\theta}^t)$ (also known as *jumping or candidate* distribution). This distribution gives a proposed probability of transition from state $\boldsymbol{\theta}^{t-1}$ to state $\boldsymbol{\theta}^t$. When the proposal distribution is such that the relation $p(\boldsymbol{\theta}^\beta \mid \mathbf{x}) q_t(\boldsymbol{\theta}^\beta, \boldsymbol{\theta}^\alpha) = p(\boldsymbol{\theta}^\alpha \mid \mathbf{x}) q_t(\boldsymbol{\theta}^\alpha, \boldsymbol{\theta}^\beta)$ is satisfied for all $\boldsymbol{\theta}^\alpha$ and $\boldsymbol{\theta}^\beta$, then iterations from q_t for $t \to \infty$ will result in a dependent sample from the target posterior $p(\boldsymbol{\theta} \mid \mathbf{x})$. This sufficient condition is called the *reversibility condition.*

Finding a proposal to satisfy this condition is difficult in general and therefore the MH algorithm introduces a transition probability α_{MH} which will ensure equality for the reversibility condition[18]. For instance, if $p(\boldsymbol{\theta}^\beta \mid \mathbf{x}) q_t(\boldsymbol{\theta}^\beta, \boldsymbol{\theta}^\alpha) > p(\boldsymbol{\theta}^\alpha \mid \mathbf{x}) q_t(\boldsymbol{\theta}^\alpha, \boldsymbol{\theta}^\beta)$ this means that transitions from state $\boldsymbol{\theta}^\beta$ to $\boldsymbol{\theta}^\alpha$ are made often while transitions from $\boldsymbol{\theta}^\alpha$ and $\boldsymbol{\theta}^\beta$ are made seldom, therefore $\alpha_{MH}(\boldsymbol{\theta}^\alpha, \boldsymbol{\theta}^\beta)$ is set equal to 1 to ensure more transitions from $\boldsymbol{\theta}^\alpha$ to $\boldsymbol{\theta}^\beta$

[17] For instance, software WinBUGS is based on Gibbs sampling.
[18] It is in this point that the Gibbs sampler becomes a special case of MH simulation. For details see Brooks (1998).

and $\alpha_{MH}(\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\alpha})$ is simply set so that the reversibility condition is satisfied, that is

$$\begin{aligned}
&p(\boldsymbol{\theta}^{\beta} \mid \mathbf{x}) q_t(\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\alpha}) \alpha_{MH}(\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\alpha}) = p(\boldsymbol{\theta}^{\alpha} \mid \mathbf{x}) q_t(\boldsymbol{\theta}^{\alpha}, \boldsymbol{\theta}^{\beta}) \alpha_{MH}(\boldsymbol{\theta}^{\alpha}, \boldsymbol{\theta}^{\beta}) \overset{\alpha_{MH}(\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\alpha})=1}{\Rightarrow} \\
&p(\boldsymbol{\theta}^{\beta} \mid \mathbf{x}) q_t(\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\alpha}) \alpha_{MH}(\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\alpha}) = p(\boldsymbol{\theta}^{\alpha} \mid \mathbf{x}) q_t(\boldsymbol{\theta}^{\alpha}, \boldsymbol{\theta}^{\beta}) \Leftrightarrow \\
&\alpha_{MH}(\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\alpha}) = \frac{p(\boldsymbol{\theta}^{\alpha} \mid \mathbf{x}) q_t(\boldsymbol{\theta}^{\alpha}, \boldsymbol{\theta}^{\beta})}{p(\boldsymbol{\theta}^{\beta} \mid \mathbf{x}) q_t(\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\alpha})}.
\end{aligned}$$

Note that $p(\boldsymbol{\theta} \mid \mathbf{x}) = p(\mathbf{x} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) / p(\mathbf{x})$ so knowledge of $p(\mathbf{x})$ is not needed as it cancels out in the ratio, therefore

$$\alpha_{MH}(\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\alpha}) = \frac{p(\mathbf{x} \mid \boldsymbol{\theta}^{\alpha}) p(\boldsymbol{\theta}^{\alpha}) q_t(\boldsymbol{\theta}^{\alpha}, \boldsymbol{\theta}^{\beta})}{p(\mathbf{x} \mid \boldsymbol{\theta}^{\beta}) p(\boldsymbol{\theta}^{\beta}) q_t(\boldsymbol{\theta}^{\beta}, \boldsymbol{\theta}^{\alpha})}.$$

That is the basic reasoning of MH simulation, the algorithmic form in order to simulate a MH sample of size M is:

1. Set initial value $\boldsymbol{\theta}^0$.
2. For iteration $t = 1, 2, \ldots, M$:
   a) Generate $\boldsymbol{\theta}^*$ from the proposal distribution $q_t(\boldsymbol{\theta}^{t-1}, \boldsymbol{\theta}^*)$.
   b) Calculate $\alpha_{MH}(\boldsymbol{\theta}^{t-1}, \boldsymbol{\theta}^*) = \min\left[\dfrac{p(\mathbf{x} \mid \boldsymbol{\theta}^*) p(\boldsymbol{\theta}^*) q_t(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{t-1})}{p(\mathbf{x} \mid \boldsymbol{\theta}^{t-1}) p(\boldsymbol{\theta}^{t-1}) q_t(\boldsymbol{\theta}^{t-1}, \boldsymbol{\theta}^*)}, 1\right]$.
   c) Generate a uniform random number u from $U(0,1)$.
   d) Set $\boldsymbol{\theta}^t = \begin{cases} \boldsymbol{\theta}^*, & \text{if } \alpha_{MH}(\boldsymbol{\theta}^{t-1}, \boldsymbol{\theta}^*) \le u, \\ \boldsymbol{\theta}^{t-1}, & \text{if } \alpha_{MH}(\boldsymbol{\theta}^{t-1}, \boldsymbol{\theta}^*) > u. \end{cases}$

It should be noted that when the proposal distribution is symmetric then $q_t(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{t-1}) = q_t(\boldsymbol{\theta}^{t-1}, \boldsymbol{\theta}^*)$ and therefore the proposal probabilities do not need to be calculated as they cancel out. That is actually the initial Metropolis algorithm of Metropolis et al. (1953) which is a special case of the MH algorithm.

Concerning the selection of the proposal distribution, the most common option is a *random-walk chain* where the candidate $\boldsymbol{\theta}^*$ is drawn from a process $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t + \mathbf{z}^{t+1}$. Thus, the candidate equals the present value plus some noise. Common options (e.g. for regression problems) are the multivariate normal distribution and the multivariate t distribution, i.e. $\boldsymbol{\theta}^* \sim \mathbf{N}(\boldsymbol{\theta}^t, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ and $\boldsymbol{\theta}^* \sim \mathbf{t}_{\mathbf{v}}(\boldsymbol{\theta}^t, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$, respectively. These densities are symmetric and simplify calculations. Random-walk chain algorithms require only specification of the covariance matrix $\mathbf{\Sigma_\theta}$. Another frequently used option is an *independence chain* where candidate values are generated based on a process $\boldsymbol{\theta}^{t+1} = \mathbf{z}^{t+1}$ which is independent of the present value and $\boldsymbol{\theta}^* \equiv \mathbf{z}^*$ can be drawn again from multivariate normal $\boldsymbol{\theta}^* \sim \mathbf{N}(\boldsymbol{\mu}_\theta, \mathbf{\Sigma}_\theta)$ or from multivariate *t* $\boldsymbol{\theta}^* \sim \mathbf{t}_\mathbf{v}(\boldsymbol{\mu}_\theta, \mathbf{\Sigma}_\theta)$ distributions. In this case both location $\boldsymbol{\mu}_\theta$ and scale $\mathbf{\Sigma}_\theta$ must be specified. In addition calculation of the proposal probabilities is required even if the proposals are symmetric.

An analytic description of the MH algorithm and further issues related to proposal distributions can be found in Chib and Greenberg (1995). Further information for MCMC in general can be found in the interesting paper of Brooks (1998) and also in the books by Gilks et al. (1996) and Gamerman and Lopes (2006).

2.6.3 Convergence diagnostics

There exists a variety of MCMC convergence diagnostics which determine whether the convergence of the Markov chain has been reached or not. Diagnostic tools come in many different forms; there are quantitative tools which result in numerical summaries and graphical tools such as time series autocorrelation plots. Most of the quantitative diagnostics are based on bias considerations, but there are also some diagnostics which address precision considerations. Finally, some approaches are suitable for occasions when one single MCMC chain is being produced while others are suitable for assessing the convergence of multiple MCMC chains starting from different initial states. Reviews for convergence diagnostics are provided by Cowles and Carlin (199) and Brooks and Roberts (1998).

In this dissertation convergence is diagnosed through the methods of Geweke (1992), Raftery and Lewis (1992) and Heidelberger and Welch (1983), which are suitable for single-chain MCMC simulation. The underlying theoretical content of some of these methods is quite complex and therefore the following brief descriptions are non-mathematical and focus on the essence of each method.

The diagnostic of Geweke (1992) is based on constructing a z-test for assessing the convergence of the mean of each element of vector $\boldsymbol{\theta}$ individually. This is achieved by splitting the MCMC sample into two parts; the first referring to some initial part of the chain and the second referring to some last part (usually the first 10% and last 50% of the observations). The estimate of the asymptotic variance is based on spectral density theory. The test is basically one of equality-of-means and the asymptotic distribution of z is standardized normal. Thus values $|z|>1.96$ lead to rejection of the null hypothesis implying that convergence has not been reached. Nevertheless, in multiparameter problems 5% of the z's are allowed to fall outside the interval $(-1.96, 1.96)$ due to type I error.

The diagnostic of Raftery and Lewis (1992) provides the information that is needed in order to estimate specific percentiles with a prespecified degree of accuracy and a certain probability, e.g. the 2.5% percentile to be estimated with an accuracy of 0.0005 and with a probability of 0.95. The diagnostic is rather intuitive based on concepts of ergodic theory and mixing properties of stationary distributions, but has proven to be a strong and reliable tool in practice. In addition, this diagnostic delivers valuable information; namely, the minimum number of iterations $N_{\min}$ which are required in order to achieve the aforementioned aim under the assumption of independence (i.e. zero autocorrelations), the total number of iterations N that the chain must run, the number of burn-in iterations N_{burnin} and finally a *dependence factor* $I = N_{\min} / N$ indicating the required increase of the total sample due to autocorrelations. Values of I close to 1 indicate that the posterior samples are almost independent while values greater than 5 indicate serious autocorrelation problems. The dependence factor is very useful for transforming a MCMC sample with high autocorrelations to an almost independent sample as it can be used to define the proper *thinning interval*, i.e. keep every I-th iteration of the sample.

The diagnostic of Heidelberger and Welch (1983) is for univariate tests and consists of two parts. The first part, based on Brownian bridge theory, initially checks stationarity for the first 10% of the observations. If stationarity is rejected, then the first 10% is discarded and the test is repeated for the following 10%. This procedure is repeated until the test is passed or until more than 50% of sample is discarded. For the latter case convergence is rejected. For the former case the diagnostic proceeds to the second test in which the precision of the mean of the retained sample is checked based on spectral time series analysis. If the half-width of the 95% interval of the mean is less than the mean multiplied by a small fraction (e.g. 0.1), then the test is passed.

Finally, in combination with a variety of diagnostic tools the Monte Carlo (MC) error is also usually reported. The MC error refers to error due to simulation and should not be confused with the posterior standard deviation or variance. Monitoring this type of error is the most basic diagnostic tool. In general, MC errors must be low so that quantities of interest are estimated with high precision. In the context of MCMC sampling the simplest way of calculating MC error is with the *batch mean* method where the MCMC sample is split into batches and then MC errors can be calculated as deviations of the batched means from the overall mean. Usually, a number of batches between 30 and 50 is sufficient (see e.g. Carlin and Louis, 1996).

2.7 Calculating quantities of interest given an MCMC sample

Having described the basic aspects of Bayesian theory and of MCMC simulation it is useful to provide some basic examples of how to calculate posterior and other summaries of interest given an MCMC sample after having successfully checked for convergence. Let us assume a final MCMC sample of size M, that is $\boldsymbol{\theta}^{(m)} = (\theta_1^{(m)}, \theta_2^{(m)}, \ldots, \theta_k^{(m)})^T$ for $m = 1, 2, \ldots, M$.

2.7.1 Calculating posterior quantities

The most basic point posterior estimate is the mean $\bar{\boldsymbol{\theta}} = (\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_k)^T$ which can be easily calculated by computing the individual means as follows

$$\bar{\theta}_j = M^{-1} \sum_{m=1}^{M} \theta_j^{(m)} ,$$

for $j = 1, 2, \ldots, k$. Having calculated the individuals means the corresponding variances are obtained by

$$Var(\theta_j) = (M-1)^{-1} \sum_{m=1}^{M} (\theta_j^{(m)} - \bar{\theta}_j)^2 ,$$

for $j = 1, 2, \ldots, k$, with the corresponding posterior standard deviations being simply $sd(\theta_j) = Var(\theta_j)^{1/2}$. The covariance of any given pair of parameters θ_i, θ_j for $i \neq j$ and $i, j = 1, 2, \ldots, k$ can be calculated as

$$Cov(\theta_i, \theta_j) = (M-1)^{-1} \sum_{m=1}^{M} (\theta_i^{(m)} - \bar{\theta}_i)(\theta_j^{(m)} - \bar{\theta}_j) .$$

The same line of reasoning applies for the calculation of any posterior point estimate, for instance different types of mean (e.g. trimmed-mean, geometric mean), median and percentile estimates. For most of the commonly used estimates, manual computations are rarely needed as the majority of modern statistical packages have standard routines for such measures.

An important aspect in any kind of statistical analysis is the calculation of the log-likelihood given the parameter estimates. As discussed in section 2.5 all model information criteria equal the log-likelihood plus some form of penalty term related to the inclusion of estimated parameters. Within the Bayesian framework and for usual univariate sampling problems where the sample is i.i.d. $x_i \sim p(x_i \mid \boldsymbol{\theta})$, M log-likelihood estimates can be calculated as

$$l(\mathbf{x} \mid \boldsymbol{\theta}^{(m)}) = \log p(\mathbf{x} \mid \boldsymbol{\theta}^{(m)}) = \sum\nolimits_{i=1}^{n} \log p(x_i \mid \boldsymbol{\theta}^{(m)}) ,$$

with the mean log-likelihood being equal to

$$\overline{l(\mathbf{x} \mid \boldsymbol{\theta})} = M^{-1} \sum_{m=1}^{M} l(\mathbf{x} \mid \boldsymbol{\theta}^{(m)}) .$$

This quantity is related to the mean deviance discussed in section 2.5 since $\overline{D(\boldsymbol{\theta})} = -2\overline{l(\mathbf{x} \mid \boldsymbol{\theta})}$. On the other hand the deviance based on the mean parameter vector is not equivalent to the above, the corresponding statistic $D(\bar{\boldsymbol{\theta}}) = -2l(\mathbf{x} \mid \bar{\boldsymbol{\theta}})$ is calculated as $l(\mathbf{x} \mid \bar{\boldsymbol{\theta}}) = \log p(\mathbf{x} \mid \bar{\boldsymbol{\theta}}) = \sum\nolimits_{i=1}^{n} \log p(x_i \mid \bar{\boldsymbol{\theta}})$.

2.7.2 Generating predictions and performing goodness-of-fit checks

As discussed in subsection 2.4.1 the predictive distribution is $p(\mathbf{x}^{pred} \mid \mathbf{x}) = \int p(\mathbf{x}^{pred} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta}$. In the majority of occasions the integral is difficult to calculate analytically, but given the MCMC sample of size M it is easy to generate M predictions as

$$\mathbf{x}^{pred(m)} \sim p(\mathbf{x}^{pred} \mid \boldsymbol{\theta}^{(m)}),$$

for $m = 1, 2, \ldots, M$. The likelihood function is usually a standard distribution from which it is easy to generate random observations, the majority of statistical software includes routines for random generators for a variety of distributions.

Similarly test quantities and Bayesian p-values are also easy to calculate through simulation. One calculates first $T(\mathbf{x}, \boldsymbol{\theta}^{(m)})$, $T(\mathbf{x}^{pred(m)}, \boldsymbol{\theta}^{(m)})$ and then the Bayesian p-value is computed by calculating the number of times that $T(\mathbf{x}^{pred(m)}, \boldsymbol{\theta}^{(m)})$ is greater than or equal to $T(\mathbf{x}, \boldsymbol{\theta}^{(m)})$ for $m = 1, 2, \ldots, M$ and by dividing this number with *M*. For instance, the Bayesian analogue of the chi-squared test is calculated through the test quantity

$$T(\mathbf{x}, \boldsymbol{\theta}) = \sum_{n=1}^{n} \frac{(x_i - E(x_i \mid \boldsymbol{\theta}))^2}{E(x_i \mid \boldsymbol{\theta})}$$

and the corresponding p-value is computed as

$$\text{Bayesian p-value} = M^{-1} \sum_{m=1}^{M} I\left(T(\mathbf{x}^{pred(m)}, \boldsymbol{\theta}^{(m)}) \geq T(\mathbf{x}, \boldsymbol{\theta}^{(m)})\right),$$

where $I\left(T(\mathbf{x}^{pred}, \boldsymbol{\theta}) \geq T(x, \boldsymbol{\theta})\right)$ is the indicator function given by

$$I\left(T(\mathbf{x}^{pred}, \boldsymbol{\theta}) \geq T(\mathbf{x}, \boldsymbol{\theta})\right) = \begin{cases} 1, & \text{if } T(\mathbf{x}^{pred}, \boldsymbol{\theta}) \geq T(\mathbf{x}, \boldsymbol{\theta}), \\ 0, & \text{if } T(\mathbf{x}^{pred}, \boldsymbol{\theta}) \geq T(\mathbf{x}, \boldsymbol{\theta}). \end{cases}$$

The chi-square discrepancy can be used as an overall goodness-of-fit test quantity, another common option is the deviance discrepancy, that is $T(\mathbf{x}, \boldsymbol{\theta}) = -2\sum_{i=1}^{n} \log(x_i \mid \boldsymbol{\theta})$.

2.8 Summary

In this chapter the key aspects of Bayesian theory were introduced and discussed in short. The main inferential tools were presented, namely location and dispersion measures, credible sets, the predictive distribution, Bayesian test quantities and p-values, and model selection criteria. In addition, the concept of MCMC simulation was discussed with emphasis given on the Metropolis-Hastings algorithm. Finally some practical guidelines for calculating quantities of interest given an MCMC sample were provided.

Bayesian theory, in general, provides a rather autonomous working framework which is strictly based on properties of conditional probabilities. As such it is relatively easy to understand as it requires only comprehension of the interrelationships of the prior, likelihood and posterior distributions. All of the subsequent aspects of Bayesian theory arise as simple by-products of these interrelationships.

3 Bayesian modeling of OD matrices

The topic of this chapter relates to a new, statistical, covariate-based approach for OD modeling under reliable information free from sampling and non-representation errors. The chapter begins with descriptions of the OD matrix derived from the 2001 Belgian census and the set of explanatory variables which is under consideration. The statistical models and methodology follow, starting from a simple comparison between the Poisson and the negative binomial models and extending to the framework of Poisson mixture models. Next, a brief but interesting note is provided which indicates a relationship between the proposed modeling approach and direct-demand, gravity modeling. The chapter concludes with a summary and discussion of results. The material presented in this chapter is based on Perrakis et al. (2011) and Perrakis et al. (2012a,c,e).

3.1 Data

This section includes descriptions for all the available data which are used for the purposes of this dissertation. Initially, an overview is provided for the OD matrix derived from the 2001 Belgian census and for the administrative structure of Flanders. Following that, the characteristics of two slightly different OD matrices for the Flemish region are described. Finally, the experimental design is discussed and descriptive statistics for the explanatory variables are provided.

3.1.1 The OD matrix from the 2001 Belgian census

The OD matrix was derived from the 2001 Belgian census which contains information about departure/arrival times and locations of commuting trips for the 10,296,350 Belgian residents. The recorded trips refer to work and school activities and are one-dimensional, that is from municipality of origin to municipality of destination. In addition, information is on a weekly basis, i.e. for the five weekdays, but without distinction concerning travel mode. Thus, OD data are for going-to-work/school trips on weekdays and for all travel modes.

Belgium consists of 589 municipalities of which 308 belong to the northern Dutch-speaking region of Flanders, 262 to the southern French-speaking region of Walloon and 19 to Brussels metropolitan area[19]. The study area in this dissertation is not the entire country of Belgium but the region of Flanders which accounts approximately for 60% of the total population in Belgium and 44% of the country's surface area. Therefore, trips within the region of Walloon and trips within Brussels metropolitan area are not under consideration. In addition, incoming trips to Flanders from Brussels metropolitan area and Walloon and outgoing trips from Flanders to Brussels metropolitan area and Walloon are also not considered. Thus, the OD matrix reflects the travel-demand for work and school activities strictly within Flanders and for residents of the Flemish region. In addition, interest lies on inferring for a "normal" or "average" weekday and not for each individual weekday. Therefore, the OD matrix was averaged across the five working days of the week.

3.1.2 The administrative structure of Flanders

The Flemish OD matrix on the level of municipalities contains 94,864 OD pairs. The level of municipalities is the lowest administrative level. Nevertheless, the matrix can also be aggregated to higher administrative levels since each Flemish municipality is also part of a canton, a district, an arrondissement and finally part of a province. Thus, there exists an inherent hierarchical administrative structure from the lowest spatial zoning level of municipalities to the highest spatial zoning level of provinces. In summary, Flanders consists of 308 municipalities, 103 cantons, 52 districts, 22 arrodissements and 5 provinces; the Flemish administrative structure is represented in Figure 3.1.

Provinces (5 zones, 25 OD pairs)
↑
Arrondissements (22 zones, 484 OD pairs)
↑
Districts (52 zones, 2704 OD pairs)
↑
Cantons (103 zones, 10609 OD pairs)
↑
Municipalities (308 zones, 94864 OD pairs)

Figure 3.1 The administrative levels of Flanders with the corresponding numbers of zones and OD pairs.

---

[19] Brussels metropolitan area although being inside the Flemish region is a separate administrative region.

The upward direction of the arrows in Figure 3.1 implies that the OD matrix of every higher level is essentially an aggregation of the OD matrix of the immediate lower administrative level. Given that the travel-demand in all administrative levels is of interest, the inherent hierarchical structure implies that modeling on the lower level of municipalities is immediately advantageous as predictive inference can then easily expand to all other levels by appropriately aggregating the OD flows.

3.1.3 OD descriptive statistics

The OD matrix derived from the Belgian census is free of sampling and non-representation errors, the most critical sources of error. Nevertheless, it is subjected to a certain non-response error since in some entries the origin is missing while in some others the destination is missing. On average the non-response rate is about 10%. The initial results, presented in section 3.2, are derived from a first OD matrix in which missing values were imputed based on the simple assumption that non-responding individuals are simply coming-from/travelling-to the largest – in terms of population – municipality within the corresponding canton. Later on, a second OD matrix was delivered where imputation was based on the assumption that non-responding individuals are essentially behaving in the same way as responding individuals, i.e. origin/destination totals from responding individuals were used as weights. Results of section 3.4 and all results after that section are based on the second OD matrix.

Regardless of differences, the flows in both OD matrices are in general scarcely distributed with the vast majority trips being zero-valued. In addition, huge outliers – especially in intra-zonal flows – characterize both matrices as extremely over-dispersed and also positively skewed. Such characteristics are in general expected on large-scale, real-world problems. In conjunction with the large data size the aforementioned attributes constitute both OD matrices difficult and complex datasets to analyze. Descriptive statistics are summarized in Table 3.1. The differences in means and medians are indicative of the positive skewness while the differences in means and variances reveal the extreme over-dispersion present in both OD matrices. The maximum in both matrices corresponds to the diagonal cell for intra-zonal flows in Antwerp which is the largest Flemish municipality in terms of size and population.

Statistic	**First OD matrix**	**Second OD matrix**
Mean	36.24	38.47
Median	0	0
Standard deviation	949.48	960.17
Variance	901,512.30	921,926.40
Maximum	222,149	211,681
Sum	3,437,168	3,649,514
% of 0-valued cells	63.13%	63.13%
% of contribution of intra-zonal trips to total trips	43.23%	51.20%

TABLE 3.1. Descriptive statistics for the first OD matrix (imputation based on largest municipality assumption) and for the second OD matrix (imputation based on behavior of responding-individuals assumption).

An important characteristic of large-scale matrices is that the OD flows are usually “inflated” across the main diagonal, that is intra-zonal flows are typically larger on average than inter-zonal flows. Apparently, that is also the case for the Flemish OD matrix; from the total travel-demand 43.23% in the first matrix and 51.2% in the second matrix corresponds to intra-zonal trips. Descriptive statistics for intra-zonal and inter-zonal trips are summarized in Table 3.2. The statistics in Table 3.2 reveal that both types of trips remain positively skewed with medians smaller than the respective means. In addition both types of trips have variances which grossly exceed the respective means and are thus extremely over-dispersed, especially the intra-zonal trips. Another finding from Table 3.2 is that all zero-valued cells – which account for 63.13% of the total number of cells in both matrices – appear in inter-zonal trips as the minimum values for intra-zonal trips and non-zero.

Graphical plots such as histograms and boxplots are not presented. Due to the complexity and size of the datasets such plots provide no particular meaningful additional information.

OD matrix	Statistic	Intra-zonal trips	Inter-zonal trips
First OD matrix	Mean	4824.04	20.64
	Median	2157.50	0
	Standard deviation	15,422.74	239.25
	Variance	237,860,909.11	57,240.56
	Minimum	9	0
	Maximum	222,149	13,426
	Sum	1,485,805	1,951,363
Second OD matrix	Mean	6,064.30	18.84
	Median	2,959.50	0
	Standard deviation	15,516.64	156.67
	Variance	240,776,116.89	24,545.49
	Minimum	12	0
	Maximum	211,681	9,542
	Sum	1,867,805	1,781,709

TABLE 3.2. Descriptive statistics for intra-zonal and inter-zonal trips in the first and second OD matrix.

3.1.4 Explanatory variables

The set of explanatory variables consists of six dummy variables and twelve covariates. The first five dummy variables are coded 0/100 so that they correspond to a difference of one hundred trips. These dummies capture intra-zonal effects by taking the value 100 if the trips are in diagonal cells (intra-zonal municipality trips) or diagonal blocks (intra-zonal canton, district, arrondissement and province trips), and 0 otherwise. In addition, the intra-zonal variables are constructed so as to capture individual effects; for instance, for intra-zonal trips in the main diagonal the municipality dummy will equal 100, whereas the province, arrondissement, district and canton dummies will equal 0. The sixth dummy is coded 0/1 and is associated with the effect of higher education institutes in destination zones; it takes the value of 1 if the destination zone supports a university college and/or a university and 0 otherwise.

The set of covariates includes four discrete-valued variables which contain the total number of neighboring municipalities on canton, district, arrondissement and province levels for each corresponding OD pair. The rest of the covariates are purely continuous, namely employment ratio, population density, relative length of road networks, perimeter length, car ownership ratio, kilometer's driven in highways and in provincial/municipal roads, and finally distance.

#	Name	Description of Variable
1	dummy.province	Dummy variable for intra-zonal province trips (0/100)
2	dummy.arron/ment	Dummy variable for intra-zonal arrondissement trips (0/100)
3	dummy.district	Dummy variable for intra-zonal district trips (0/100)
4	dummy.canton	Dummy variable for intra-zonal canton trips (0/100)
5	dummy.municipality	Dummy variable for intra-zonal municipality trips (0/100)
6	dummy.education	Dummy variable for destination zones with a college or university (0/1)
7	munic.in.cantons	Number of municipalities between the cantons of origin and destination
8	munic.in.districts	Number of municipalities between the districts of origin and destination
9	munic.in.arron/ments	Number of municipalities between the arrondissements of origin and destination
10	munic.in.provinces	Number of municipalities between the provinces of origin and destination
11	employment.ratio.o	Employment ratio of origin-zone
12	employment.ratio.d	Employment ratio of destination-zone
13	population.density.o	Population density of origin-zone (thousand inhabitants per square km)
14	population.density.d	Population density of destination-zone (thousand inhabitants per square km)
15	road.length.o	Length of road network relative to surface of origin-zone (km per square km)
16	road.length.d	Length of road network relative to surface of destination-zone (km per square km)
17	perimeter.length.o	Perimeter of origin-zone (in km's)
18	perimeter.length.d	Perimeter of destination-zone (in km's)
19	car.ownership.ratio.o	Car ownership ratio of origin-zone
20	car.ownership.ratio.d	Car ownership ratio of destination-zone
21	highway.traffic.o	Km's driven per year in highway roads of origin-zone (traffic in millions)
22	highway.traffic.d	Km's driven per year in highway roads of destination-zone (traffic in millions)
23	prov/munic.traffic.o	Km's driven per year in provincial and municipal roads of origin-zone (traffic in millions)
24	prov/munic.traffic.d	Km's driven per year in provincial and municipal roads of destination-zone (traffic in millions)
25	distance	Distance between origin-zone and destination-zone (in km's)

TABLE 3.3. Names and descriptions of the 25 explanatory variables.

All covariates are used in logarithmic scale. Distance, of course, is zero for intra-zonal municipality flows and in order to use the logarithm it is set equal to 0.1, a value which for most practical purposes refers to negligible distance (100 meters). In addition, highways do not pass through all municipalities and therefore the kilometer's-driven-in-highways variable is also zero-valued for certain OD pairs. In order to use the logarithm this variable was set to equal to 10^{-7} when being zero. Highway traffic is in million's of kilometers, therefore this value corresponds to one driven kilometer. Furthermore, due to the particularity of the OD problem the purely continuous variables come in pairs, i.e. each is used twice, one time for the origin-zone and one time for the destination-zone, except of distance which is unique for each corresponding OD pair. Reasons for adopting this particular experimental design are discussed next. The full set of 25 explanatory variables is listed in Table 3.3.

The arguments in favor of employing the continuous variables in pairs are the following:

a. Preliminary research revealed that it is better to use full information for origin and destination zones separately rather than average, for instance, between origin and destination zones[20].
b. Having separate parameters estimates for origin and destination zones allows for elementary comparison with trip-production and trip-attraction studies.
c. Using pairs on logarithmic scale and including distance provides an alternative interpretation of the Poisson log-linear models – presented next – as stochastic gravity-type, direct-demand models[21].

Descriptive statistics for the 12 covariates in natural-scale and log-scale are provided in Tables 3.4 and 3.5, respectively. For the continuous variables which are used in pairs, distinction between origin and destination is not necessary as these variables are essentially the same, i.e. they are simply on different ordering.

[20] In addition, as we will see in sections 3.2.4 and 3.3.5, some of the signs of parameter estimates are opposite for origin and for destination, implying an inverse relationship with respect to OD flows.
[21] This point is further explained in section 3.4.

Variable	Mean	St.dev.	Median	Min	Max
munic.in.cantons	8.052	2.883	8	2	18
munic.in.districts	16.857	6.836	16	2	40
munic.in.arron/ments	38.857	14.176	40	10	70
munic.in.provinces	125.857	11.367	130	88	140
empoymentl.ratio	0.422	0.031	0.428	0.233	0.571
population.density	0.518	0.444	0.381	0.052	3.153
road.length	4.809	1.912	4.378	2.062	17.135
perimeter.length	36.621	13.580	34.984	6.265	97.023
car.ownership.ratio	0.470	0.019	0.469	0.415	0.551
highway.traffic	66.426	128.476	0	0	1427.96
prov/munic.traffic	110.649	135.707	81.85	2.5	1418.8
distance	49.507	28.575	44.440	0	155.16

TABLE 3.4. Mean, standard deviation, median, minimum and maximum for each of the 12 covariates on natural-scale.

Variable	Mean	St.dev.	Median	Min	Max
munic.in.cantons	2.018	0.378	2.079	0.693	2.890
munic.in.districts	2.739	0.424	2.773	0.693	3.689
munic.in.arron/ments	3.584	0.406	3.689	2.303	4.248
munic.in.provinces	4.831	0.097	4.868	4.477	4.942
empoymentl.ratio	-0.866	0.078	-0.849	-1.457	-0.560
population.density	-0.924	0.714	-0.966	-2.952	1.148
road.length	1.507	0.346	1.477	0.724	2.841
perimeter.length	3.528	0.397	3.555	1.835	4.575
car.ownership.ratio	-0.755	0.040	-0.756	-0.880	-0.597
highway.traffic	-4.937	9.149	-13.816	-13.816	7.264
prov/munic.traffic	4.340	0.859	4.405	0.916	7.258
distance	3.693	0.754	3.794	-2.303	5.044

TABLE 3.5. Mean, standard deviation, median, minimum and maximum for each of the 12 covariates on log-scale.

As shown in Table 3.3, most of the continuous explanatory variables are transformed to ratios relative to populations or surface areas. The specific transformations are chosen in order to maintain reasonable interpretations, but also in order to solve multi-collinearity problems which are evidently present in the raw variables. For instance, variables such as working and total populations are highly correlated. Analysis of multi-collinearity based on variance inflation factors (VIF) indicates no serious multi-collinearity for the transformed variables. The VIF values of the explanatory variables are presented in Table 3.6, as shown the highest VIF value equals 3.877[22].

Variable	Variance inflation factor
dummy.province	1.442
dummy.arron/ment	1.251
dummy.district	1.153
dummy.canton	1.191
dummy.municipality	1.566
dummy.education	1.299
munic.in.cantons	1.891
munic.in.districts	3.877
munic.in.arron/ments	2.430
munic.in.provinces	1.956
employment.ratio.o	1.089
employment.ratio.d	1.094
population.density.o	3.043
population.density.d	3.195
road.length.o	1.828
road.length.d	1.829
perimeter.length.o	2.671
perimeter.length.d	2.771
car.ownership.ratio.o	1.263
car.ownership.ratio.d	1.266
highway.traffic.o	1.248
highway.traffic.d	1.262
prov/munic.traffic.o	2.393
prov/munic.traffic.d	2.404
distance	2.555

TABLE 3.6. Variance inflation factors for the 25 explanatory variables.

In summary, the selection of explanatory variables listed in Table 3.3 is a combination of variables that can be derived immediately from the hierarchical structure of the OD matrix and of continuous explanatory variables. Variables 1 to 9 are extracted directly from the hierarchical structure of the OD matrix.

[22] A common rule of thumb is that VIF values greater than 5 are suggestive of high multi-collinearity. Some authors propose the value of 10 as a threshold value (e.g. Kutner et al., 2004). For an interesting discussion see also O'Brien (2007).

Variables related to population, surface area, perimeter, car ownership, yearly traffic and distance are typically available in transportation research centers and institutes. Finally, variables related to length of road networks were obtained by the Belgian governmental website for statistics (FOD Economie, 2010). In general, the explanatory set is relatively simple, economical and also easy to obtain.

3.2 Modeling over-dispersion: Poisson versus negative binomial regression

The models presented in this section were applied on the first OD matrix and the set of explanatory variables is the one presented in Table 3.3, but without variable 6 (dummy.education) which was included on a later phase of the analysis. Therefore, the current models have in total 24 regression parameters. In addition, due to a lack of specific prior information about the parameters, non-informative priors are adopted. Nevertheless, the descriptions in subsections 3.1.1 and 3.1.2 start from general prior-forms which can be utilized in an informative framework as well.

For computational and notational convenience the vector notation introduced in section 1.1 is used, i.e. $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ is the vector of OD flows from the census and n is the number of OD pairs. In addition, let p denote the number of explanatory variables, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)^T$ the vector of unknown parameters and $\mathbf{X}$ the *design matrix* of dimensionality $n \times (p+1)$, containing a column of 1's for the intercept and also the p explanatory variables, with $\mathbf{x}_i = (1, x_{1i}, x_{2i}, \ldots, x_{pi})^T$ being the i-th row of $\mathbf{X}$ related to OD flow y_i and $i = 1, 2, \ldots, n$. In general, the models presented next, relate the expected value of the vector of OD flows $\mathbf{y}$ to the explanatory variables in $\mathbf{X}$ through the parameter vector $\boldsymbol{\beta}$ and an invertible function g, i.e. $g(E(\mathbf{y})) = \mathbf{X}^T\boldsymbol{\beta}$ and consequently $E(\mathbf{y}) = g^{-1}(\mathbf{X}^T\boldsymbol{\beta})$.

In the context of Generalized Linear Models (GLMs) the function g is known as the *link function* (McCullagh and Nelder, 1989). The common assumption for function g in Poisson and negative binomial regression is the logarithmic function, this assumption is also adopted here. The log-link function implies the assumption that the effects of the explanatory variables are additive and linear to the log-mean of y_i, that is $\log E(y_i) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}$. Consequently, the effects are multiplicative and exponential on natural scale since $E(y_i) = e^{\beta_0} \cdot (e^{\beta_1})^{x_{1i}} \cdot \ldots \cdot (e^{\beta_p})^{x_{pi}}$.

3.2.1 The Poisson model

The likelihood assumption is that the OD flows are i.i.d. (identically and independently distributed) Poisson realizations, that is $y_i \sim Pois(\mu_i)$ for $i = 1, 2, \ldots, n$, where μ_i is the Poisson mean which is related to the explanatory variables through the log-link function $\log \mu_i = \boldsymbol{\beta}^T \mathbf{x}_i$. The data likelihood is given by

$$p(\mathbf{y} \mid \boldsymbol{\beta}) = \prod_{i=1}^{n} p(y_i \mid \boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{e^{-e^{\boldsymbol{\beta}^T \mathbf{x}_i}} (e^{\boldsymbol{\beta}^T \mathbf{x}_i})^{y_i}}{y_i!}. \quad (3.1)$$

Poisson regression is a common option when modeling count data and it is frequently used in practice. Nevertheless, the Poisson model usually does not perform well for cases of over-dispersed datasets due to the property of the Poisson distribution that the mean is equal to the variance, that is $E(\mathbf{y} \mid \boldsymbol{\beta}) = \mathrm{var}(\mathbf{y} \mid \boldsymbol{\beta}) = e^{\mathbf{X}^T \boldsymbol{\beta}}$. Properties and estimation procedures for Poisson regression can be found in McCullagh and Nelder (1989) and Agresti (2002), Bayesian applications are presented in Ntzoufras (2009).

Proceeding with the specification of the prior distribution for parameter vector $\boldsymbol{\beta}$, a common choice within the framework of generalized linear models (GLMs) is a multivariate normal prior (Gelfand and Ghosh, 2000). This prior has the form

$$p(\boldsymbol{\beta}) = \frac{1}{(2\pi)^{(p+1)/2} \left|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right|} e^{-\frac{(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})}{2}}. \quad (3.2)$$

If prior knowledge for vector $\boldsymbol{\beta}$ is available, then the location vector $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and/or the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ can be set accordingly. Alternatively when prior information is not available, a diffuse prior may be adopted, by setting for instance $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = n \cdot (\mathbf{X}^T\mathbf{X}) \cdot 10^{-3}$, which is analogue to one of the "benchmark" priors suggested in Fernández et al. (2001).

From Bayes theorem, the posterior distribution of $\boldsymbol{\beta}$ is $p(\boldsymbol{\beta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta})p(\boldsymbol{\beta})$. From expressions (3.1) and (3.2) the resulting posterior distribution is

$$p(\boldsymbol{\beta} \mid \mathbf{y}) \propto \prod_{i=1}^{n}\left[e^{-e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i}}(e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i})^{y_i}\right] \times e^{-(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})/2} . \quad (3.3)$$

Direct inference from the posterior distribution is not feasible, since expression (3.3) does not result to a known distributional form.

3.2.2 The negative binomial model

The likelihood assumption for this model is that the OD flows are negative binomially distributed, i.e. $y_i \sim NB(\mu_i, \theta)$ where again $\log \mu_i = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i$ for $i = 1, 2, \ldots, n$. The data likelihood is

$$p(\mathbf{y} \mid \boldsymbol{\beta}, \theta) = \prod_{i=1}^{n} \frac{\Gamma(y_i + \theta)}{\Gamma(\theta)y_i\,!} \frac{(e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i})^{y_i}\theta^{\theta}}{(e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i} + \theta)^{y_i+\theta}} . \quad (3.4)$$

The mean of the data in this case is $E(\mathbf{y} \mid \boldsymbol{\beta}) = e^{\mathbf{X}^T\boldsymbol{\beta}}$, whereas the variance is $\mathrm{var}(\mathbf{y} \mid \boldsymbol{\beta}, \theta) = e^{\mathbf{X}^T\boldsymbol{\beta}} + (e^{\mathbf{X}^T\boldsymbol{\beta}})^2\theta^{-1}$. Note that the variance in this case is a quadratic function of the mean. Thus, negative binomial regression incorporates over-dispersion since the assumed variance always exceeds the assumed mean. It is easy to see that when $\theta^{-1} \to 0$ the negative binomial distribution converges to the Poisson distribution since $\mathrm{var}(\mathbf{y} \mid \boldsymbol{\beta}, \theta) \to E(\mathbf{y} \mid \boldsymbol{\beta})$. Therefore, θ^{-1} is also referred to as the *dispersion* parameter (Agresti, 2002).

For parameter vector $\boldsymbol{\beta}$, the multivariate normal distribution defined in equation (3.2) is used. Regarding parameter θ, a $Gamma(\alpha, \delta)$ distribution with shape $\alpha > 0$ and rate $\delta > 0$ is chosen. The prior of θ is given by

$$p(\theta) = \frac{\delta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\delta\theta} . \quad (3.5)$$

Under the parameterization in equation (3.5) $E(\theta) = \alpha / \delta$ and $\mathrm{var}(\theta) = \alpha / \delta^2$. Parameters α and δ may be set accordingly if prior knowledge regarding the expectation or the variance of θ is available. Otherwise, a common option is to set α=δ=0.001 which results to a diffuse prior with mean equal to 1 and variance equal to 1000, as in Ntzoufras (2009).

The joint posterior distribution of both $\boldsymbol{\beta}$ and θ is now $p(\boldsymbol{\beta},\theta \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta},\theta)p(\boldsymbol{\beta})p(\theta)$ which leads to expression

$$p(\boldsymbol{\beta},\theta \mid \mathbf{y}) \propto \prod_{i=1}^{n}\left[\frac{\Gamma(y_i+\theta)}{\Gamma(\theta)y_i!}\frac{(e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i})^{y_i}\theta^{\theta}}{(e^{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i}+\theta)^{y_i+\theta}}\right]\times e^{-(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})/2}\times\theta^{\alpha-1}e^{-\delta\theta}\,. \quad (3.6)$$

Inference from the posterior distribution is once again not straightforward since the normalizing constant of the density in (3.6) is not known.

It should be noted that the over-dispersed negative binomial model is derived from a *Poisson-gamma mixture* model or else stated *hierarchical* model which is of the form $y_i \sim Pois(\mu_i u_i)$ where $\log\mu_i = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i$ and $u_i \sim Gamma(\theta,\theta)$ for $i = 1,2,\ldots,n$. The negative binomial distribution in expression (3.4) is the corresponding *marginal sampling likelihood* and is obtained by integrating over the *random vector* $\mathbf{u} = (u_1,u_2,\ldots,u_n)^T$, i.e. $p(\mathbf{y} \mid \boldsymbol{\beta},\theta) = \int p(\mathbf{y} \mid \boldsymbol{\beta},\mathbf{u})p(\mathbf{u} \mid \theta)d\mathbf{u}$. A more in-depth discussion about the relation between the hierarchical Poisson-gamma and marginal negative binomial models is provided in section 3.3 under the general context of Poisson mixture models.

For the time being it is worth noting that by means of Bayesian methodology one might choose to fit either the hierarchical or the marginal form of the model. In both cases the estimates for the parameters of main scientific interest $\boldsymbol{\beta}$ and θ will be the same due to the equivalence of the two models. The hierarchical Poisson-gamma model provides additional information over the posterior distribution of $\mathbf{u}$ but also requires sampling the full set of parameters $(\boldsymbol{\beta},\mathbf{u},\theta)$. The marginal negative binomial model on the other hand is simpler to fit, especially for large datasets, since estimation is restricted to the reduced set of parameters $(\boldsymbol{\beta},\theta)$. In addition, due to the convenient conjugate relationship between the Poisson and the gamma distribution vector $\mathbf{u}$ can be sampled subsequently from its respective *full conditional* distribution which is again a gamma distribution, namely $\mathbf{u} \mid \boldsymbol{\beta},\theta,\mathbf{y} \sim Gamma(\mathbf{y}+\theta, e^{\mathbf{X}^T\boldsymbol{\beta}}+\theta)$.

3.2.3 Implementation and performance of Metropolis-Hastings simulation

A MH algorithm is utilized in order to generate posterior samples from the corresponding normalized densities of expressions (3.3) and (3.6). In particular, we employ an independence chain algorithm where the location and scale parameters of the proposal distributions remain fixed. Large data sizes result to considerable time-consuming calculations and independence chain MH simulation performs faster than random-walk chain MH or other types of Metropolis-within-Gibbs algorithms due to its simplified algorithmic form.

For the proposal distribution of parameter $\boldsymbol{\beta}$, common in both the Poisson and the negative binomial model, a multivariate normal distribution is adopted, i.e. $q(\boldsymbol{\beta}) \sim \mathbf{N}_{p+1}(\boldsymbol{\beta}^{ML}, \mathbf{V}_{\boldsymbol{\beta}}^{ML})$, where $\boldsymbol{\beta}^{ML}$ is the maximum-likelihood (ML) estimate of $\boldsymbol{\beta}$ and $\mathbf{V}_{\boldsymbol{\beta}}^{ML}$ is the estimated covariance matrix of $\boldsymbol{\beta}^{ML}$. Regarding parameter θ of the negative binomial model, the proposal distribution is defined as $q(\theta) \sim Gamma(\tilde{a}, \tilde{b})$, where parameters $\tilde{a}$ and $\tilde{b}$ are set so to satisfy $\tilde{a} / \tilde{b} = \theta^{ML}$ and $\tilde{a} / \tilde{b}^2 = \mathrm{var}(\theta^{ML})$ with θ^{ML} being the ML estimate of θ. Regarding the specification of starting values, random samples may be generated first from the proposal distributions and then the mean, the median or a specific percentile point from these random samples may be used as starting point $\boldsymbol{\beta}^0$ and θ^0. The algorithms for both models are presented in Appendix A.

Thus, the ML estimates of the 2 models were calculated first and samples of 5,000 draws were generated subsequently from the proposal distributions. The corresponding 90th percentile estimates from these samples were set as starting values. One single MH chain was utilized for each model. After certain preliminary tests, the final MH simulations ran for 5,000 iterations of the Poisson model and for 21,000 iterations of the negative binomial model with resulting acceptance ratios of 97% and 68%, respectively. The first 1,000 iterations were discarded as the “burn-in” part for both models. Convergence checks were based on the methods of Raftery and Lewis (1992), Geweke (1992) and Heidelberger and Welch (1983). The posterior sample of the Poisson model passed all the diagnostics. Regarding the negative binomial model, the diagnostic of Raftery and Lewis (1992) indicated autocorrelation problems. In order to break the strong autocorrelations, every 4^{th} draw of the sample was kept. In the final posterior sample of 4,000 draws, all lag 1 autocorrelations were in absolute value below 0.05. Results concerning convergence diagnostics and also MC errors for the final posterior samples of both models are presented in Appendix A.

3.2.4 Posterior inference from the Poisson and negative binomial models

The results of this section concern the regression parameters β_j, for $j = 0, 1, 2, \ldots, 24$, and dispersion parameter θ. The effects of the regression parameters on the mean OD flows is additive on logarithmic scale and therefore interpretation is straightforward. Posterior means greater than 0 correspond to an increasing additive effect, whereas posterior means less than 0 have a decreasing additive effect. Specifically, the effects of the dummy variables, with parameters β_1 to β_5, on the logarithm of the mean OD flows are simply the posterior mean multiplied by 100. For the rest of the explanatory variables, with corresponding parameters β_6 to β_{24}, a one-unit increase of an explanatory variable, given that the rest remain constant, would result to a change in the logarithm of the mean OD flows equal to the posterior mean of the corresponding parameter. The posterior means, standard deviations and 95% credible intervals for parameters β_j and parameter θ, calculated from the 4,000 posterior draws of the Poisson and negative binomial models, are summarized in Table 3.7.

Model comparison is based on the DIC introduced by Spiegelhalter et al. (2002). As in all information criteria support is given to models with lower values of DIC. The DIC values for the two models are also presented in Table 3.7. According to these values the negative binomial model provides a significant improvement in model fit as it has a DIC value which is extremely lower than the corresponding value of the Poisson model. This difference also explains the striking differences between regression estimates of the two models. Evidently, the Poisson distribution is a completely inappropriate assumption under the presence of extreme over-dispersion. On the other hand the negative binomial distribution with just one extra parameter – parameter θ accounting for over-dispersion – is a far more appropriate starting assumption for the problem at hand. The negative binomial model becomes our main focus of interest for the remainder of this section, issues of goodness-of fit are pursued further next.

Parameter	**Poisson**			**Negative binomial**		
	Mean	**St.dev**	**95% C.I.**	**Mean**	**St.dev**	**95% C.I.**
β_0 intercept	7.1351	0.03294	(7.0696,7.2001)	3.6431	0.41321	(2.8415,4.4550)
β_1 dummy.province	0.0167	0.00003	(0.0167,0.0168)	0.0070	0.00014	(0.0067,0.0073)
β_2 dummy.arron/ment	0.0169	0.00003	(0.0168,0.0169)	0.0079	0.00023	(0.0075,0.0083)
β_3 dummy.district	0.0155	0.00004	(0.0154,0.0156)	0.0092	0.00033	(0.0085,0.0098)
β_4 dummy.canton	0.0174	0.00004	(0.0173,0.0174)	0.0087	0.00038	(0.0080,0.0095)
β_5 dummy.municipality	-0.0281	0.00009	(-0.0283,-0.0280)	-0.0771	0.00084	(-0.0788,-0.0755)
β_6 munic.in.cantons	0.0033	0.00176	(-0.0002,0.0067)	0.5843	0.01854	(0.5482,0.6201)
β_7 munic.in.districts	0.2073	0.00189	(0.2037,0.2110)	-0.4653	0.02426	(-0.5128,-0.4160)
β_8 munic.in.arron/ments	-0.3761	0.00135	(-0.3788,-0.3735)	-0.2023	0.01772	(-0.2375,-0.1681)
β_9 munic.in.provinces	-1.5395	0.00576	(-1.5510,-1.5280)	-0.9058	0.07296	(-1.0470,-0.7619)
β_{10} employment.ratio.o	-0.4541	0.00750	(-0.4686,-0.4396)	-1.0723	0.07474	(-1.2227,-0.9289)
β_{11} employment.ratio.d	0.2097	0.00669	(0.1968,0.2230)	0.5700	0.06990	(0.4352,0.7080)
β_{12} population.density.o	0.1497	0.00181	(0.1462,0.1532)	0.4775	0.01417	(0.4492,0.5056)
β_{13} population.density.d	0.9442	0.00215	(0.9400,0.9484)	0.8243	0.01429	(0.7962,0.8519)
β_{14} road.length.o	-0.3204	0.00272	(-0.3257,-0.3152)	-0.2833	0.02274	(-0.3296,-0.2387)
β_{15} road.length.d	-0.1939	0.00328	(-0.2003,-0.1874)	0.2510	0.02203	(0.2093,0.2949)
β_{16} perimeter.length.o	0.6353	0.00264	(0.6301,0.6404)	1.2918	0.02370	(1.2453,1.3383)
β_{17} perimeter.length.d	1.2313	0.00329	(1.2249,1.2380)	0.8831	0.02258	(0.8396,0.9269)
β_{18} car.ownership.ratio.o	0.2061	0.01889	(0.1678,0.2426)	2.9009	0.15321	(2.5961,3.1967)
β_{19} car. ownership.ratio.d	1.1075	0.02115	(1.0659,1.1485)	-1.8642	0.15933	(-2.1750,-1.5525)
β_{20} highway.traffic.o	-0.0011	0.00021	(-0.0015,-0.0007)	0.0080	0.00173	(0.0046,0.0114)
β_{21} highway.traffic.d	0.0218	0.00025	(0.0214,0.0223)	0.0529	0.00179	(0.0494,0.0564)
β_{22} prov/munic.traffic.o	0.0197	0.00125	(0.0172,0.0222)	0.2419	0.01043	(0.2216,0.2625)
β_{23} prov/munic.traffic.d	0.7323	0.00173	(0.7289,0.7356)	0.8884	0.01027	(0.8683,0.9086)
β_{24} distance	-1.6599	0.00151	(-1.6629,-1.6569)	-2.8404	0.01097	(-2.8623,-2.8197)
θ theta	–	–	–	1.0139	0.00951	(0.9952,1.0324)
DIC		2526726			263198.4	

TABLE 3.7. Posterior means, standard deviations and 95% credible intervals for the parameters of the Poisson and negative binomial models with the corresponding values of DIC.

In addition to posterior point estimates and intervals, direct examination of the posterior distribution often provides extra information and a more comprehensive view regarding the random nature of parameters. Kernel smoothed estimates of the 26 posterior distributions for the parameters of the negative binomial model are presented in Figure 3.2.

Statistical significance of regression parameters may be checked by examining the 95% credible intervals in Table 3.7 or the posterior distribution of Figure 3.2. For the negative binomial model none of the intervals includes the value of 0 and therefore all regression parameters have statistically significant effects. In lack of prior knowledge, it is not possible to compare the parameter estimates to older estimates. Nonetheless, the current regression estimates provide useful insights into how the explanatory variables affect OD flows on overall and also illustrate some relationships with traditional travel demand modeling variable selection during the trip-generation and trip-distribution steps. This discussion is deferred to subsection 3.3.4 where the sensitivity of the negative binomial estimates is additionally tested under slightly different distributional assumptions.

At present it is interesting to assess the impact of the explanatory variables on the OD flows and determine which variables are the most influential. To this end, one can simply divide the posterior means, presented in Table 3.7, by the corresponding standard deviations and examine which values are larger in absolute terms. Distance (β_{24}) is the most influential variable of the model with a mean over standard deviation value equal to -258.93. This result is not surprising and confirms the importance of distance with respect to travel demand modeling theory in which distance is perhaps the most crucial factor. Following distance, the most influential effects among the continuous variables are those of yearly kilometers-driven in provincial/municipal roads (β_{23}) and population density (β_{13}) for destination-zones, with respective values 86.50 and 57.68, while perimeter length of origin-zones (β_{16}) is the fourth most influential variable with a value of 54.51. Overall, the general variables of population density, perimeter-length and yearly kilometers-driven are more influential in comparison to the specific variables of employment ratio, relative length of road networks and car-ownership ratio. Among the latter category of variables the

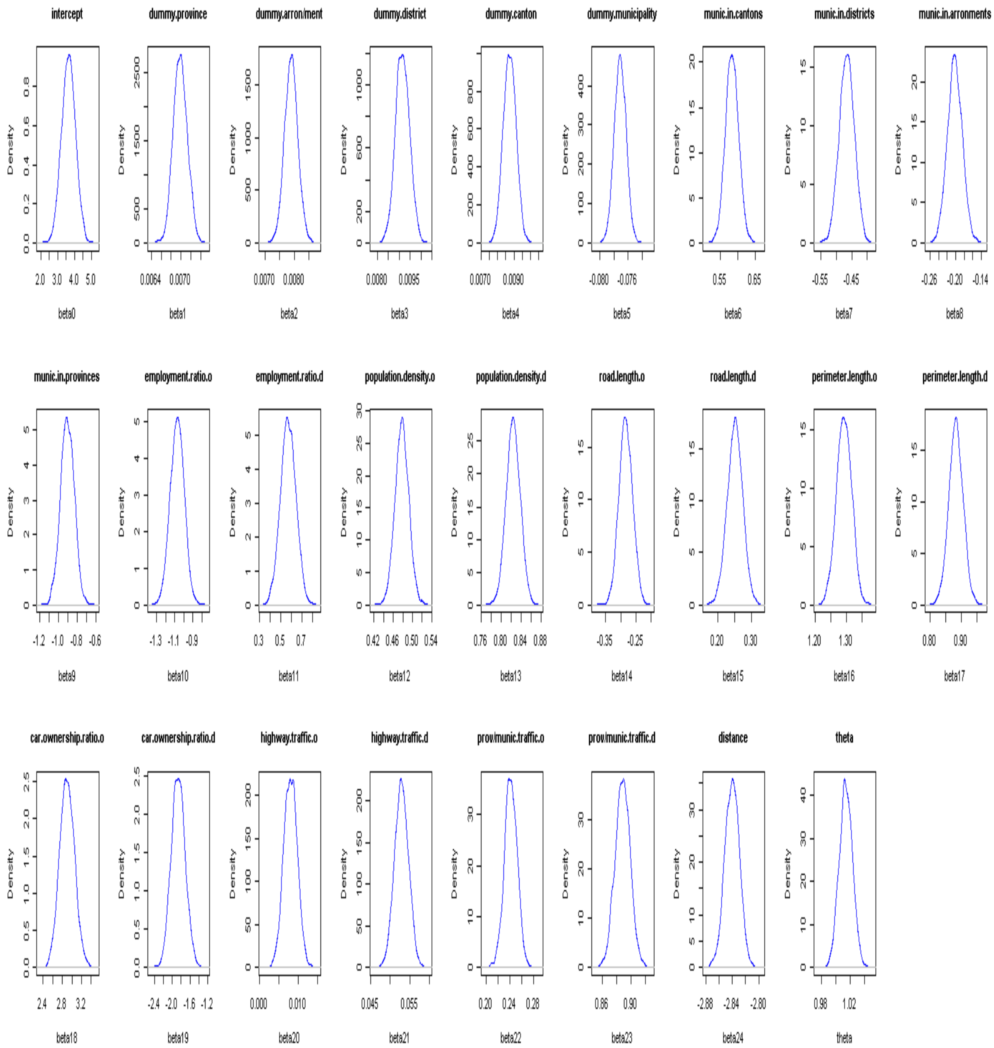

Figure 3.2 Kernel distribution estimates from 4,000 posterior draws of the 26 parameters of the negative binomial model.

most influential variable is car-ownership ratio for origin-zones (β_{18}) with a value of 18.93. Finally, it is interesting to note that the influence of the dummy variables is not only substantially strong but also analogous to their hierarchical structure; parameters β_1 to β_4, which have a positive effect and correspond to the levels of provinces, arrondissements, districts and cantons, are decreasingly influential with values equal to 50, 34.35, 27.88 and 22.90, respectively. On the other hand, the negative effect[23] of intra-zonal municipality trips (β_5) has a posterior mean over standard deviation value equal to -91.79, thus exceeding in absolute terms the corresponding values of the other dummy variables.

3.2.5 Predictive inference from the Poisson-gamma structure

The overall goodness-of-fit of the negative binomial model is initially assessed by the generalized version of the coefficient of determination R^2. The generalized R^2 measure provides the proportion of variability in a dataset which is accounted for by a given statistical model (Nagelkerke, 1991). The measure is generally defined as

$$R^2 = 1 - (L(0) \,/\, L(\hat{\boldsymbol{\theta}}))^{2/n},$$

where the term $L(0)$ is the likelihood of the *null* model – containing only the intercept – and $L(\hat{\boldsymbol{\theta}})$ is the likelihood of an estimated model, with n denoting sample size. In our context, the measure is calculated as

$$R^2 = 1 - (p(\mathbf{y} \mid \beta_0^{null}, \theta^{null}) \,/\, p(\mathbf{y} \mid \boldsymbol{\beta}, \theta))^{2/n},$$

where $p(\mathbf{y} \mid \cdot)$ is the negative binomial likelihood and β_0^{null}, θ^{null} are the intercept and over-dispersion parameters of the null model. The measure was calculated over 4,000 posterior draws of the null model and the model with the explanatory variables[24]. The average value of R^2 is equal to 0.73056 and the corresponding 95% credible interval is (0.73047, 0.73062). Thus, a significant amount of variation, approximately 73%, in the OD matrix is accounted for by the explanatory variables[25].

Evaluation of model fit is supplemented by posterior predictive tests based on Bayesian p-values. As mentioned in section 3.2.2 prediction is based on the hierarchical Poisson-gamma structure which includes random effects. Due to

---

[23] Possible explanations and interpretations for the signs of the parameters are discussed in subsection 3.3.4.

[24] An additional MH simulation was executed for the null model. Acceptance ratio for this simple model was very high, equal to 99%. A final sample of 4000 draws was kept for parameters β_0^{null} and θ^{null}.

[25] According to Nagelkerke (1991), variation is defined in general "as the extent to which a distribution is not degenerate". The generalized R^2 is consistent with the classical R^2 for normal linear regression cases.

memory limitations, the complete posterior sample could not be utilized for predictive purposes. Therefore, a random sample of 500 posterior draws was selected without replacement from the complete sample. Following that, 500 posterior draws of the random effect vector were sampled first and then 500 predictive vectors were generated from the hierarchical Poisson likelihood. Specifically, $\mathbf{u}^{(m)} \sim Gamma(\mathbf{y}+\theta^{(m)}, e^{\mathbf{X}^T\boldsymbol{\beta}^{(m)}}+\theta^{(m)})$ and $\mathbf{y}^{pred(m)} \sim Pois(e^{\mathbf{X}^T\boldsymbol{\beta}^{(m)}}\mathbf{u}^{(m)})$ for $m=1,2,\dots,500$[26].

Two general measures of discrepancy between observed and predicted data are under consideration. Namely, the *absolute distances* and the *squared distances* of observed and predicted data from the corresponding expected values of the negative binomial model. The test quantity for absolute distances is denoted by T_1 and the test quantity for squared distance by T_2, the two measures are defined as

$$T_1(\mathbf{y},\boldsymbol{\beta},\theta) = \sum_{i=1}^{n}\left|y_i - E(y_i \mid \boldsymbol{\beta},\theta)\right|,$$
$$T_2(\mathbf{y},\boldsymbol{\beta},\theta) = \sum_{i=1}^{n}\left(y_i - E(y_i \mid \boldsymbol{\beta},\theta)\right)^2.$$

In general, the absolute distance is a strict measure which assigns greater penalty to small deviations from the data, whereas the squared distance measure gives more weight to large deviations from the data. The resulting Bayesian p-values are 0.522 and 0.462 for absolute distance and squared distance, respectively. The p-values are close to the ideal value of 0.5, therefore the fit of the model is found to be satisfactory under both test quantities.

In combination with overall goodness-of-fit tests one can additionally check the proximity of predictions from any aspect that is of particular interest. As mentioned in section 3.1.2, modeling on the level of municipalities allows for predictive inference on various levels of aggregation. For instance, predictions for OD flows between districts can be derived directly as aggregations of predictive flows between municipalities. Thus, predictive inference is not necessarily restricted on the level of municipalities and can be applied on any other hierarchical level. In addition, prediction may also focus on specific trip-types such as strictly in-coming trips, strictly out-coming trips or internal trips.

[26] For notational simplicity $\mathbf{u}^{(m)} \equiv \mathbf{u}^{(m)} \mid \boldsymbol{\beta}^{(m)}, \theta^{(m)}, \mathbf{y}$.

Applications of prediction on different levels of aggregation and for different types of trips are demonstrated in Figure 3.3. The applications correspond to predictions for (a) total number of in-coming, going-to-work/school trips from all other municipalities to the capital of Flanders, Antwerp, (b) total number of going-to-work/school trips that occur daily within Flanders and (c-g) daily internal going-to-work/school trips that take place in each of the five Flemish provinces. It is worth noting that these predictions also serve as further goodness-of-fit tests, since in every case there is an observed quantity to compare with, resulting in a corresponding Bayesian p-value. As illustrated in Figure 3.3, all observed quantities are within acceptable density regions of the predictive distributions, an indication that the predictions are not extreme with respect to the initial data. The most extreme p-value, which corresponds to internal going-to-work/school trips for the province of Flemish Brabant, equals 0.89 and is still within tolerable limits[27].

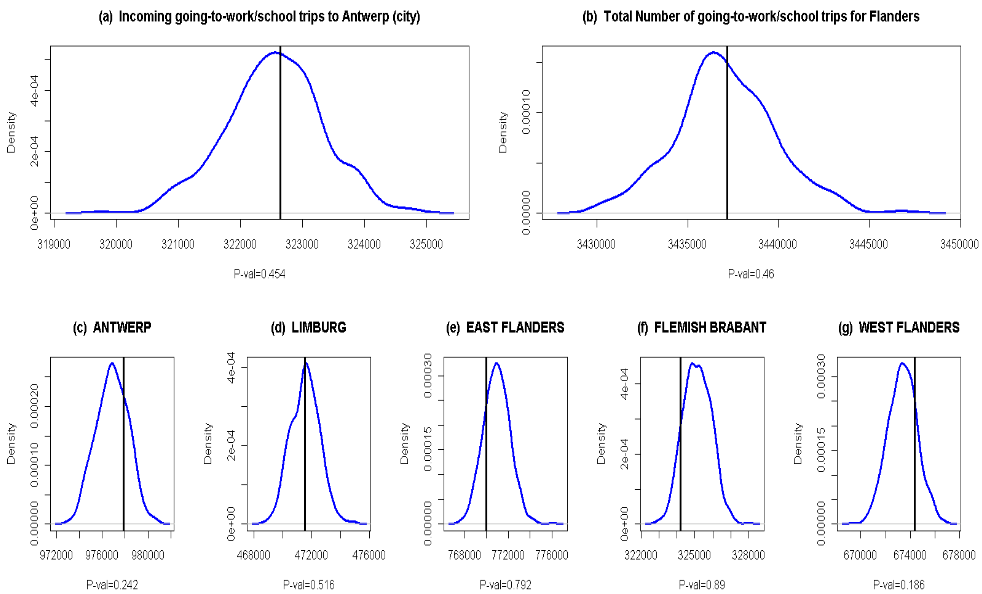

Figure 3.3 Kernel estimates of going-to-work/school trip predictive distributions for incoming trips to Antwerp (a), for total number of trips in Flanders (b) and for internal trips within each of the five Flemish provinces; Antwerp (c), Limburg (d), East Flanders (e), Flemish Brabant (f) and West Flanders (g). The vertical lines correspond to observed quantities, the p-values to the probabilities of exceeding the observed quantities.

[27] For a reminder on interpretation of Bayesian p-values, see section 2.4.2.

Similar predictive distributions can be derived for any case of specific OD flows that might be of particular interest. In general, predictive distributions provide all the necessary information concerning short-term future variability. In Table 3.8, some of the information provided by the distributions in Figure 3.3 is summarized in the form of predictive means and specific percentile points.

Type of trips	**Observed**	**Predictions**					
		Mean	**Min**	**2.5%**	**Median**	**97.5%**	**Max**
Incoming-trips Antwerp (municipality)	322,644	322,549	319,764	320,968	322,563	323,985	324,866
Total-trips for Flanders	3,437,168	3,437,092	3,429,849	3,431,615	3,436,970	3,442,768	3,447,193
Internal-trips Antwerp (province)	977,891	976,876	972,975	974,225	976,884	979,476	980,798
Internal-trips Limburg (province)	471,559	471,571	468,657	469,737	471,588	473,457	475,013
Internal-trips East Flanders (province)	769,973	770,943	767,470	768,663	770,928	773,259	776,077
Internal-trips Flemish Brabant (province)	324,181	325,132	322,840	323,705	325,135	326,722	328,002
Internal-trips West Flanders (province)	674,388	673,395	669,382	671,122	673,391	675,739	676,835

TABLE 3.8. Observed quantities, predicted means and predicted percentile points for incoming trips to the municipality of Antwerp, total number of trips in Flanders and internal trips for each of the five Flemish provinces.

Predictive effects, as those presented in Table 3.8, may be examined under different assumptions; one might choose to infer based on conservative summaries such as the predictive mean or median, or might be interested in examining the effect of more extreme summaries such as the 97.5th percentile or the maximum value. In cases of extreme predictions with p-values close to 0.01 or 0.99, one might choose where to place more trust on; the predictions of the model, the observed values or perhaps test the effects of both, depending on the specific case. These alternative options reduce overall uncertainty and may serve as predictive scenarios for transportation policy-makers, e.g. in decisions concerning infrastructure expansion.

3.2.6 The need for random effects

It is worth clarifying why predictive inference is based on the hierarchical Poisson-gamma structure and not on the marginal negative binomial structure. In Figure 3.4 the predictive distributions for incoming trips to the municipality of Antwerp, total travel-demand in Flanders and internal trips within the five

Flemish provinces resulting from the marginal negative binomial model are presented.

The predictions of Figure 3.4 can be contrasted to the predictions of Figure 3.3 from the corresponding hierarchical structure which includes random effects. It is immediately obvious that predictions from the marginal negative binomial structure are neither accurate nor realistic. For instance, one simply has to notice the range of x-axes in the graphs of Figure 3.4 where the magnitude of differences between the minimum and the maximum ranges from hundreds of thousands of trips in Figure 3.4(a) to millions of trips in Figures 3.4(b-g). On the other hand, the predictions from the hierarchical PG model which take into account the random effect of each individual OD pair result in the predictive distributions of Figure 3.3 which have reasonable range and when compared to the observed quantities yield acceptable p-values.

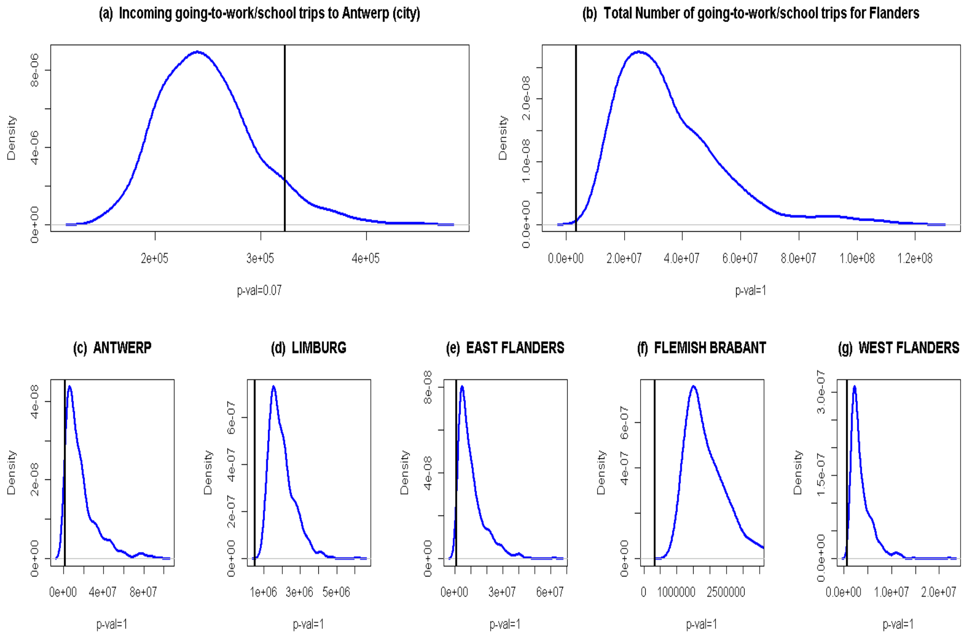

Figure 3.4 Kernel estimates of going-to-work/school trip predictive distributions from the marginal negative binomial model for incoming trips to Antwerp (a), total number of trips in Flanders (b) and internal trips within each of the five Flemish provinces; Antwerp (c), Limburg (d), East Flanders (e), Flemish Brabant (f) and West Flanders (g). The vertical lines correspond to observed quantities, the p-values to the probabilities of exceeding the observed quantities.

Regardless whether these random components are considered as random effects or intercepts or even errors, the point is that they capture “unobserved” effects which relate to the extreme heterogeneity of OD flows. One might argue that it would be much better to capture these unobserved effects by adding appropriate observed explanatory variables. This would indeed be preferable, nevertheless in current experience this has been found to be a very difficult task. Many other explanatory variables have been considered during the course of this research besides the 24 variables presented here. For instance, geographical coordinates, income levels, number of elementary/secondary schools and student populations among others. Nevertheless, the resulting gains in marginal fit were always small and not significant[28].

Needless to say that by continuously adding explanatory variables high correlations among them eventually appear, since additional socio-economic and infrastructure explanatory variables essentially reflect the same things, e.g. economic welfare, growth and so forth. Simply stated, large scale OD matrices are very complex datasets to model accurately without some kind of random component accounting for heterogeneity[29].

3.3 Extending the modeling framework: Poisson mixture models

Previously, the negative binomial model was compared to the simple Poisson model and was found to be clearly a better option due to the extreme over-dispersion. As briefly discussed the negative binomial model is a marginal model which is derived by a hierarchical Poisson model. In this section we present the general framework of hierarchical Poisson models or Poisson mixture models and also extend model comparison. The models presented here are applied on the second OD matrix and include all 25 explanatory variables presented in Table 3.3.

With Poisson mixture models it is assumed again that the OD flows y_i are i.i.d. Poisson realizations only this time the rate of the Poisson distribution is $\lambda_i = \mu_i u_i$ for $i = 1, 2, \ldots, n$. The rate λ_i is split in two parts; μ_i is the part which is

[28] As we will see in the relevant discussion of section 5.2.1, even the variable from the radiation model (Simini et al., 2012) does not lead to significant improvements.

[29] The results and conclusions in West (1994) seem to be in agreement with this statement.

related to the vector of $p+1$ unknown parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ and the set of explanatory variables $\mathbf{x}_i = (1, x_{1i}, \ldots, x_{pi})^T$ through the log-link function $\log \mu_i = \boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_i$, and u_i is a random component – interpreted as a multiplicative random effect – which is attributed with a density $g_1(u_i)$. The Poisson mixture modeling formulation is summarized as follows

$$
\begin{aligned}
& y_i \sim Pois(\lambda_i), \text{ with } \lambda_i = \mu_i u_i \text{ and} \\
& \mu_i = e^{\boldsymbol{\beta}^T \mathbf{x}_i}, \\
& u_i \sim g_1(u_i) \text{ and } E(u_i) = 1.
\end{aligned}
$$

The density g_1 is known as the *mixing* density and can be continuous, discrete or even a finite-step distribution. The assumption that the expected value of the random component u_i equals 1 ensures *scale*-identifiability. Poisson mixture models are employed as overdispersed alternatives to the simple Poisson model since the random component of the Poisson rate accounts for heterogeneity within the population. If the mixing density is degenerate, i.e. it assigns positive probability in only one point, then the simple Poisson model arises which assumes equality of mean and variance. A general framework for overdispersion models is provided by Hinde and Demétrio (1998). Alternatively, from a generalized linear mixed model (GLMM) perspective the above model can be expressed as

$$
\begin{aligned}
& y_i \sim Pois(\lambda_i), \text{ with } \log\lambda_i = \boldsymbol{\beta}^T \mathbf{x}_i + \varepsilon_i, \\
& \varepsilon_i \sim g_2(\varepsilon_i) \text{ and } E(\varepsilon_i) = 0.
\end{aligned}
$$

where ε_i is an additive random error term, namely an observation random effect or random intercept as it is most commonly known. Here, the constraint on the expected value ensures *location*-identifiability. The Poisson likelihood is the *conditional* likelihood given the unobserved random effect vector $\mathbf{u} = (u_1, u_2, \ldots, u_n)^T$. Integration with respect to $\mathbf{u}$ leads to the marginal sampling likelihood, i.e. $p(\mathbf{y} \mid \boldsymbol{\beta}) = \int p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{u}) g_1(\mathbf{u}) d\mathbf{u}$. The Poisson mixture and GLMM formulations are equivalent, however the resulting intercept estimates and the interpretations of marginal means $E(\mathbf{y} \mid \boldsymbol{\beta})$ are different due to the identifiability constraints (Lee and Nelder, 2004). Frequentist inference usually focuses on the marginal structure of the model under ML, restricted-maximum-likelihood (REML), quasi-likelihood (QL) and pseudo-likelihood (PL) estimation procedures. Estimation of random effects, when needed, is usually based on empirical Bayes estimates; the posterior expectation of **u**, for instance[30].

Different assumptions on the mixing density result to different marginal sampling distributions. When the mixing density g_1 is a gamma distribution we have the Poisson-gamma (PG) model which is the most frequently used Poisson mixture model due to the property that the resulting marginal likelihood is a negative binomial distribution. Properties and estimation procedures for negative binomial regression can be found in Lawless (1987). Negative binomial modeling has been investigated in many studies as, for instance, in Thall (1988), Xue and Deddens (1992), McNeney and Petkau (1994), Dean (1994) and Van de Ven and Weber (1995). The PG model is also included in the family of hierarchical generalized linear models (HGLM's) introduced by Lee and Nelder (1996) who provide ML estimates for regression parameters as well as random effects based on a hierarchical likelihood (h-likelihood) and also examine various theoretical aspects of the model (see also the associated discussion).

The Poisson-lognormal (PLN) model arises when g_1 is a lognormal distribution. The resulting marginal distribution of this model, known simply as Poisson-lognormal (Shaban, 1988), does not have a closed form expression and thus numerical integration is needed for marginal estimation. Nevertheless, the PLN model is regularly used in practice due to its distinct historical development as a GLMM for count data based on the assumption that g_2 is a normal distribution (Hinde, 1982; Breslow 1984). The density g_1 is lognormal, consequently. Estimation of the model through Gaussian quadrature and the expectation-maximization (EM) algorithm is handled in Anderson and Hinde (1988) and Aitkin (1996).

An inverse Gaussian density for g_1 results in the Poisson-inverse Gaussian (PIG) model which leads to a Poisson-inverse Gaussian marginal density. This distribution, unlike the Poisson-lognormal case, does have a closed form

[30] A convenient property in Poisson mixtures which has been to a large degree overlooked is that the posterior expectation always has a closed-form expression regardless of the choice of mixing density (see Sapatinas, 1995).

expression, but it involves a modified Bessel function which poses some computational difficulties in comparison to the negative binomial distribution. The PIG model was first presented by Holla (1967). The first ML estimates for simple cases without covariates were provided by Shaban (1981). ML estimates under a different parameterization of asymptotically uncorrelated parameters were presented by Stein et al. (1987) and Stein and Juritz (1988) for cases without and with covariates, respectively. Finally, Dean et al. (1989) provided ML and QL estimates for another re-parameterization which is not asymptotically orthogonal but has a convenient interpretation in terms of multiplicative random effects. Despite the fact that the theoretical properties of this model have been thoroughly explored, the PIG model has started only recently to be considered as a competing alternative to the PG and PLN models.

A first consideration of all three models is presented in Chen and Ahn (1996) who compare the bias and efficiency of semi-parametric quasi estimates in simulation experiments. Later, Karlis (2001) provided a generally applicable EM algorithm for Poisson mixtures and compared the three models on a real dataset. The algorithm of Karlis significantly relaxes the computational demands for obtaining ML estimates, particularly for the case of the PIG model. In Boucher and Denuit (2006) the performance of three models is investigated from a random-effect versus fixed-effect perspective on motor insurance claims. Finally, in Nikoloulopoulos and Karlis (2008) the models are compared with respect to distributional properties such as skewness and kurtosis under simulation experiments. This study verifies some theoretical expectations (see e.g. Willmot, 1990), namely that the PLN and PIG models allow for longer right tails and are thus more appropriate than the PG model for modeling highly positive-skewed data.

Within the Bayesian framework, Poisson mixture models have a natural interpretation as hierarchical or multilevel models, since the mixing distribution may be regarded in fact as a first-level prior distribution of which the parameters are subsequently assigned with a second-level or hyper-prior distribution. Within this context, the question between hierarchical or marginal model fitting depends on experimental design assumptions, inferential scope and practical considerations regarding computational costs in terms of simulation time. Bayesian applications of negative binomial modeling as well as hierarchical

PG and PLN modeling from a Poisson mixture perspective can be found in the book of Ntzoufras (2009). Hierarchical treatment of PG and PLN models is also available in the books of Lawson et al. (2003) for disease mapping (see also Mollié, 1996) and of Gelman and Hill (2006) which deals with hierarchical/multilevel modeling in general. Some applications of the two models are also presented in Congdon (2001).

The PIG model is not included, as yet, in the relative Bayesian literature which as we will see has no particular justification. Regarding probability calculations from the marginal Poisson-inverse Gaussian distribution, computer routines are readily available, e.g. in Stasinopoulos and Rigby (2007). Furthermore and perhaps more important, the inverse Gaussian distribution is a special case of the three-parameter generalized inverse Gaussian distribution which is generally conjugate to the family of exponential distributions (Jorgensen, 1982).

In the remainder of this section, Bayesian forms of the three models are presented for hierarchical and marginal modeling with non-informative prior distributions. Given the size of the OD dataset, the parameters of scientific interest – the regression and dispersion parameters – are estimated from the marginal models. Nevertheless predictive inference is based on the hierarchical models. The Bayesian framework is particularly useful for large datasets and designs of observational random effects. If the prior of **u** depends on a hyper-parameter ω, then from properties of conditional probabilities the joint posterior is $p(\boldsymbol{\beta}, \mathbf{u}, \omega \mid \mathbf{y}) = p(\mathbf{u} \mid \boldsymbol{\beta}, \omega, \mathbf{y}) p(\boldsymbol{\beta}, \omega \mid \mathbf{y})$. Thus, for hierarchical inference one can use marginal models for sampling from $p(\boldsymbol{\beta}, \omega \mid \mathbf{y})$ and generate **u** subsequently from $p(\mathbf{u} \mid \boldsymbol{\beta}, \omega, \mathbf{y})$. As illustrated previously, this is straightforward for the PG model and as shown next, it is also straightforward for the PIG model due to a conjugate prior. Empirical evidence for the equivalence of the estimates can be found in the interesting research of Fahrheim and Osuna (2003) where estimates from a marginal negative binomial and from the corresponding hierarchical PG are presented. The authors conclude that the results from the two models are indistinguishable.

3.3.1 The PG model

For the PG model the following likelihood and prior assumptions are made;

$y_i \mid \boldsymbol{\beta}, u_i \sim Pois(e^{\boldsymbol{\beta}^T \mathbf{x}_i} u_i)$,
$\boldsymbol{\beta} \sim \mathbf{N}_{p+1}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$ with $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = n(\mathbf{X}^T\mathbf{X})^{-1}\sigma^2_{\boldsymbol{\beta}}$ and $\sigma^2_{\boldsymbol{\beta}} = 10^3$,
$u_i \sim Gamma(\theta, \theta)$ and
$\theta \sim Gamma(a_\theta, a_\theta)$ with $a_\theta = 10^{-3}$.

For the multivariate normal prior on the regression vector a g-prior similar structure (Zellner, 1986) is adopted which as mentioned in section 3.2.2 is analogue to one of the benchmark priors discussed in Fernández et al. (2001) for normal linear models, but for $\sigma^2_{\boldsymbol{\beta}}$ fixed. The same weakly informative multivariate prior is also adopted for the PLN and PIG models. The gamma prior for u_i is defined in terms of shape and rate parameters which both equal θ, so that $E(u_i) = 1$ and $\mathrm{var}(u_i) = \theta^{-1}$. The gamma hyper-prior for dispersion parameter θ with both shape and rate equal to 0.001 is a diffuse prior as in Ntzoufras (2009). The joint posterior distribution of all parameters is $p(\boldsymbol{\beta}, \mathbf{u}, \theta \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{u}) p(\boldsymbol{\beta}) p(\mathbf{u} \mid \theta) p(\theta)$, that is

$$p(\boldsymbol{\beta}, \mathbf{u}, \theta \mid \mathbf{y}) \propto \prod_{i=1}^{n} \left[(e^{\boldsymbol{\beta}^T \mathbf{x}_i} u_i)^{y_i} e^{-e^{\boldsymbol{\beta}^T \mathbf{x}_i} u_i} \right] \times e^{-\frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta}} \times \theta^{n\theta} \Gamma(\theta)^{-n} \prod_{i=1}^{n} \left[u_i^{\theta-1} e^{-\theta u_i} \right] \times \\ \times \theta^{a_\theta - 1} e^{-a_\theta \theta} \quad . \quad (3.7)$$

The only conditional distribution with a known form, based on expression 3.7, is the one of the random effect vector which is namely a gamma distribution, i.e. $u_i \mid \boldsymbol{\beta}, \theta, y_i \sim Gamma(y_i + \theta, e^{\boldsymbol{\beta}^T \mathbf{x}_i} + \theta)$ (Gelman and Hill, 2006). This implies that fitting the hierarchical model through MCMC would require a Metropolis-within-Gibbs type of algorithm with Metropolis steps for the joint conditional of $\boldsymbol{\beta}, \theta \mid \mathbf{u}, \mathbf{y}$ or for the conditionals of $\boldsymbol{\beta} \mid \mathbf{u}, \mathbf{y}$ and $\theta \mid \mathbf{u}, \mathbf{y}$. Alternatively, *adaptive-rejection* sampling can also be used.

Nevertheless, as discussed in section 3.2.2 integration over $\mathbf{u}$ leads to a negative binomial marginal likelihood, i.e. $y_i \mid \boldsymbol{\beta}, \theta \sim NB(e^{\boldsymbol{\beta}^T \mathbf{x}_i}, \theta)$ where the marginal mean and variance are given by $E(\mathbf{y} \mid \boldsymbol{\beta}) = e^{\mathbf{x}^T \boldsymbol{\beta}}$ and $\mathrm{var}(\mathbf{y} \mid \boldsymbol{\beta}) = e^{\mathbf{x}^T \boldsymbol{\beta}} + (e^{\mathbf{x}^T \boldsymbol{\beta}})^2 \theta^{-1}$, respectively, with the variance being a quadratic function of the mean. The un-normalized posterior distribution is shown in expression 3.6 and does not result in any closed form expressions. Nevertheless, generating samples from the joint posterior of $\boldsymbol{\beta},\theta \mid \mathbf{y}$ is feasible through MH simulation.

3.3.2 The PLN model

The assumptions for the PLN model are the following;

$$\begin{aligned}
&y_i \mid \boldsymbol{\beta}, u_i \sim Pois(e^{\boldsymbol{\beta}^T \mathbf{x}_i} u_i),\\
&\boldsymbol{\beta} \sim \mathbf{N}_{p+1}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \text{ with } \boldsymbol{\Sigma}_{\boldsymbol{\beta}} = n(\mathbf{X}^T\mathbf{X})^{-1}\sigma^2_{\boldsymbol{\beta}} \text{ and } \sigma^2_{\boldsymbol{\beta}} = 10^3,\\
&u_i \sim LN(-\sigma^2 / 2, \sigma^2) \text{ and}\\
&\sigma^2 \sim InvGamma(a_{\sigma^2}, a_{\sigma^2}) \text{ with } a_{\sigma^2} = 10^{-3}
\end{aligned}$$

The prior distribution of u_i has location parameter equal to $-\sigma^2/2$ and scale σ^2, so that $E(u_i)=1$ and $\mathrm{var}(u_i) = (e^{\sigma^2}-1)$ [31]. The inverse gamma hyper-prior for σ^2 is the common option for this model; for $a_{\sigma^2} = 10^{-3}$ the distribution of σ^{-2} is a diffuse gamma with a mean equal to 1 and a variance equal to 1000 (Ntzoufras, 2009). The posterior is $p(\boldsymbol{\beta},\mathbf{u},\sigma^2 \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta},\mathbf{u})p(\boldsymbol{\beta})p(\mathbf{u} \mid \sigma^2)p(\sigma^2)$ which leads to

$$\begin{aligned}
p(\boldsymbol{\beta},\mathbf{u},\sigma^2 \mid \mathbf{y}) \propto & \prod_{i=1}^{n}\left[(e^{\boldsymbol{\beta}^T\mathbf{x}_i}u_i)^{y_i} e^{-e^{\boldsymbol{\beta}^T\mathbf{x}_i}u_i}\right] \times e^{-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}} \times (\sigma^2)^{-n/2}\prod_{i=1}^{n}\left[u_i^{-1}e^{-\frac{(\log u_i+\sigma^2/2)^2}{2\sigma^2}}\right]\times \\
& \times (\sigma^2)^{-a_{\sigma^2}-1}e^{-a_{\sigma^2}/\sigma^2}
\end{aligned} \quad (3.8)$$

In this case none of the full conditional distributions are of known form. MCMC sampling for the hierarchical PLN model is in general more convenient in the respective GLMM form where the full conditional distribution of σ^2 is known, namely $\sigma^2 \mid \mathbf{u},\mathbf{y} \sim InvGamma(a_{\sigma^2}+n/2, a+\sum_{i=1}^{n}(\log u_i)^2/2)$. Sampling from the conditionals of $\boldsymbol{\beta}$ and $\mathbf{u}$ is possible with Metropolis steps (Browne, 2003) or rejection-sampling (Zeger and Karim, 1991).

[31] Interested readers should refer first to Lee and Nelder (2004) for this parameterization. Most Bayesian applications are focusing on the additive random effect structure which has a different parameterization with a 0 mean on non-logarithmic scale. Nevertheless, the additive structure is not immediately comparable to the multiplicative structure.

For the case of the PLN model the marginal likelihood $p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2)$ is not known analytically, nevertheless the mean and variance of the PLN distribution are available and given by $E(\mathbf{y} \mid \boldsymbol{\beta}) = e^{\mathbf{X}^T\boldsymbol{\beta}}$ and $\mathrm{var}(\mathbf{y} \mid \boldsymbol{\beta}) = e^{\mathbf{X}^T\boldsymbol{\beta}} + (e^{\mathbf{X}^T\boldsymbol{\beta}})^2(e^{\sigma^2} - 1)$. Once again the variance is a quadratic function of the mean. The joint posterior density is $p(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2)p(\boldsymbol{\beta})p(\sigma^2)$, specifically

$$p(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}) \propto \prod_{i=1}^{n} \left[\int (e^{\boldsymbol{\beta}^T \mathbf{x}_i} u_i)^{y_i} e^{-e^{\boldsymbol{\beta}^T \mathbf{x}_i} u_i} (\sigma^2)^{-1/2} u_i^{-1} e^{-\frac{(\log u_i + \sigma^2/2)^2}{2\sigma^2}} du_i \right] \times e^{-\frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta}} \times \\ \times (\sigma^2)^{-a_{\sigma^2} - 1} e^{-a_{\sigma^2}/\sigma^2} \quad . \; (3.9)$$

MH simulation is employed in order to sample from the joint posterior density of $\boldsymbol{\beta}$ and σ^2. The integral in expression 3.9 can be evaluated through numerical integration, e.g. with Gauss-Hermite quadrature estimation which is also frequently employed in frequentist practice for marginal estimation (see for example Hedeker and Gibbons, 1994; Rabe-Hesketh et al., 2001). Another alternative which is investigated here is MC integration from the lognormal prior of the random effect vector $\mathbf{u}$ within the Metropolis kernel. That is, for a given MH iteration t and draws $\boldsymbol{\beta}^{(t)}$, $\sigma^{2(t)}$, the integral above can be evaluated by generating first L MC draws $\left\{u_i^{(t,l)}, l = 1, 2, \ldots, L\right\}$ from $u_i^{(t,L)} \sim LN(-\sigma^{2(t)}/2, \sigma^{2(t)})$ and then by calculating the marginal likelihood probability as $p(y_i \mid \boldsymbol{\beta}^{(t)}, \sigma^{2(t)}) = L^{-1}\sum_{l=1}^{L} p(y_i \mid \boldsymbol{\beta}^{(t)}, u_i^{(t,l)})$.

The disadvantage of fitting the marginal PLN model is that draws from the posterior distribution of $\mathbf{u}$ cannot be generated straightforwardly and therefore exact inference from the hierarchical structure is not easy. It is possible, for instance, to calculate posterior moments r of $\mathbf{u}$ from the general formula provided by Sapatinas (1995) as

$$E(u_i^r \mid \boldsymbol{\beta}, \sigma^2, y_i) = \frac{(y_i + r)!}{e^{\boldsymbol{\beta}^T \mathbf{x}_i} y_i!} \frac{p(y_i + r \mid \boldsymbol{\beta}, \sigma^2)}{p(y_i \mid \boldsymbol{\beta}, \sigma^2)} .$$

Of course, the formula requires once again calculation of probabilities from the marginal PLN distribution which leads to additional numerical or MC integration.

3.3.3 The PIG model

For the PIG hierarchical formulation the following assumptions are adopted;

$$
\begin{aligned}
& y_i \mid \boldsymbol{\beta}, u_i \sim Pois(e^{\boldsymbol{\beta}^T \mathbf{x}_i} u_i), \\
& \boldsymbol{\beta} \sim \mathbf{N}_{p+1}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \text{ with } \boldsymbol{\Sigma}_{\boldsymbol{\beta}} = n(\mathbf{X}^T\mathbf{X})^{-1}\sigma_{\boldsymbol{\beta}}^2 \text{ and } \sigma_{\boldsymbol{\beta}}^2 = 10^3, \\
& u_i \sim IG(1, \zeta) \text{ and} \\
& \zeta \sim Gamma(a_\zeta, a_\zeta) \text{ with } a_\zeta = 10^{-3}.
\end{aligned}
$$

For the inverse Gaussian prior the initial parameterization of Holla (1967) is followed, denoted as $IG(\mu, \zeta)$ with mean μ and shape ζ. Specifically,

$$
p(u_i \mid \mu, \zeta) = \left(\frac{\zeta}{2\pi u_i^3} \right)^{1/2} e^{-\frac{\zeta (u_i - \mu)^2}{2\mu^2 u_i}}.
$$

For $\mu = 1$ we have that a-priori $E(u_i) = 1$ and $\text{var}(u_i) = \zeta^{-1}$. The inverse Gaussian distribution is a special case of the three-parameter generalized inverse Gaussian (GIG) distribution which is investigated in detail in Jorgensen (1982). The p.d.f. of a $GIG(\lambda, \psi, \chi)$ distribution with parameters $\lambda \in \mathbb{R}$, $\psi, \chi > 0$ is given by

$$
f(x) = \frac{(\psi / \chi)^{1/2}}{2K_\lambda(\sqrt{\psi\chi})} x^{\lambda-1} e^{-\frac{1}{2}(\psi x + \chi x^{-1})},
$$

where K_λ is the modified Bessel function of the third kind with order λ. The inverse Gaussian distribution arises for $\lambda = -1/2$. Interestingly, the gamma distribution is also a special case of the GIG distribution for $\chi = 0$. For shape parameter ζ a gamma hyper-prior is adopted, similarly to the PG model. The posterior distribution is $p(\boldsymbol{\beta}, \mathbf{u}, \zeta \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{u}) p(\boldsymbol{\beta}) p(\mathbf{u} \mid \zeta) p(\zeta)$, resulting in

$$
\begin{aligned}
p(\boldsymbol{\beta}, \mathbf{u}, \zeta \mid \mathbf{y}) \propto & \prod_{i=1}^{n} \left[(e^{\boldsymbol{\beta}^T \mathbf{x}_i} u_i)^{y_i} e^{-e^{\boldsymbol{\beta}^T \mathbf{x}_i} u_i} \right] \times e^{-\frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta}} \times \zeta^{n/2} \prod_{i=1}^{n} \left[u_i^{-3/2} e^{-\frac{\zeta(u_i - 2 - u_i^{-1})}{2}} \right] \times \\
& \times \zeta^{a_\zeta - 1} e^{-a_\zeta \zeta}
\end{aligned}
. \quad (3.10)
$$

It can be easily shown that the conditional of u_i is $GIG(y_i - 1/2, 2e^{\boldsymbol{\beta}^T \mathbf{x}_i} + \zeta, \zeta)$ and that the conditional of ζ is $Gamma(a_\zeta + n/2, a_\zeta + \sum_{i=1}^{n} (u_i - 1)^2 / 2u_i)$, with the parameterization of the GIG distribution as presented above. Athreya (1986) was the first to notice the specific conjugate relationship between the inverse

Gaussian and Poisson distribution. This relationship is also noted in Karlis (2001) under a different parameterization of the inverse Gaussian. Thus, in comparison to the PG and PLN models, the hierarchical PIG model is actually the simplest in terms of MCMC, since all that is needed is a MH or rejection-sampling algorithm for the conditional of $\boldsymbol{\beta}$. Regarding simulation from the GIG distribution, random generators are readily available (see for instance Atkinson, 1982; Dagpunar, 1988).

Marginally, we have that $y_i \mid \boldsymbol{\beta}, \zeta \sim PIG(e^{\boldsymbol{\beta}^T \mathbf{x}_i}, \zeta)$ for $i = 1, 2, \ldots, n$ with p.d.f. given by

$$p(y_i \mid \boldsymbol{\beta}, \zeta) = K_{y_i - 1/2}\left(\sqrt{2\zeta\left(1 + \frac{\zeta}{2e^{\boldsymbol{\beta}^T \mathbf{x}_i}}\right)}\right)\left(\frac{2\zeta}{\pi}\right)^{1/2} \frac{e^{\zeta / e^{\boldsymbol{\beta}^T \mathbf{x}_i}}}{y_i!}\left(2\left(1 + \frac{\zeta}{2e^{\boldsymbol{\beta}^T \mathbf{x}_i}}\right)\zeta^{-1}\right)^{1/2(y_i - 1/2)}.$$

In this case, the marginal mean and variance are $E(\mathbf{y} \mid \boldsymbol{\beta}) = e^{\mathbf{X}^T \boldsymbol{\beta}}$ and $\mathrm{var}(\mathbf{y} \mid \boldsymbol{\beta}) = e^{\mathbf{X}^T \boldsymbol{\beta}} + (e^{\mathbf{X}^T \boldsymbol{\beta}})^3 \zeta^{-1}$, which are very similar to those of the negative binomial model only that the variance is now a cubic function of the mean. The posterior distribution is $p(\boldsymbol{\beta}, \zeta \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \zeta) p(\boldsymbol{\beta}) p(\zeta)$, i.e.

$$\begin{aligned} p(\boldsymbol{\beta}, \zeta \mid \mathbf{y}) \propto \prod_{i=1}^{n}\left[K_{y_i - 1/2}\left(\sqrt{2\zeta\left(1 + \frac{\zeta}{2e^{\boldsymbol{\beta}^T \mathbf{x}_i}}\right)}\right) e^{\zeta / e^{\boldsymbol{\beta}^T \mathbf{x}_i}}\left(2\left(1 + \frac{\zeta}{2e^{\boldsymbol{\beta}^T \mathbf{x}_i}}\right)\zeta^{-1}\right)^{1/2(y_i - 1/2)}\right] \times \\ \times \zeta^{n/2} \times e^{-\frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta}} \times \zeta^{a_\zeta - 1} e^{-a_\zeta \zeta} \end{aligned} \quad (3.11)$$

In this case, a sample from the posterior of $\boldsymbol{\beta}$ and ζ can be obtained through MH simulation on the joint un-normalized posterior in expression 3.11. As with the PG model, when posterior draws of $\boldsymbol{\beta}$ and ζ are available predictive inference from the hierarchical structure is possible by generating from the conditional distribution of $\mathbf{u}$.

3.3.4 Implementation and performance of MH simulation

In order to bypass the daunting task of sampling and storing 94864 random effects for each MCMC iteration, MH simulation is implemented once again on the marginal structures. Although sampling $\mathbf{u}$ is relatively straightforward for the hierarchical PG and PIG models, memory requirements pose significant limitations if terms of storing. MCMC for the marginal PG and PIG structures is far more efficient with $\boldsymbol{\beta}$, θ and ζ being easy to sample and with $\mathbf{u}$ generated subsequently. The PLN model is more problematic since an additional Metropolis step or rejection sampling is required within MCMC for the hierarchical structure which is an obvious burden for 94864 random effects. On the other hand simulation for the marginal PLN structure requires numerical or MC integration within MCMC and – in addition – vector $\mathbf{u}$ is not easy to sample.

As before, an independence chain algorithm is employed where the location and scale of the proposals are fixed to the corresponding ML estimates. For regression vector $\boldsymbol{\beta}$ a multivariate normal proposal is used, i.e. $q(\boldsymbol{\beta}) = \mathbf{N}_{p+1}(\boldsymbol{\beta}^{ML}, \mathbf{V}_{\boldsymbol{\beta}}^{ML})$ with $\boldsymbol{\beta}^{ML}$, $\mathbf{V}_{\boldsymbol{\beta}}^{ML}$ being the respective ML estimate of $\boldsymbol{\beta}$ and the estimated variance-covariance matrix of $\boldsymbol{\beta}^{ML}$ for each model. For parameters θ, σ^2 and ζ the following proposals are adopted; $q(\theta) \sim Gamma(a_{PG}, b_{PG})$, $q(\sigma^2) \sim Gamma(a_{PLN}, b_{PLN})$ and $q(\zeta) \sim Gamma(a_{PIG}, b_{PIG})$ with proposal parameters set to satisfy the conditions $a_{PG} / b_{PG} = \theta^{ML}$, $a_{PG} / b_{PG}^2 = \mathrm{var}(\theta^{ML})$, $a_{PLN} / b_{PLN} = \sigma^{2ML}$, $a_{PLN} / b_{PLN}^2 = \mathrm{var}(\sigma^{2ML})$, $a_{PIG} / b_{PIG} = \zeta^{ML}$ and $a_{PIG} / b_{PIG}^2 = \mathrm{var}(\zeta^{ML})$. Regarding probability calculations from the PLN distribution both techniques of numerical and MC integration were investigated. It was found that the MC sample L should be preferably 2,000 in order to obtain stable results which were similar to the results from numerical integration, while numerical integration was already two-times faster than MC integration with a sample of 200. Therefore, numerical integration was preferred.

One single MH chain of 21,000 iterations was executed for each model. Random samples of size 5,000 were initially generated from the aforementioned proposals and the 90th percentile points were used as starting values. The simulations for the PG and PIG models required approximately 1 and 2 hours, respectively, whereas the PLN model required significant time due to numerical integration, almost 3.6 days[32]. The resulting acceptance ratios were 72% for the PG model, 67% for the PLN model and 33% for the PIG model[33]. The first thousand iterations were discarded as the burn-in part of the chains.

[32] All simulations were executed on a 64bit Windows Server 2003 R2 with 32GB of RAM.

[33] For the PIG model we were not able to obtain an estimate for $V_{\boldsymbol{\beta}}^{ML}$ and therefore the variance-covariance matrix of the PG model was used instead.

Convergence checks were based on the methods of Heidelberger and Welch (1983), Raftery and Lewis (1992) and Geweke (1992). The diagnostic of Raftery and Lewis indicated autocorrelation problems which were more serious for the cases of the PLN and PIG models. In order to weaken the strong autocorrelations 4,000 draws were selected randomly without replacement in the PLN and PIG models, while the sample of the PG model was thinned by an interval of 5. The final samples of 4,000 posterior draws passed all diagnostics, with lag 1 autocorrelations not exceeding in absolute terms the value of 0.05. The MH algorithms and results from convergence diagnostics are presented in Appendix A. All simulations and related calculations were implemented in the R programming environment v. 2.8.2. A variety of R libraries was used for ML estimation, random sampling, density calculations and MCMC convergence checks. A list of the R routines and corresponding libraries can be found in Appendix C.

3.3.5 Posterior inference from Poisson mixture models

Posterior summaries are presented in Table 3.9 for means and standard deviations and in Table 3.10 for 95% credible intervals. Based on the 95% intervals all regression parameters in the three models are statistically significant. A first significant observation is that posterior means are more similar for the PLN and PIG models, for instance parameters β_0, β_6, β_9, β_{10}, β_{14} and β_{18} of the PG model are substantially different from the corresponding estimates of the other two models, especially the intercept estimate. On the other hand parameters β_{11}, β_{12} and β_{20} differ in general across models. Standard deviations are in overall slightly lower in the PG model.

Several remarks can be made based on the signs of the posterior estimates regarding interpretation of parameters independently and also in conjunction to traditional trip-production and trip-attraction modeling employed within the trip-generation step. The parameters of dummy variables β_1 to β_5 are positive except of the last parameter for intra-zonal municipality trips. The small posterior means of these parameters are due to the fact that the corresponding dummy variables are coded 0/100. The positive effects of β_1 to β_4 are to be expected, since the OD flows are generally larger in diagonal blocks of cells of the OD matrix corresponding to intra-zonal flows for the various administrative levels. The negative sign of β_5 is not expected but it might be explained as simply counterbalancing the absence of the strong negative effect of distance which is set almost equal to zero for intra-zonal municipality trips. Parameter β_6 is positive and leads to the consistent interpretation that destination zones which support a college or a university are more likely to attract trips than zones without a college/university.

Parameters β_7 to β_{10} quantify the influence of the total number of surrounding municipalities on the four levels of cantons, districts, arrondissements and provinces, respectively. This effect is in general not straightforward to predict. On one hand it can be argued that the more municipalities there are the more likely it is to have an increase of trips due to a general boost of socio-economic activities, but on the other hand it is also likely to observe less trips for a given OD pair since it is probable that a significant proportion of trips will spread across the surrounding municipalities. Under this consideration, the parameter estimates are interesting and provide some insights. On the small-scale level of cantons parameter β_7 has a positive sign, whereas on the large-scale levels of districts, arrondissements and provinces – where the total number of municipalities increase and a spread-out of trips is more likely – the corresponding parameters β_8, β_9 and β_{10} are negative. This implies that the effect changes from positive to negative when exceeding a specific radius threshold of distance. Recent transportation studies discuss resembling ideas such as the *neighborhood-effect* concept investigated in Sohn and Kim (2010) where it is suggested that the attractiveness of a specific zone is affected by its neighboring municipalities depending on boundaries related to the influence of distance.

Regarding the continuous variables which come in pairs, the more general explanatory variables, namely population density (β_{13}, β_{14}), perimeter length (β_{17}, β_{18}) and yearly kilometers-driven in highways (β_{21}, β_{22}) and provincial/municipal roads (β_{23}, β_{24}) have parameters with positive signs. The uniformly positive effects for origin and destination zones do not come as a surprise, since these four variables are expected to have a positive influence on trip-production (origin zones) as well as trip-attraction (destination zones). In contrast, the parameters of employment ratio (β_{11}, β_{12}), relative length of road network (β_{15}, β_{16}) and car ownership ratio (β_{19}, β_{20}) have opposite signs for origin and destination effects.

In transportation studies employment ratio is commonly associated with trip-attraction models (see e.g. Yao and Morikawa, 2005). In accordance, the posterior estimate of employment ratio is positive for destination zones and negative for origin zones which leads to the rational interpretation that zones with high employment ratios are more likely to attract trips rather than to generate trips. The relative length of road networks is associated with the concept of *accessibility* (see e.g. Odoki et al., 2001), a concept which is present primarily in trip-attraction studies and, to a lesser degree, in trip-production studies. The posterior mean is positive for destination zones and negative for origin zones. Consistently, this implies that zones with high levels of accessibility are more likely to attract trips than low-accessible zones. Conversely, high-accessible zones are less likely to produce trips than low-accessible zones. A possible explanation for the negative origin effect is that high levels of accessibility within a zone might encourage intra-zonal trips and reduce outgoing trips. Finally, car ownership is traditionally used as an explanatory variable with positive impact in trip-production models. In agreement, the posterior mean for car ownership is positive for origin zones, which means that zones with high car ownership ratios also have high trip-production rates. The estimate is negative for destination zones implying that high car ownership ratios are negatively correlated with trip-attraction. The negative destination effect may be attributed to congestion issues.

Distance with parameter β_{25} is the final variable. Distance is the key variable in gravity-type and direct-demand models, since it is directly related to the costs of the deterrence function used within the trip-distribution step (see e.g. Ortúzar and Willumsen, 2001). In the models presented here distance has a negative posterior mean which accords with the basic deterrent gravitational assumption of trip-distribution models. Furthermore, based on the posterior mean-over-standard deviation ratio distance is also the most significant explanatory variable in all three models.

Parameter	PG		PLN		PIG	
	Mean	St.dev	Mean	St.dev	Mean	St.dev
β_0 intercept	4.034	0.4101	6.124	0.4381	6.841	0.4453
β_1 dummy.province	0.005	0.0001	0.005	0.0002	0.006	0.0002
β_2 dummy.arron/ment	0.007	0.0002	0.007	0.0002	0.008	0.0002
β_3 dummy.district	0.008	0.0003	0.008	0.0003	0.009	0.0003
β_4 dummy.canton	0.008	0.0004	0.006	0.0004	0.006	0.0004
β_5 dummy.municipality	-0.082	0.0008	-0.086	0.0009	-0.084	0.0008
β_6 dummy.education	0.424	0.0189	0.535	0.0194	0.535	0.0198
β_7 munic.in.cantons	0.472	0.0187	0.461	0.0206	0.451	0.0204
β_8 munic.in.districts	-0.494	0.0236	-0.441	0.0247	-0.442	0.0254
β_9 munic.in.arron/ments	-0.087	0.0174	-0.188	0.0195	-0.211	0.0194
β_{10} munic.in.provinces	-0.493	0.0720	-0.740	0.0752	-0.781	0.0759
β_{11} employment.ratio.o	-1.062	0.0741	-0.483	0.0739	-0.241	0.0751
β_{12} employment.ratio.d	0.328	0.0692	0.491	0.0761	0.611	0.0764
β_{13} population.density.o	0.505	0.0139	0.499	0.0150	0.499	0.0145
β_{14} popolation.density.d	0.577	0.0147	0.627	0.0158	0.632	0.0159
β_{15} road.length.o	-0.315	0.0222	-0.319	0.0230	-0.334	0.0217
β_{16} road.length.d	0.280	0.0219	0.267	0.0234	0.264	0.0235
β_{17} perimeter.length.o	1.254	0.0233	1.289	0.0245	1.282	0.0241
β_{18} perimeter.length.d	0.429	0.0233	0.500	0.0255	0.512	0.0261
β_{19} car.ownership.ratio.o	3.454	0.1524	3.420	0.1629	3.520	0.1634
β_{20} car.ownership.ratio.d	-1.464	0.1520	-1.253	0.1653	-1.075	0.1656
β_{21} highway.traffic.o	0.011	0.0017	0.011	0.0018	0.010	0.0018
β_{22} highway.traffic.d	0.052	0.0017	0.050	0.0018	0.050	0.0019
β_{23} prov/munic.traffic.o	0.270	0.0101	0.278	0.0107	0.276	0.0100
β_{24} prov/munic.traffic.d	0.870	0.0100	0.875	0.0114	0.869	0.0111
β_{25} distance	-2.906	0.0104	-2.984	0.0118	-2.937	0.0110
θ theta	0.965	0.0093	–		–	
σ^2 sigma square	–		1.065	0.0113	–	
ζ zeta	–		–		0.375	0.0095

TABLE 3.9. Posterior means and standard deviations from 4,000 posterior draws of the Poisson-gamma (PG), Poisson-lognormal (PLN) and Poisson-inverse Gaussian (PIG) models.

Parameter	PG	PLN	PIG
	95% C.I.	**95% C.I.**	**95% C.I.**
β_0 intercept	(3.223, 4.842)	(5.244, 6.962)	(5.937, 7.708)
β_1 dummy.province	(0.005, 0.005)	(0.005, 0.006)	(0.005, 0.006)
β_2 dummy.arron/ment	(0.007, 0.008)	(0.007, 0.008)	(0.007, 0.008)
β_3 dummy.district	(0.008, 0.009)	(0.008, 0.009)	(0.008, 0.009)
β_4 dummy.canton	(0.007, 0.009)	(0.006, 0.007)	(0.006, 0.007)
β_5 dummy.municipality	(-0.083, -0.080)	(-0.088, -0.084)	(-0.086, -0.082)
β_6 dummy.education	(0.386, 0.460)	(0.497, 0.573)	(0.497, 0.573)
β_7 munic.in.cantons	(0.436, 0.510)	(0.422, 0.501)	(0.411, 0.491)
β_8 munic.in.districts	(-0.540, -0.448)	(-0.489, -0.393)	(-0.493, -0.394)
β_9 munic.in.arron/ments	(-0.122, -0.052)	(-0.225, -0.149)	(-0.248, -0.174)
β_{10} munic.in.provinces	(-0.635, -0.355)	(-0.886, -0.588)	(-0.923, -0.634)
β_{11} employment.ratio.o	(-1.203, -0.915)	(-0.632, -0.337)	(-0.386, -0.096)
β_{12} employment.ratio.d	(0.192, 0.465)	(0.343, 0.639)	(0.457, 0.753)
β_{13} population.density.o	(0.479, 0.532)	(0.470, 0.529)	(0.472, 0.528)
β_{14} popolation.density.d	(0.548, 0.606)	(0.597, 0.658)	(0.601, 0.664)
β_{15} road.length.o	(-0.359, -0.271)	(-0.363, -0.273)	(-0.378, -0.293)
β_{16} road.length.d	(0.237, 0.324)	(0.222, 0.312)	(0.218, 0.309)
β_{17} perimeter.length.o	(1.209, 1.300)	(1.241, 1.336)	(1.236, 1.331)
β_{18} perimeter.length.d	(0.385, 0.476)	(0.452, 0.551)	(0.461, 0.563)
β_{19} car.ownership.ratio.o	(3.155, 3.747)	(3.093, 3.735)	(3.195, 3.837)
β_{20} car.ownership.ratio.d	(-1.761, -1.172)	(-1.590, -0.916)	(-1.405, -0.764)
β_{21} highway.traffic.o	(0.007, 0.014)	(0.008, 0.015)	(0.007, 0.014)
β_{22} highway.traffic.d	(0.049, 0.056)	(0.047, 0.054)	(0.046, 0.053)
β_{23} prov/munic.traffic.o	(0.250, 0.289)	(0.257, 0.299)	(0.255, 0.295)
β_{24} prov/munic.traffic.d	(0.850, 0.889)	(0.853, 0.897)	(0.848, 0.891)
β_{25} distance	(-2.926 -2.885)	(-3.007, -2.961)	(-2.960, -2.915)
θ theta	(0.947, 0.983)	–	–
σ^2 sigma square	–	(1.043, 1.087)	–
ζ zeta	–	–	(0.357, 0.393)

TABLE 3.10. 95% credible intervals from 4,000 posterior draws of the Poisson-gamma (PG), Poisson-lognormal (PLN) and Poisson-inverse Gaussian (PIG) models.

It is also worth noting that despite some differences in parameters estimates obtained by the first OD matrix (Table 3.7 – negative binomial) and the second OD matrix (Table 3.9) the main inferences concerning the signs and the significance of parameters do not alter. Based on mean-over-standard deviation ratios the most significant variable, as mentioned, is distance followed by yearly kilometers-driven in provincial/municipal roads, perimeter length of origin zones, population density in destination zones and so forth.

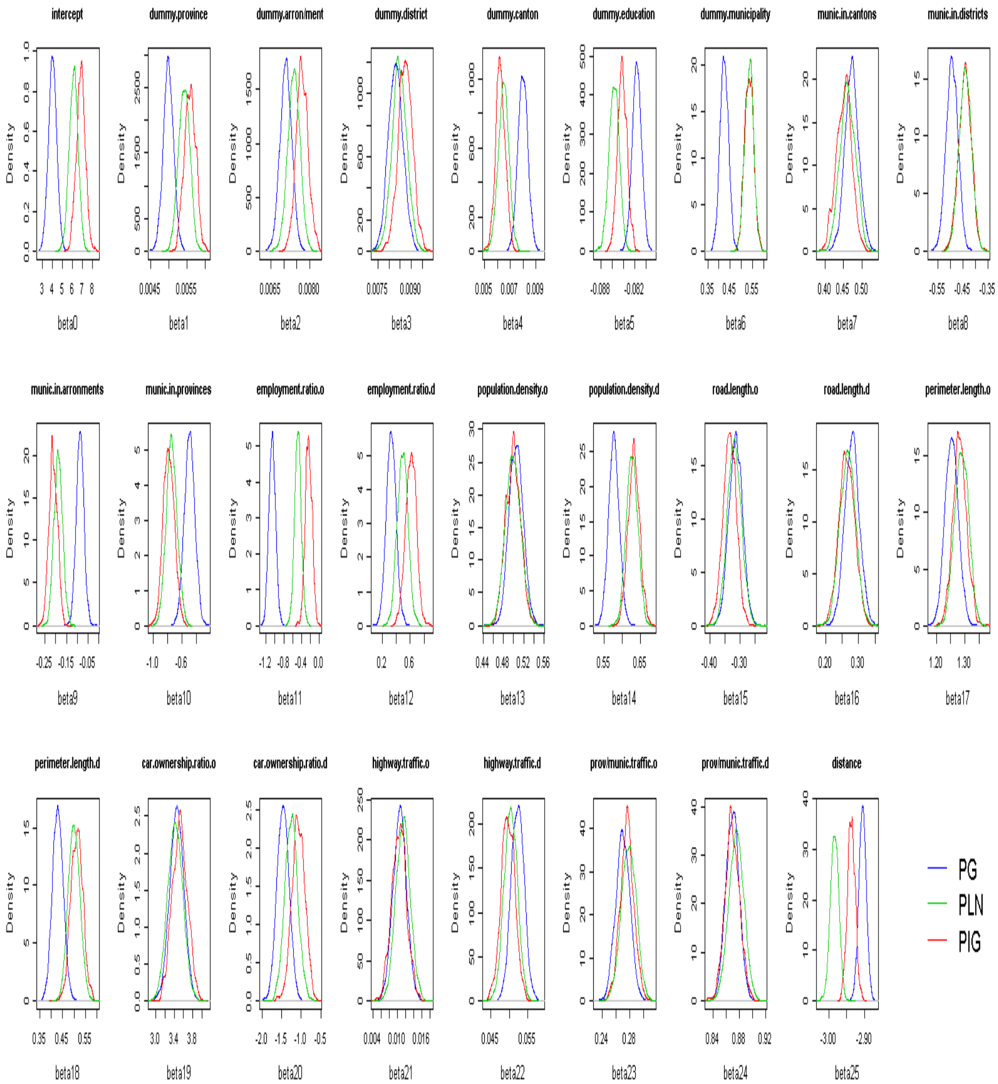

Figure 3.5 Kernel posterior distributions from 4,000 draws of the 26 regression parameters for the PG model (in blue), PLN model (in green) and PIG model (in red).

Kernel smoothed estimates of the posterior distributions of the regression parameters are plotted in Figure 3.5. Differences in the PG regression parameters are evident, especially for the intercept. It is also shown that the marginal posterior distributions are all relatively symmetrical, close to normal distributions. Histograms of the dispersion parameters are presented in Figure 3.6.

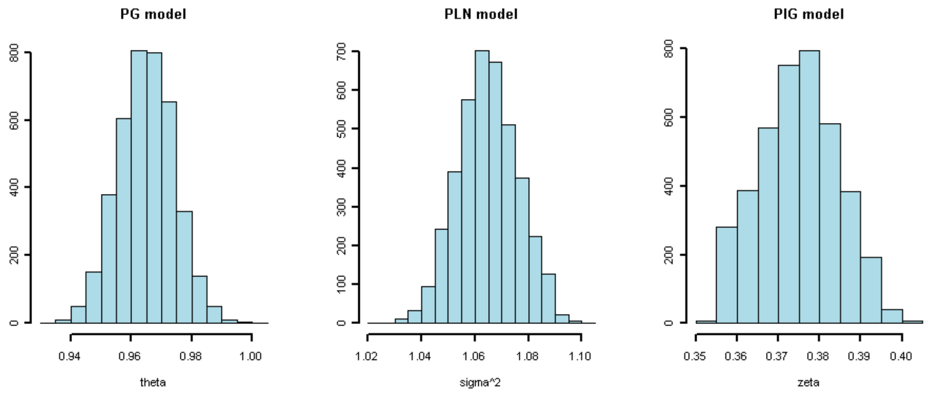

Figure 3.6 Histograms from 4,000 posterior draws of parameters theta, sigma square and zeta.

Model comparison based on information criteria is summarized Table 3.11. The AIC, BIC and marginal DIC statistics were calculated from the marginal negative binomial, PLN and PIG likelihoods based on the 4,000 posterior draws of each model. AIC and BIC are based on the posterior mean of the deviance[34]. Marginally, all three criteria give more support to both the PLN and PIG models over the PG model which also provides a justification for the similarity of the posterior estimates from the two models. The result is partially anticipated since the PLN and PIG are in theory more appropriate for cases of highly positive-skewed count data (Willmot, 1990). Furthermore, all three criteria favor the PIG marginal likelihood more than the PLN marginal likelihood, indicating that the PIG distribution is the most appropriate marginal sampling distribution.

The hierarchical DIC's are calculated from the hierarchical Poisson likelihoods based on random-without-replacement samples of 500 parameter draws from the PG and PIG models. Reduced samples were used again due to memory limitations and the random effect vectors were generated as

[34] Deviances based on posterior vector means and on posterior vectors maximizing the marginal likelihood were also considered for the calculation of AIC and BIC. The resulting AIC and BIC values are different, but do not alter the conclusion discussed here.

$\mathbf{u}^{(m)} \sim Gamma(\mathbf{y} + \theta^{(m)}, e^{\mathbf{X}^T \boldsymbol{\beta}^{(m)}} + \theta^{(m)})$ from the PG model and as $\mathbf{u}^{(m)} \sim GIG(\mathbf{y} - 1/2, 2e^{\mathbf{X}^T \boldsymbol{\beta}^{(m)}} + \zeta^{(m)}, \zeta^{(m)})$ from the PIG model, for $m = 1, 2, \ldots, 500$. The hierarchical DIC for the PLN model could not be calculated, since direct sampling of random effects is not possible for this model. Based on the hierarchical DIC values distinguishing a "better" hierarchical model is not as clear as with the marginal models. The difference is marginal and does not provide sufficient evidence in favor of one of the two models. Therefore, a solid conclusion cannot be drawn with respect to which of the two is more appropriate for predictive purposes.

Model selection criterion	**PG**	**PLN**	**PIG**
AIC	281519.2	279386.9	278469.1
BIC	281774.6	279642.3	278724.5
DIC (marginal)	281492.2	279362.4	278442.2
DIC (hierarchical)	224141.4	-	224146.1

TABLE 3.11. The AIC, BIC and marginal/hierarchical DIC values for the three models. AIC and BIC are based on the posterior mean of deviance. The marginal DIC is calculated from the marginal likelihoods, whereas the hierarchical DIC is calculated from the hierarchical Poisson likelihoods.

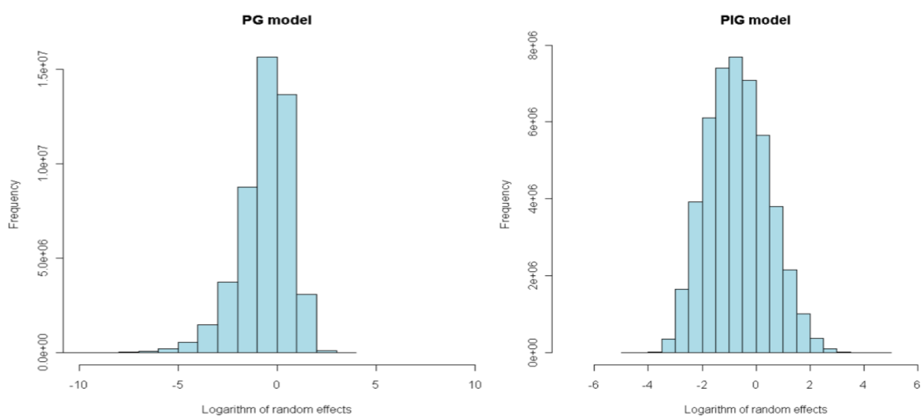

Figure 3.7 Histograms of random effects on log-scale from the reduced PG and PIG samples of 500 posterior draws.

Finally, it interesting to note that the posterior distributions of the random effects from the two models present some dissimilarity. The approximate range of the PG random effects is from 1.39×10^{-8} to 43.95 and from -18.09 to 3.78 on log-scale, while the approximate range of the PIG random effects is from 8.86×10^{-3} to 123.75 and from -4.73 to 4.82 on log-scale. Due to the GIG posterior distribution, the PIG random effects exhibit a longer right-tail than the PG random effects which are gamma distributed. On logarithmic scale the PIG random effects are relatively more symmetrical near 0 whereas the PG random effects have a longer left-tail. Histograms of the random effects on logarithmic scale are presented in Figure 3.7.

3.3.6 Predictive inference from the PG and PIG models

Due to the insufficient evidence regarding the most appropriate predictive model, predictions of OD flows were generated from both models. For overall goodness-of-fit, the absolute and squared distances are utilized and also the deviance. Each test quantity is calculated for observed and predicted data over the 500 posterior draws. The expected values and the deviance are those of the Poisson distribution conditional on the regression and random effect vectors. Results are summarized in Table 3.12. In general, both models provide satisfactory Bayesian p-values – close to the ideal value of 0.5 – for squared distances. On the other hand, predictions from the PIG seem to replicate better the observed data for small deviations from the expected values and also with respect to the Poisson distributional assumption.

Test quantity	**Formula**	**Bayesian p-value**	
		PG	**PIG**
Absolute distance	$\sum_{i=1}^{n}\lvert y_i - E(y_i \mid \boldsymbol{\beta}, u_i)\rvert$	0.276	0.424
Squared distance	$\sum_{i=1}^{n}(y_i - E(y_i \mid \boldsymbol{\beta}, u_i))^2$	0.468	0.512
Deviance	$-2\sum_{i=1}^{n}\log p(y_i \mid \boldsymbol{\beta}, u_i)$	0.992	0.620

TABLE 3.12. Bayesian p-values for the absolute distance, squared distance and deviance test quantities from 500 posterior draws of the PG and PIG models.

Nevertheless, as discussed in section 2.4.2, Bayesian p-values provide certain indications but are not to be formally compared across models. Therefore, we proceed with a more detailed investigation of the differences between the predictions of the two models. To this end, the particular case-specific tests presented in section 3.2.5 provide an interesting insight. The corresponding predictive distributions from both models are presented in Figure 3.8. Once again, all p-values are within acceptable limits. Nevertheless, it is interesting to note that although the distributions in Figures 3.8(a) and (c-d) more or less concur, the predictions for the total number of flows from the two models in Figure 3.8(b) are quite different in location with the PIG distribution positioned more to the left.

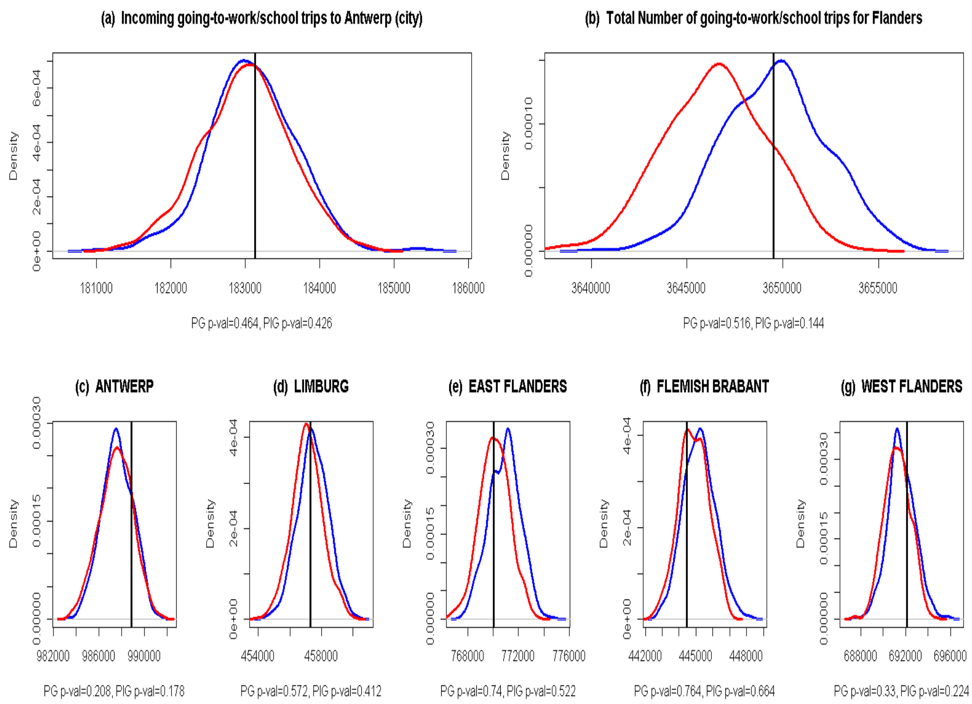

Figure 3.8 Kernel estimates of predictive distributions for going-to-work/school trips from the PG model (in blue) and the PIG model (in red) for (a) incoming trips to the municipality of Antwerp, (b) total number of trips in Flanders and intra-zonal trips for the five Flemish provinces; (c) Antwerp, (d) Limburg, (e) East Flanders, (f) Flemish Brabant and (g) West Flanders. The vertical black lines indicate the observed quantities.

In order to explain the discrepancy in Figure 3.8(b) a cell by cell examination of the extremeness of the predictions for the 94,864 OD pairs was performed. In overall, the predictions from both models are close to the observed flows and provide a satisfactory fit. The cell by cell examination reveals that approximately 88% of the p-values lie within 0.025 and 0.975 for both models. The 12% of the cells which is not predicted as well are mainly zero and low-valued cells. Following this result, a closer examination regarding how well the models replicate the total number of zero-valued cells resulted to Bayesian p-values equal to 0 for the two models which means that both underestimate the total number of zero-valued OD pairs. The predictive distributions for the total number of zero-valued OD pairs are shown in Figure 3.9.

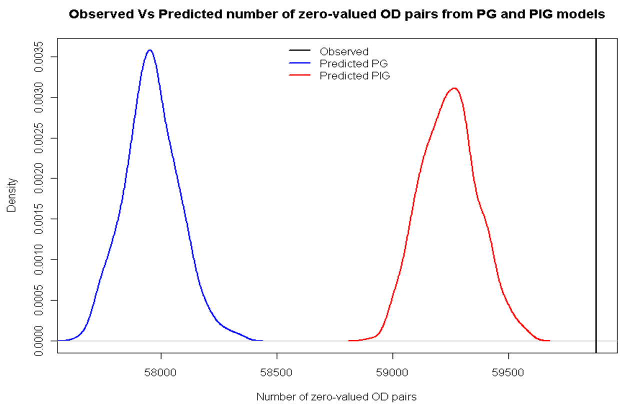

Figure 3.9 Predictive distributions for number of zero-valued OD pairs from the PG model (in blue) and PIG model (in red), the vertical black line corresponds to the observed number of zero-valued OD pairs.

As illustrated in Figure 3.9, both models underestimate the number of zero-valued cells, nevertheless underestimation is more extreme for the PG model. Given that the OD matrix is approximately 63% zero-valued, the finding is interesting as it implies two things. First, that the PIG model predicts relatively better than the PG model in this particular aspect which is also in accordance with the lower p-value for absolute distances in Table 3.12, and second that the distribution from the PG model in Figure 3.8(b) is more well-centered not due to model consistency but due to more extreme underestimation of the total number of zero-valued cells which have a strong cumulative influence on the total number of OD flows.

3.4 A note on a direct-demand gravity type perspective

All models described previously assume a log-linear relationship between the Poisson rate and the 25 explanatory variables and random effects. The multiplicative relationship for $i = 1, 2, \ldots, n$ is

$$\lambda_i = e^{\beta_0} e^{\beta_1 x_{1i}} \cdot \ldots \cdot e^{\beta_6 x_{6i}} e^{\beta_7 x_{7i}} \cdot \ldots \cdot e^{\beta_{10} x_{10i}} e^{\beta_{11} x_{11i}} e^{\beta_{12} x_{12i}} \cdot \ldots \cdot e^{\beta_{23} x_{23i}} e^{\beta_{24} x_{24i}} e^{\beta_{25} x_{25i}} u_i \,.$$

As discussed in section 3.1.4, variables $\mathbf{x}_1$ to $\mathbf{x}_6$ are dummy variables while variables $\mathbf{x}_7$ to $\mathbf{x}_{25}$ are discrete/continuous variables in log scale. Let us denote with $\mathbf{z}_j$ the untransformed variables, i.e. $\mathbf{x}_j = \log \mathbf{z}_j$, for $j = 7, 8, \ldots, 25$. From this category of variables, $\mathbf{z}_7$ to $\mathbf{z}_{10}$ (total number of municipalities between the administrative levels) and $\mathbf{z}_{25}$ (distance) have a one-to-one correspondence with the vector of OD flows **y**. On the other hand, variables $\mathbf{z}_{11}$ to $\mathbf{z}_{24}$ come in pairs since they distinguish between origin and destination attributes, for instance $\mathbf{z}_{11}$ refers to employment ratio of origin while $\mathbf{z}_{12}$ refers to employment ratio of destination. Thus, $\mathbf{z}_{12}$ is actually $\mathbf{z}_{11}$ reordered, or vice versa.

Then, if we set $B_s = e^{\beta_s}$ for the intercept and the dummies, i.e. for $s = 0, 1, \ldots, 6$, re-denote the distance variable with **d**, parameters $(\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \ldots, \beta_{23}, \beta_{24}, \beta_{25})$ as $(\beta_{11}^O, \beta_{11}^D, \beta_{12}^O, \beta_{12}^D, \ldots, \beta_{17}^O, \beta_{17}^D, \beta^{Dst})$ – with the superscripts *O*, *D*, *Dst* corresponding to origin, destination and distance effects – and finally switch back to the matrix representation, i.e. $T_{od} \sim Pois(\lambda_{od})$ for $o, d = 1, 2, \ldots, m$, we have that

$$\lambda_{od} = B_0 B_1^{x_{1od}} B_2^{x_{2od}} \cdot \ldots \cdot B_6^{x_{6od}} z_{7od}^{\beta_7} \cdot \ldots \cdot z_{10od}^{\beta_{10}} z_{11o}^{\beta_{11}^D} z_{11d}^{\beta_{11}^D} z_{12o}^{\beta_{12}^D} z_{12d}^{\beta_{12}^D} \cdot \ldots \cdot z_{17o}^{\beta_{17}^D} z_{17d}^{\beta_{17}^D} f(d_{od})^{\beta^{Dst}} u_{od} \,,$$

where $f(\mathbf{d})$ is a function of distance in our case the identity function, i.e. $f(\mathbf{d}) = \mathbf{d}$. This representation apart from the effects of the dummy variables $\mathbf{x}_1$ to $\mathbf{x}_6$, the variables $\mathbf{z}_7$ to $\mathbf{z}_{10}$ and the random effect vector **u**, bears a strong resemblance to traditional gravity-type and direct-demand transportation models (see for instance chapters 5 and 6 in Ortúzar and Willumsen, 2001). Of course, instead of marginal origin and destination totals which depend on balancing factors that need to be estimated (synthetic or gravity models) or only population and income effects (direct-demand models) a wider range of geographic and socio-economic attributes is utilized together with distance and exponential effects, which are estimated within a statistical framework, are assumed. In addition, the usual transportation modeling assumptions for the function of distance are either $f(\mathbf{d}) = \mathbf{d}^{-1}$ or $f(\mathbf{d}) = \mathbf{d}^{-2}$ and not the identity function[35]. Nevertheless, the resemblance is worthwhile noting as it provides a more familiar modeling interpretation within the transportation field.

It should be noted that the material presented throughout this chapter is not the first Bayesian OD modeling approach. In fact, in the discussion paper by West (1994) a Bayesian gravity-model for transportation forecasting is presented. The main difference is that in West (1994) the functional gravity form is used directly and without explanatory variables, whereas our approach starts from an additive log-linear form with explanatory variables. Nevertheless, the similarities in methodology and also conclusions are striking. The Bayesian gravity-model of West (1994) is based on the Poisson distribution and also on proposed extensions of the Poisson with random effects which are either lognormal or gamma distributed. The concluding discussion mentions the possibility of extending the gravity-model by including explanatory variables and also emphasizes on the necessity of random effects for applications on large-scale OD matrices.

Finally, it should be noted that similar types of gravity models are also commonly used in econometrical and trade-flow studies (see for instance the studies of Anderson, 1979; Bergstrand, 1985).

3.5 Summary

In this chapter we introduced a statistical modeling approach with covariates for the Flemish OD derived from the 2001 Belgian census study. Poisson mixtures and Bayesian methods are advocated as a suitable working framework for modeling large, over-dispersed and highly-skewed OD datasets. Initially, a comparison was performed between the simple Poisson model – a modeling assumption which is frequently utilized for OD flows – and the over-dispersed negative binomial model. Model comparison indicated that the negative binomial distribution is clearly a more suitable distributional assumption. Furthermore, the negative binomial model was presented from the hierarchical or Poisson mixture perspective as a Poisson-gamma model with random effects accounting

[35] Initially, all three functional forms were considered and eventually the identity function was selected as it resulted to lower values of model selection information criteria.

for heterogeneity across OD pairs. The Poisson-gamma model was then compared to the Poisson-lognormal model which up to date is the predominant alternative option and also to the Poisson-inverse Gaussian model. The latter – a model not as popular as its competing alternatives – was found not only to provide the best marginal fit, but that it also has desired distributional properties very much alike the Poisson-gamma model and unlike the rather cumbersome Poisson-lognormal model.

A set of 25 explanatory variables was used to model the expected OD trips under the assumption of a log-link function. All parameters proved to be statistically significant and the related inference led to consistent interpretations which demonstrate a correspondence with traditional trip-production and trip-attraction studies as well as recent transportation studies. In addition, the particular experimental design allows for an alternative interpretation of the Poisson mixture models as direct-demand, gravity models. An interpretation which brings this current research closer to the traditional and still frequently used four-step modeling approach.

One advantage of the Bayesian approach is the provision of a well defined predictive framework. The benefit from a transportation perspective is that one can predict the short-term distribution of any type and/or combination of trips that are of interest. Predictive distributions may serve as predictive scenarios useful in transport planning and policy evaluation. The benefit from a statistical perspective is that goodness-of-fit can be assessed in many different ways. Predictions of OD flows were generated from both Poisson-gamma and Poisson-inverse Gaussian models and the proximity of these predictions to the observed data was evaluated according to several measures of discrepancy. The overall fit was found in general to be satisfactory. Nevertheless, one important finding is that both models tend to underestimate zero-valued cells. Although replicating the number of zero-valued cells was initially not one of the primary goals of the analysis, it is shown that zero-valued cells have a strong cumulative influence on total travel-demand.

It is also worth noting that Bayesian methods provide in general a flexible framework for improving and keeping up-to-date the estimates of a given model. If further data become available, e.g. from a new census study, then the non-informative prior distributions presented here would be replaced by informative prior distributions based on the current posterior estimates, and a new application of Bayes' theorem would then result to updated posterior estimates that borrow strength from both the old and the new OD datasets. In addition, the new distributional assumptions regarding prior and likelihood densities would not be necessarily constrained to the distributional assumptions presented in this research.

Finally, it is arguable that the proposed methodology may serve as an effective alternative to the traditional four-step transportation model for cases in which historical OD data exist. From this point of view the methodology may be seen as a joint trip generation and trip distribution model which integrates the first phases of a four-step model in statistical models which provide a wider inferential horizon.

4 Bayesian inference on traffic assignment and network congestion

The methods introduced in this chapter provide a framework which is suitable for enhancing information from traffic assignment in terms of delivering stochastic estimates for traffic flows on links. Stochastic variability is associated to the travel-demand uncertainty related to the origin-destination OD matrix used as input into a given assignment model and therefore the methodology is not constrained by the choice of the latter. In general, the proposed framework relates closely with two important contemporary research directions; the first is traffic assignment under demand uncertainty, the second is identification of congested links.

Naturally, the methodology is based on the Bayesian predictive framework which, as demonstrated previously, provides a suitable working framework for generating multiple OD matrices from the corresponding predictive distribution of a given statistical model. As a consequence predictive inference for link flows is straightforward to implement either by assigning summarized OD information for approximate network inference or by performing multiple assignments for full network inference. The proposed methodology is tested on the Flemish road under the assumption of a deterministic user equilibrium assignment model.

The chapter begins with literature reviews concerning the topics of traffic assignment under demand uncertainty and congested/critical link identification. A description of the Flemish road network follows. Next, the two OD inputting methods are presented. Results for the Flemish network are discussed and analyzed initially by considering only the predictions from the Poisson-gamma model. These results are subsequently compared to results based on Poisson-inverse Gaussian predictions. The chapter ends with a summary of results and conclusions. The material of this chapter is based on Perrakis et al. (2012b,d,e).

4.1 Traffic assignment under demand uncertainty

As discussed in the introduction of this dissertation, traffic assignment is perhaps the most crucial part of transportation analysis. Traffic assignment models take into account the dependencies among OD demand, link flows and path costs, and simulate the interactions between transportation supply and travel demand in order to deliver an output which describes the mean state of a transportation system and its corresponding overall variability (Cascetta, 2009). Traffic assignment methods flourish in the relative literature; ranging from simple deterministic/stochastic uncongested network models and deterministic/stochastic user equilibrium and system optimum models, for the cases of congested networks, to more advanced methods such as equilibrium assignment with variable demand, multiclass assignment and dynamic process models. Descriptions of models and algorithms can be found in numerous books as in Thomas (1991), Patriksson (1994) and Cascetta (2009), to name a few. Extensive information for deterministic and stochastic user equilibrium (UE) assignment is also found in the articles of Florian and Hearn (1995) and Cantarella and Cascetta (1998), respectively.

The output of traffic assignment models is generally vital to decisions related to infrastructure expansion and transport policy measures, and simple point estimates – even if they refer to the most likely values – are not sufficient for a proper and safe assessment of the risks associated with such decisions. For instance, commonly used assignment models such as deterministic user equilibrium (DUE) and stochastic user equilibrium (SUE) deliver by definition deterministic solutions, i.e. point estimates of link flows without corresponding measures of statistical dispersion[36]. Therefore, despite the wide range of available assignment models, the need of quantifying precisely the uncertainty, which is related to traffic assignment estimates, is strongly present.

This necessity did not pass unnoticed as in recent years many studies are orientated towards this research direction. Initially, Waller et al. (2001) showed that in DUE assignment the expected performance of a network is, not only, not equivalent to the performance of the system under the expected value of travel-demand, but that the latter case is also suboptimal. Ukkusuri and Waller (2006) extended previous work and investigated the performance of seven point estimates of OD demand under the UE assumption. In the studies of Gardner et al. (2008, 2010), the impact of demand uncertainty on the pricing of transportation networks is explored by evaluating network performance resulting

[36] Despite the stochastic formulation, the equilibrium solution for the SUE model is deterministic based on a large sample approximation. More details can be found in section 5.1.1.

from single point demand-approximations, multiple points of inflated/deflated demand and meta-heuristic approaches. Sampling approaches are demonstrated in Duthie et al. (2011) where correlated OD demand realizations are sampled from multivariate truncated-at-zero normal and multivariate lognormal distributions, and iteratively used as input into the user equilibrium model. The authors further investigate the performance of different sampling techniques – Monte Carlo, quasi-Monte Carlo, antithetic, latin hypercube and control variates sampling – in terms of relative bias and error considerations. Other approaches aim to incorporate long-term demand stochasticity directly in the assignment problem, resulting to modifications of the DUE formulation as a bi-level, non-linear, non-convex mathematical optimization problem. For instance, Ukkusuri et al. (2007) propose a robust network design problem and utilize a genetic algorithm for the solution, while Ukkusuri and Patil (2009) formulate a flexible network design problem which can be solved under complementarity constraints.

4.2 Congested link identification and critical links

Congestion analysis is one of the primary aims of traffic assignment as it often serves as a basis for transport planning and investment decisions concerning road expansions, re-routing schemes and toll pricing among other issues. Congestion on a given link is typically measured by the *volume-over-capacity* (V/C) ratio. According to the Highway Capacity Manual (1994) congestion is in a state of “under-capacity”, ”at-capacity”, “near-capacity” or “over-capacity” for respective V/C ratios smaller than 0.85, between 0.85 and 0.95, between 0.95 and 1 or for values greater than 1. Although different ranking criteria are frequently adopted depending on the nature and scope of specific projects and case studies, a V/C ratio greater than 1 is generally accepted as an indicator of severe congestion since in this case the traffic volume on a given link exceeds the theoretical capacity of the link. In general, the V/C ratio is the most basic and commonly used measure for congestion inference. It is also worth noting that the V/C ratio is also commonly used as an explanatory variable for prediction of traffic accidents in traffic studies as for instance in Frantzeskakis and Iordanis (1987), Zhou and Sisiopiku (1997) and Lord et al. (2005). Within the context of traffic assignment, V/C ratios are estimated quantities and are consequently subjected to the deterministic limitations of the assignment models which are common in practice. Thus, in many real world applications inference is based on point estimates of V/C ratios without relative measures concerning the variability of the estimates.

Uncertainty in congestion estimation becomes particularly influential when related to *critical link* identification which is customarily the subject of *vulnerability analysis*. The concept of vulnerability of a road network is defined by Berdica (2002) as the susceptibility to incidents that result in considerable reductions of network performance. Thus, analysis of vulnerability focuses on traffic incidents and the consequent responses to those incidents, and evaluates the impact on a network by means of properly defined vulnerability indices which typically rely on traffic assignment procedures. Within this domain, state-of-the-art approaches are based on the full network scan approach (Jenelius et al., 2006; Taylor et al., 2006; Taylor, 2008) which – in short – involves iteratively removing every link of a network and assessing the impact of its removal through a traffic assignment procedure. Nevertheless, the associated computational burden constitutes the implementation of full network scan algorithms prohibitive, as yet, for applications on large-scale, congested networks. Other studies aim to reduce this computational burden by finding appropriate strategies for pre-selecting the potentially critical links based on minimum OD cost paths or high-choice probability links through SUE assignment (D'Este and Taylor, 2003; Taylor and D'Este, 2004), or on measures based on V/C ratio and various variations of V/C (Knoop et al., 2007). Recently, Chen et al. (2012) introduced critical link identification for large-scale, congested networks under demand uncertainty with a reduced-scan approach on what is defined as an *impact area*, i.e. a local area of influence surrounding a critical link.

4.3 The road network of Flanders

The road network of Flanders runs a total length of 65,296.72 kilometers. The road network with the corresponding borders of the 5 Flemish provinces, Antwerp, Limburg, East Flanders, Flemish Brabant and West Flanders, is presented in Figure 4.1. The circled areas indicate the capital-municipality of each province, the size of each circle is an approximate representation of the population of each municipality. Antwerp is the most populated capital, followed by Ghent, Leuven, Bruges and finally Hasselt. It should be reminded that trips originating from or terminating within Brussels metropolitan area, which is also marked in Figure 4.1, are not included in the proceeding analysis. Brussels metropolitan area, although being situated within the Flemish region, is a separate administrative region.

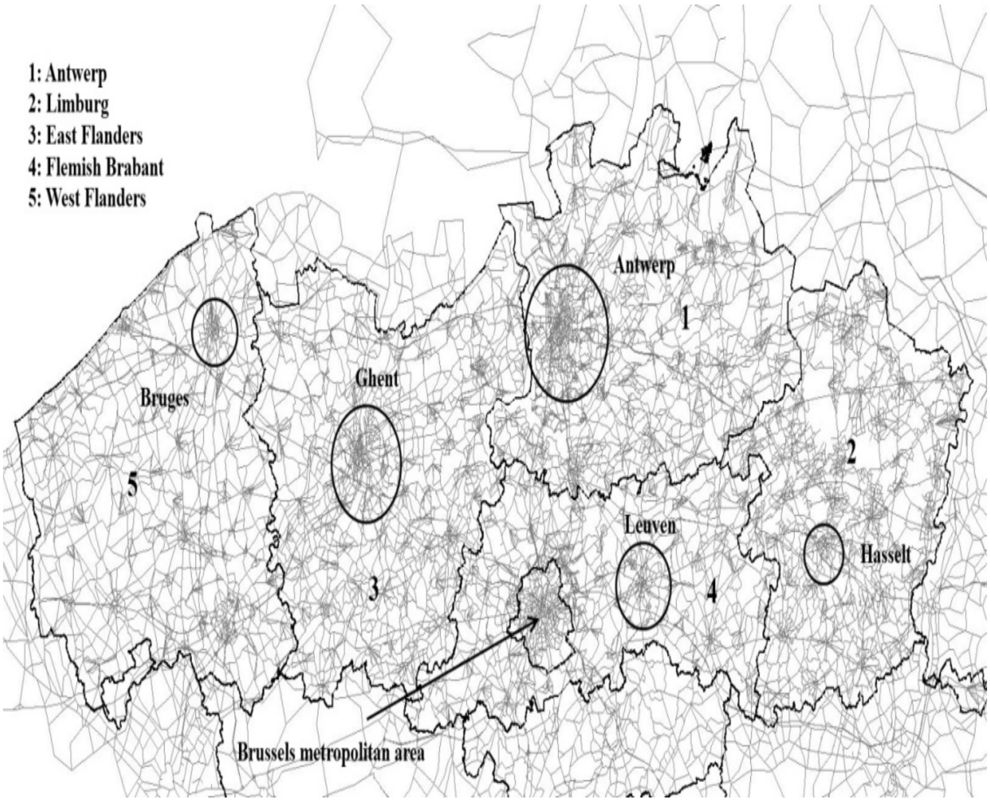

Figure 4.1 The road network of Flanders and the 5 Flemish provinces of Antwerp, Limburg, East Flanders, Flemish Brabant and West Flanders with corresponding capitals; Antwerp, Hasselt, Ghent, Leuven and Bruges.

In overall the network contains 97,450 links which can be categorized into highways (including entrance/exit road segments), main regional roads, small regional roads, local municipal roads and walk/bicycle paths. The majority of links corresponds to local municipal road segments (52.91%), followed by links belonging to small regional roads (21.1%), main regional roads (15.48%), highways (8.58%) and finally walk/bicycle paths in municipalities, countryside and near train stations (1.93%). The latter category is by default not taken into account during traffic assignment[37].

Link category	Total number of links	Percentage of links
Highways (including exits & entrance)	8,357	8.58%
Main regional roads	15,091	15.48%
Small regional roads	20,562	21.1%
Local municipal roads	51,557	52.91%
Walk/bicycle paths	1,883	1.93%
Sum	**97,450**	**100%**

TABLE 4.1. Categories of links with the corresponding totals and percentages of each category for the 97,450 links comprising the Flemish road network.

The total number of links for each of the aforementioned categories and the corresponding percentages with respect to the total number of links are summarized in Table 4.1.

4.4 Quantifying input-uncertainty with Bayesian OD predictions

Following the notation which was introduced in section 1.1, let I denote the total number of links in a network and let $\mathbf{v} = (v_1, v_2, \dots, v_I)^T$ be the corresponding vector of link flows or volumes. In addition, let **A** represent the assignment operator. All of the estimates discussed next will be of course conditional on **A** which implies that alteration of the parameters of the assignment model or – more important – of the choice of the assignment model itself will result to potentially different estimates. Finally, in order to simplify notation the OD predictions are now denoted by $\mathbf{y}^{(m)}$ instead of $\mathbf{y}^{pred(m)}$, for $m = 1, 2, \dots, M$.

4.4.1 Method 1 for approximate network inference – assigning OD summary statistics

In the first method the predictive vectors $\mathbf{y}^{(m)}$, for $m = 1, 2, \dots, M$, are utilized in calculating a summary statistic of the OD matrix denoted by $S(\mathbf{y})$. In general, the OD summary vector is a function of the M predictions, that is

[37] In addition, for traffic assignment 2,451 connecting centroids are used. These are not actual road links, but are needed for implementing traffic assignment. When considering the connectors, the total number of links becomes 99,901, 2.5% of which are connectors.

$S(\mathbf{y}) = f(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(M)})$. By assigning $S(\mathbf{y})$ to a network, a corresponding estimate for link flows, denoted by $S(\mathbf{v})$, is obtained.

The most common point estimate is the mean $S(\mathbf{y}) = \bar{\mathbf{y}}$ which can be calculated as follows:

$$\bar{\mathbf{y}} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{pmatrix} = \frac{1}{M}\left[\begin{pmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_n^{(1)} \end{pmatrix} + \begin{pmatrix} y_1^{(2)} \\ y_2^{(2)} \\ \vdots \\ y_n^{(2)} \end{pmatrix} + \ldots + \begin{pmatrix} y_1^{(M)} \\ y_2^{(M)} \\ \vdots \\ y_n^{(M)} \end{pmatrix}\right] = \frac{1}{M}\begin{pmatrix} y_1^{(1)} + y_1^{(2)} + \ldots + y_1^{(M)} \\ y_2^{(1)} + y_2^{(2)} + \ldots + y_2^{(M)} \\ \vdots \quad \vdots \quad \vdots \\ y_n^{(1)} + y_n^{(2)} + \ldots + y_n^{(M)} \end{pmatrix}.$$

The $n \times 1$ mean vector $\bar{\mathbf{y}}$ can then be used as the OD-input in an assignment model which will yield a $l \times 1$ vector $\bar{\mathbf{v}}$ corresponding to the mean estimate of link flow vector $\mathbf{v}$, i.e. $\mathbf{A}\bar{\mathbf{y}} = \bar{\mathbf{v}}$ or in a more illustrative notation:

$$\bar{\mathbf{y}} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{pmatrix} \xrightarrow{\mathbf{A}} \bar{\mathbf{v}} = \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \\ \vdots \\ \bar{v}_l \end{pmatrix}.$$

Specifically, $\bar{\mathbf{v}}$ is an estimate of $E(\mathbf{v} \mid \bar{\mathbf{y}}, \mathbf{A})$ the expected vector of link flows conditional on the predictive expectations of OD flows but also conditional on the assignment model.

For interval estimates the appropriate summary statistics are percentile vectors, i.e. $S(\mathbf{y}) = \mathbf{y}^p$ for the p-th percentile. Estimation of a percentile vector is not as straightforward as the calculation of the mean vector. Calculating individually the corresponding percentile of each OD pair would result in a percentile vector which will be highly unlikely to occur, especially for percentiles near 0 or 100 and for a large number of OD pairs, i.e. for a large n. For example, if the 0^{th} and 100^{th} percentile vectors – corresponding to the minimum and maximum – are calculated as mentioned, then that implies the assumption that each OD pair is occurring exactly at its minimum/maximum, an assumption which is not realistic and which would lead initially to an erroneously small/large total demand figure and consequently to an erroneously wide interval estimate. Therefore, it is preferable to derive percentile vectors based on a function of vectors $\mathbf{y}^{(m)}$ which will operate as a criterion and constrain the estimates within realistic limits.

A natural selection for the criterion is the sum or total demand of each vector $\mathbf{y}^{(m)}$, i.e. $s^{(m)} = \sum_{i=1}^{n} y_i^{(m)}$, for $m = 1, 2, \ldots, M$. Under this approach, the percentile vector $\mathbf{y}^p$ is derived by calculating first the corresponding percentile $s^{(m)}$ from the values $\left\{s^{(1)}, s^{(2)}, \ldots, s^{(M)}\right\}$. Then, the vector which has the closest sum to s^p is chosen as $\mathbf{y}^p$. In case there are two or more sums which satisfy that condition, then $\mathbf{y}^p$ is set as the average of the corresponding vectors. In general,

$$\mathbf{y}^p = \left(\overline{\overline{K_p}}\right)^{-1} \sum_{k=1}^{\overline{\overline{K_p}}} \mathbf{y}^{(k)},$$

where $K_p = \left\{y^{(k)} : \left|s^{(k)} - s^p\right| = \min_m \left\{\left|s^{(m)} - s^p\right|\right\},\ m = 1, 2, \ldots, M\right\}$ and $\overline{\overline{K_p}}$ is the cardinality of set K_p. A common percentile pair is $(\mathbf{y}^{2.5}, \mathbf{y}^{97.5})$ which corresponds to a 95% interval estimate. Note that $\mathbf{y}^{50}$, the vector corresponding to the median, can also be calculated through this procedure as an additional point estimate to the mean. For the special cases of the minimum and maximum vectors we simply have to find $s^{\min}$ and $s^{\max}$, respectively, and then calculate vectors $\mathbf{y}^{\min}$ and $\mathbf{y}^{\max}$ as follows

$$\mathbf{y}^{\min} = \left(\overline{\overline{K_{\min}}}\right)^{-1} \sum_{k=1}^{\overline{\overline{K_{\min}}}} \mathbf{y}^{(k)} \text{ and } \mathbf{y}^{\max} = \left(\overline{\overline{K_{\max}}}\right)^{-1} \sum_{k=1}^{\overline{\overline{K_{\max}}}} \mathbf{y}^{(k)},$$

where

$$K_{\min} = \left\{\mathbf{y}^{(k)} : s^{(k)} = \min_m \left\{s^{(m)}\right\},\ m = 1, 2, \ldots, M\right\},$$
$$K_{\max} = \left\{\mathbf{y}^{(k)} : s^{(k)} = \max_m \left\{s^{(m)}\right\},\ m = 1, 2, \ldots, M\right\},$$

with $\overline{\overline{K_{\min}}}$ and $\overline{\overline{K_{\max}}}$ being the corresponding cardinalities of sets $K_{\min}$ and $K_{\max}$. After calculating a specific percentile vector $\mathbf{y}^p$ that is of interest, the corresponding estimate for link flows $\mathbf{v}^p$ is obtained by assigning $\mathbf{y}^p$ to the network. The results will once again depend on the choice of the assignment model.

It should be noted that the recommended percentile derivation approach based on the total demand requires OD predictions which are well-validated for the majority of OD pairs, at least for the non-negligible, inter-zonal OD pairs which have a significant contribution to the total demand. From a Bayesian perspective well-validated implies that predictions include the observed quantities within certain credible intervals (e.g. 95%) with a reasonable range. Obviously, the same approach is not applicable for random permutations across the cells of a given OD matrix or similar sampling approaches which will eventually result to the same total demand but will alter completely the traffic flow dynamics induced by the demand distribution of each OD pair[38].

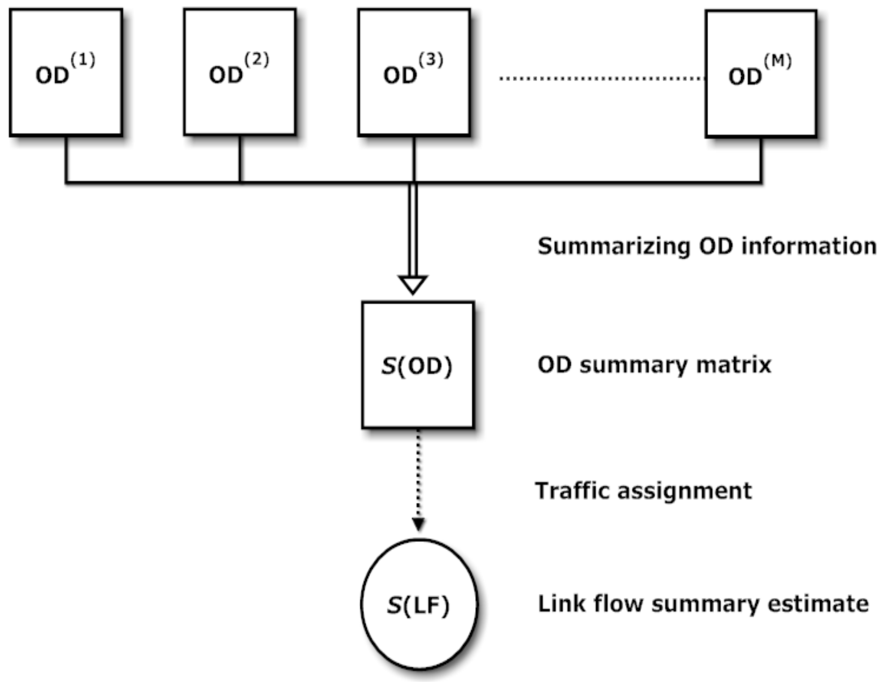

Figure 4.2 Using M predictive OD datasets to calculate an OD summary matrix S(OD) which is used as input in traffic assignment, resulting to a link flow summary estimate S(LF).

[38] The point seems trivial, but is worth mentioning in order to avoid misconceptions with relative research applied on a different context. For instance, a similar problem is observed in OD estimation from link counts (discussed in section 1.2) where the goal is to find the most plausible OD matrix among many candidate OD matrices which can potentially replicate an observed, reduced vector of link counts.

A graphical representation of the method is provided in Figure 4.2. Similar approaches exist with respect to calculation of centrality point estimates; for instance, besides of the mean and the median the additional point estimates discussed in Ukkusuri and Waller (2006) can also be utilized. On the other hand derivation of percentile estimates is not pursued to a large degree in the relative literature, therefore the approach presented here does contribute to that direction to some extent. Nevertheless, the main contributions of this paper rely on the method presented next.

4.4.2. Method 2 for full network inference – assigning multiple OD's

Method 2 involves assigning all M OD predictions individually in order to obtain M corresponding vectors of link flows. This method is computationally more intense than method 1, but also delivers full information for link flows in the form of distributional estimates. In this case, an individual assignment must be implemented for each $\mathbf{y}^{(m)}$, for $m = 1, 2, \ldots, M$. In vector notation, we have that

$$\mathbf{y}^{(1)} = \begin{pmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_n^{(1)} \end{pmatrix} \xrightarrow{\mathbf{A}} \mathbf{v}^{(1)} = \begin{pmatrix} v_1^{(1)} \\ v_2^{(1)} \\ \vdots \\ v_l^{(1)} \end{pmatrix}, \mathbf{y}^{(2)} = \begin{pmatrix} y_1^{(2)} \\ y_2^{(2)} \\ \vdots \\ y_n^{(2)} \end{pmatrix} \xrightarrow{\mathbf{A}} \mathbf{v}^{(2)} = \begin{pmatrix} v_1^{(2)} \\ v_2^{(2)} \\ \vdots \\ v_l^{(2)} \end{pmatrix}, \ldots, \mathbf{y}^{(M)} = \begin{pmatrix} y_1^{(M)} \\ y_2^{(M)} \\ \vdots \\ y_n^{(M)} \end{pmatrix} \xrightarrow{\mathbf{A}} \mathbf{v}^{(M)} = \begin{pmatrix} v_1^{(M)} \\ v_2^{(M)} \\ \vdots \\ v_l^{(M)} \end{pmatrix}.$$

Overall summary statistics can be calculated, only this time directly from the vectors $\mathbf{v}^{(m)}$. For instance, the mean is calculated as follows

$$\bar{\mathbf{v}} = \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \\ \vdots \\ \bar{v}_l \end{pmatrix} = \frac{1}{M}\left[\begin{pmatrix} v_1^{(1)} \\ v_2^{(1)} \\ \vdots \\ v_l^{(1)} \end{pmatrix} + \begin{pmatrix} v_1^{(2)} \\ v_2^{(2)} \\ \vdots \\ v_l^{(2)} \end{pmatrix} + \ldots + \begin{pmatrix} v_1^{(M)} \\ v_2^{(M)} \\ \vdots \\ v_l^{(M)} \end{pmatrix}\right] = \frac{1}{M}\begin{pmatrix} v_1^{(1)} + v_1^{(2)} + \ldots + v_1^{(M)} \\ v_2^{(1)} + v_2^{(2)} + \ldots + v_2^{(M)} \\ \vdots \quad \vdots \quad \vdots \\ v_l^{(1)} + v_l^{(2)} + \ldots + v_l^{(M)} \end{pmatrix}.$$

Note that this an estimate of the expectation of vector $\mathbf{v}$ conditional on *all* OD predictions and on the assignment model, that is $E(\mathbf{v} \mid \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots \mathbf{y}^{(M)}, \mathbf{A})$. A graphical representation of method 2 is provided in Figure 4.3.

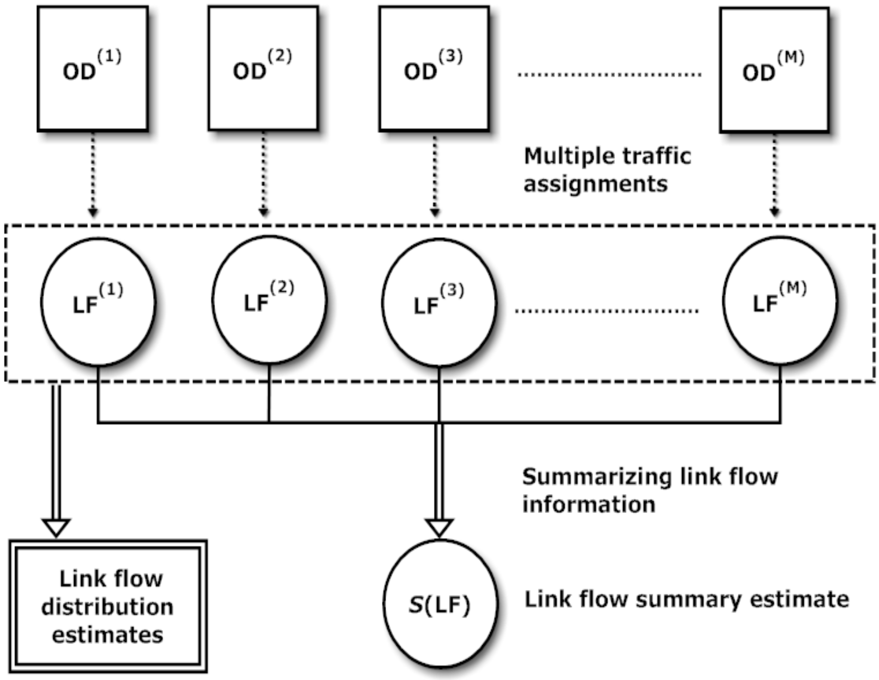

Figure 4.3 Performing M individual assignments results to link flow summary estimates S(LF) as well as distribution estimates.

With this method, estimates of percentile vectors are straightforward to calculate. Any percentile vector $\mathbf{v}^p = (v_1^p, v_2^p, \ldots, v_l^p)$ is calculated directly from the M vectors $\mathbf{v}^{(m)}$, i.e. v_j^p, for $j = 1, 2, \ldots, l$, is estimated individually as the p-th percentile obtained from the corresponding sample values $\left\{v_j^{(1)}, v_j^{(2)}, \ldots, v_j^{(M)}\right\}$. In general, the vectors $\mathbf{v}^{(m)}$ contain all necessary information for the links of the network. For inference on a specific link v_j, point, interval and dispersion estimates, such as the variance and the standard deviation, or even the distribution of v_j can be estimated directly from the sample $\left\{v_j^{(1)}, v_j^{(2)}, \ldots, v_j^{(M)}\right\}$. Consider for instance the calculation of V/C probabilities; let us denote by c_j the theoretical capacity of link j and let us assume that interest lies in calculating the probability that v_j / c_j equals or exceeds a specific threshold value h. If $p_j(\cdot)$ is the probability density of v_j, then an estimate of the following quantity is needed

$$\Pr(v_j / c_j \geq h) = \int_{v_j \geq hc_j} p_j(v_j) dv_j \, .$$

Given the sample $\{v_j^{(1)}, v_j^{(2)}, \ldots, v_j^{(M)}\}$, this probability is straightforward to estimate through direct Monte Carlo (MC) integration, that is

$$\widehat{\Pr}(v_j / c_j \geq h) = M^{-1} \sum_{m=1}^{M} I_j^{(m)} \text{ , where } I_j^{(m)} = \begin{cases} 1, \text{ if } v_j^{(m)} \geq hc_j, \\ 0, \text{ otherwise.} \end{cases}$$

4.5 Network inference under DUE assignment and PG predictions

In this section, results from the two inputting methods previously discussed are presented for the Flemish road network under the assumptions of DUE assignment and PG predictions. The main purpose of this section is to highlight the differences between full network inference (method 2) and approximate network inference (method 1). Thus, the focus of interest in this section lies on the impact of input-uncertainty. Model-uncertainty originating from the choice of statistical OD modeling and traffic assignment modeling is further investigated in sections 4.6 and 5.1, respectively.

The DUE model is utilized which is one of the most commonly used traffic assignment models. The model is based on Wardrop's 1st principle (Wardrop, 1952), also known as the *equilibrium principle*, which states that:

"The journey times on all the routes actually used are equal, and less than those that would be experienced by a single vehicle on any unused route".

The principle leads to a user-optimal approach where the users of a network aim to minimize their individual travel costs or travel times. The approach can also be interpreted in terms of game theory. Interesting notes on related theoretical and mathematical aspects can be found in Patriksson (1994), extensive information for UE assignment is also available in Florian and Hearn (1995). The model is deterministic in the sense that it is based on the assumption that users have perfect knowledge of the travel costs.

In short, DUE assignment uses an iterative process for reaching to a convergent solution in which travelers cannot reduce their travel times by switching routes. At each iteration link capacity restraints and link flow-dependent travel times are taken into account in order to calculate link flows. As a mathematical program DUE assignment is solved with the Frank-Wolfe algorithm (Frank and Wolfe, 1956). As *link performance function* the common BPR formulation (Bureau of Public Roads, 1964) is adopted which relates link travel times to volume over capacity (V/C) ratios, specifically

$$t = t_f\left[1 + a\left(\frac{v}{c}\right)^{\beta}\right],$$

where t is link travel time, t_f is link free-flow travel time, v is link volume (flow), c is link capacity and a, β are calibration parameters which according to the BPR formulation are set equal to 0.15 and 4, respectively.

Concerning implementation of inputting method 2, 500 individual assignments were executed each taking as input one predictive OD from the PG model[39], thus resulting in 500 corresponding link flow or link volume vectors. That is, $\mathbf{A}^{DUE}\mathbf{y}^{(m)} = \mathbf{v}^{(m)}$, for $m = 1, 2, \ldots, 500$, where $\mathbf{v}^{(m)} = (v_1^{(m)}, v_2^{(m)}, \ldots, v_l^{(m)})$ and l is the total number of network links which equals to 97,450 for the Flemish network[40]. The assignments concern the morning peak-hour interval between 7 am and 8 am for a normal weekday. It should be noted that since the initial OD matrix concerns work/school trips made by Flemish residents the assignment-related inference that follows is also restricted to work/school trips of Flemish residents without considering potential traffic originating from the metropolitan area of Brussels or from the French-speaking region of Walloon to the south of Belgium. In addition, as already mentioned, links within the metropolitan area of Brussels are not included in the analysis. Nevertheless, the highway ring around Brussels metropolitan area is included.

4.5.1 Sample size considerations

Initially, the adequacy of the sample size is checked by examining the behavior of the cumulative average of certain OD attributes, namely the total demand (sum), the mean and also the standard deviation. Intra-zonal municipal flows are not taken into account any more as they do not influence the traffic assignment.

---

[39] See section 3.2.5.

[40] Software TransCAD (Caliper Corporation, 2008) was used to execute all assignments either based on OD summaries (method 1) or for repeated assignments (method 2).

Results are presented in Figure 4.4, as illustrated the the three attributes start to stabilize roughly after 100 OD predictions and become almost constant after 200 predictions.

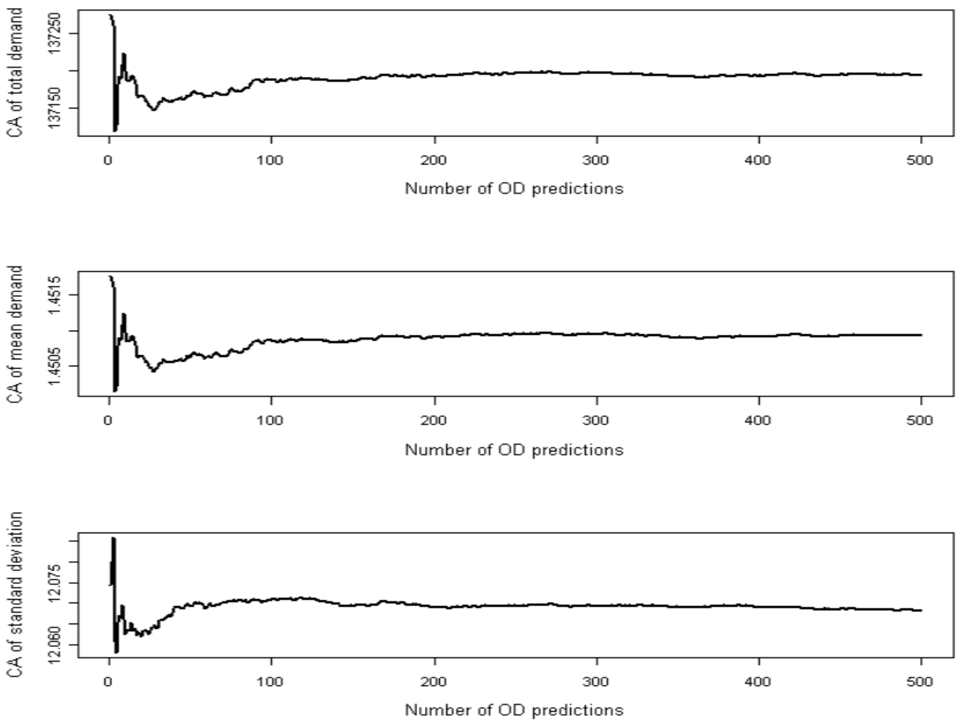

Figure 4.4 Cumulative averages (CA) of total demand, mean demand and standard deviation for 500 PG predictions of inter-zonal OD flows between 7 am and 8 am.

4.5.2 The average state of the Flemish network

Global visualization is a first illustrative step which provides a general idea for the overall state of the network. The mean link flow vector is the most suitable summary for describing the average state of the network. In Figure 4.5, the results from DUE model and the mean link flow vector obtained from method 1 are visualized. In order to make Figure 4.5 simpler to comprehend only flows and V/C ratios for highway links are highlighted, the main findings are the following.

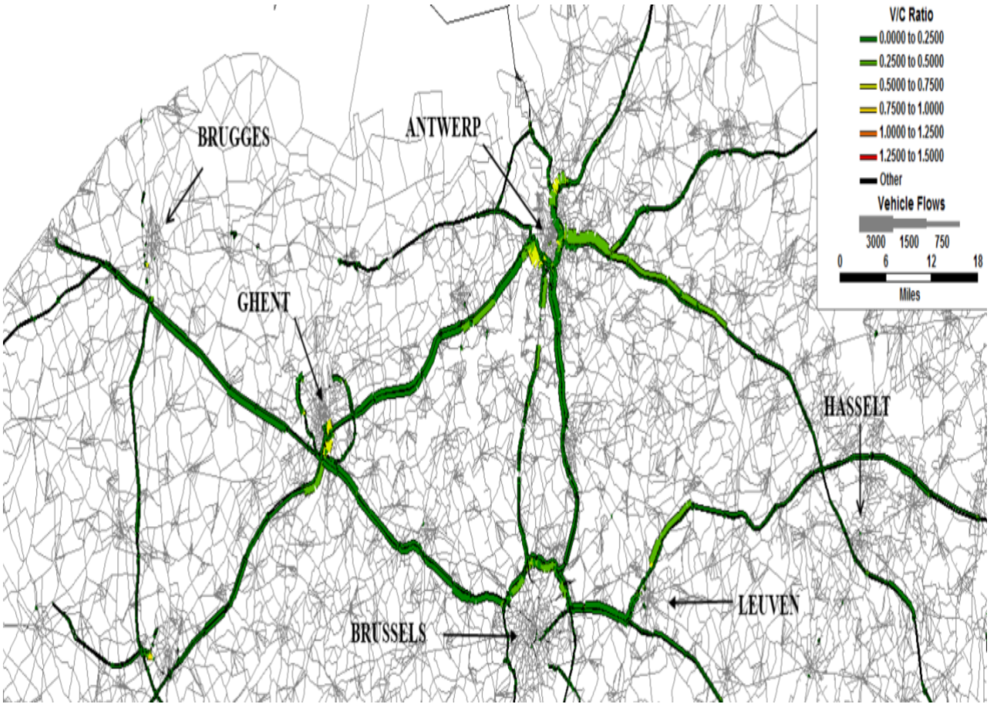

Figure 4.5 Mean link traffic flows in highways for work/school trips between 7 am and 8 am from DUE assignment and PG predictions based on the mean link flow vector from method 1.

V/C ratios are higher in specific segments on or near the highways rings of Antwerp (R1) and Ghent (R4) which can be identified by the yellow spots indicating V/C ratios between 0.5 and 0.75. Relatively high V/C ratios (light green color) also occur on the northern part of highway ring R0 around Brussels, on highway E40 near Leuven, highway E313 which connects Antwerp with Hasselt and to a lesser degree on highways E17 and E19 which connect Antwerp with Ghent and Brussels, respectively. In terms of link volumes, that is judging by the thickness of the lines, the busiest highway seems to be E313 as it approaches Antwerp from the east.

The corresponding assignment map for mean link flows from method 2 seems almost identical to the map of Figure 4.5 and therefore not presented here, since differences between the two methods are difficult to mark on a global scale[41]. The differences are examined in more detail next.

[41] The assignment map from method 2 is presented in section 4.6.1.

4.5.3 Jensen's inequality, centrality and percentile estimates

An interesting topic in the relative literature concerns the behavior of the total system travel time (TSTT) under uncertain demand. In Waller et al. (2001) it is discussed that in UE assignment the TSTT as a function of OD demand can be approximated by a convex function. The implications are first that TSTT is a random variable due to the stochastic OD demand and second that it is subjected to Jensen's inequality due to the fact that it is represented by a convex function. The OD demand vector is denoted by **y**, so if g is the objective function of UE which is approximately convex, then from Jensen's inequality we have that $g(E(\mathbf{y})) \leq E(g(\mathbf{y}))$. The quantity on the left side of the inequality is the TSTT based on $E(\mathbf{y})$ which is the mean demand, while the quantity on the right side is the mean of TSTT. What the inequality implies is that the "true" mean TSTT is potentially underestimated when the mean OD demand is used in UE assignment and since this is a very frequent strategy this theoretical consideration is alarming with respect to planning and investment decisions. An analytical description is provided by Waller et al. (2001).

Our working framework provides the opportunity to address this consideration empirically and see what the impact on the Flemish network is. The quantity $g(E(\mathbf{y}))$ is easily calculated from assigning the mean OD matrix (method 1) and adding up the resulting vehicle-hours travelled for each link, while the quantity $E(g(\mathbf{y}))$ is calculated by adding the vehicle-hours vectors for each one of the 500 assignments (method 2) and then calculating the average of the 500 TSTT values. Results are presented in Figure 4.6 where the two quantities of interest are superimposed on a kernel estimate of the distribution of TSTT. As illustrated in Figure 4.6, TSTT based on mean OD demand is indeed smaller than the expected TSTT given the OD variability, nevertheless the difference is relatively small; in numbers we have that TSTT[E(OD)] equals 1,387,009 vehicle-hours, while E(TSTT) equals 1,387,129 vehicle-hours. The 95% interval for TSTT is from 1,383,152 to 1,391,169. In practical terms and in the application presented here the estimate TSTT[E(OD)] is quite close to the target quantity. Nevertheless, a generalization of this conclusion is not warranted as the results depend to a large extent on network topology, level of congestion (as illustrated in Figure 4.5 the overall congestion level is low to medium) and also on the degree of variability between the OD predictions.

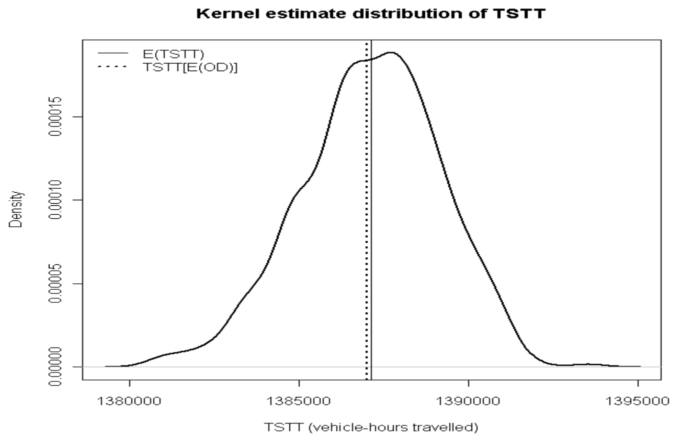

Figure 4.6 Kernel estimate of TSTT (vehicle-hours) distribution derived from the 500 values of TSTT. The solid line represents the expected TSTT estimate (method 2), the dotted line represents TSTT derived from the expected demand (method 1).

We proceed with investigating point and interval estimates derived from the two methods. In order to summarize results the link flows are aggregated according to road-type. The corresponding estimates are rounded and presented in Table 4.2. Centrality measures, i.e. the means and the medians, derived from the two methods are relatively close, since relative to the large magnitudes of the estimates the differences are in terms of hundreds. This provides some evidence that method 1 might be adequate if the purpose is solely to have an estimate of the central state of the network. Once again a generalization is not warranted. In principle, the expectation function is not exchangeable in UE due to non-linearity and factors such as congestion and the variability in the OD predictions are influential. From this point of view we emphasize on the importance of well-validated OD predictions for the majority of inter-zonal OD pairs[42]. Regarding the 95% interval and range estimates, the ones derived from method 1 prove to be clearly narrower than the corresponding intervals of

[42] As discussed in section 3.36, 88% of the predictions for the 94864 OD cells result to p-values which lie within a 95% credible interval. The remaining 12% of the cells which are not predicted as well are zero or low valued cells. As we saw, these are influential on the total demand, but on the other hand they do affect traffic flow interrelationships and dynamics.

method 2. The result is partially anticipated since the percentile estimates of method 1 are approximate and are derived by a further conditioning upon total demand. Therefore, compared to the intervals of method 2 the intervals of method 1 fail to capture a large part of the variability induced by the assignment procedure. Conclusively, method 2 is more reliable for interval estimation. In addition, method 2 provides link flow dispersion estimates. The standard deviation for each category of link volumes is also presented in Table 4.2.

Links types	Method 1				
	Mean	**Median**	**95% interval**	**Range**	**St.dev.**
Highways	1,565,723	1,565,536	(1,564,267 - 1,566,497)	(1,561,068 - 1,570,341)	-
Main regional roads	1,292,689	1,293,363	(1,291,685 - 1,294,257)	(1,288,723 - 1,295,259)	-
Small regional roads	796,803	797,499	(796,357 -798,115)	(793,479 - 800,656)	-
Local roads	450,905	450,263	(449,754 - 450,293)	(449,653 - 451,522)	-
	Method 2				
	Mean	**Median**	**95% interval**	**Range**	**St.dev.**
Highways	1,565,784	1,565,614	(1,559,724 - 1,572,337)	(1,554,757 - 1,576,062)	3,312
Main regional roads	1,292,461	1,292,363	(1,288,519 - 1,296,716)	(1,285,404 - 1,298,399)	2,143
Small regional roads	797,313	797,061	(794,092 - 802,817)	(793,451 - 805,032)	1,991
Local roads	450,659	450,704	(448,947 - 452,236)	(447,668 - 452,842)	835

TABLE 4.2. Mean, median, 95% interval (2.5%-97.5%), range and standard deviation estimates obtained from the two methods for highway, main regional road, small regional road and local road traffic flows.

Finally, link flow distributions can be estimated directly from the 500 link flow vectors obtained from method 2. Kernel estimates for flows in highways, main regional roads, small regional roads and local roads are presented in Figure 4.7. Except of the small regional road distribution, the rest are symmetrical and close to normal distributions. This is to be expected since these distributions refer to aggregated flows which according to the central limit theorem should converge to normality, asymptotically. On the other hand, the distribution of small regional roads exhibits a long right-tail. A possible explanation is that individual small regional road irregularities do not completely “wash-out” because such roads often demonstrate all-or-nothing oscillations during UE assignment due to the fact that they are closer to zonal loading points where there are few re-routing alternatives.

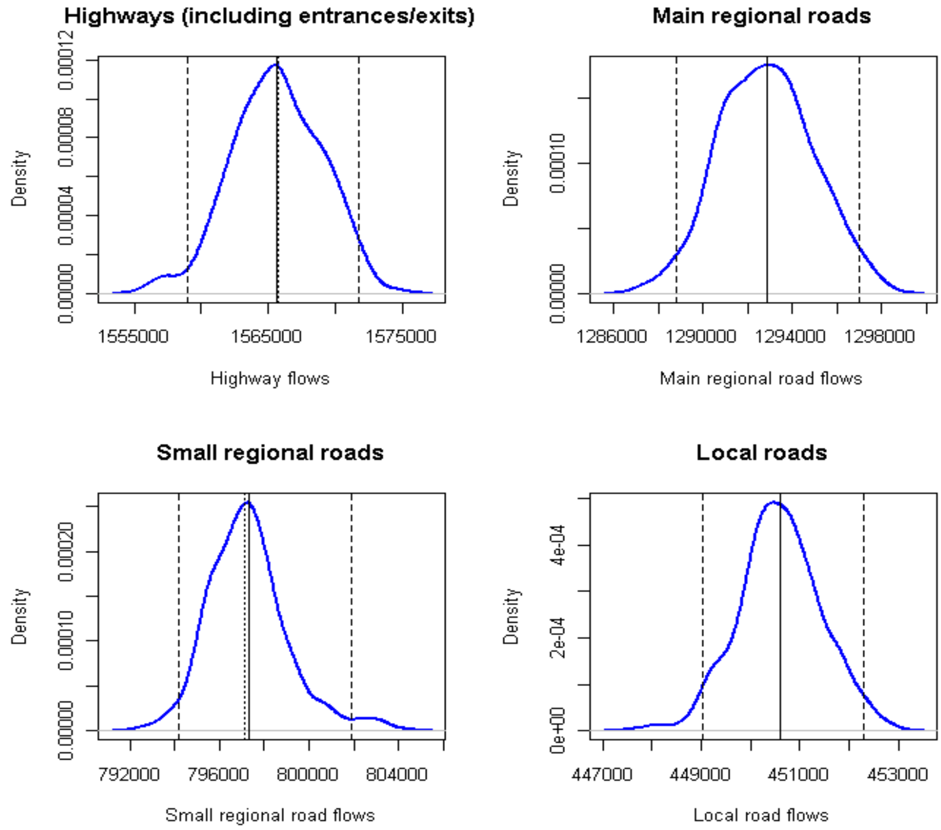

Figure 4.7 Kernel distributions from method 2 for traffic flows on highways, main regional roads, small regional roads and local roads. The solid, dotted and dashed vertical lines correspond to the mean, median and 95% interval of each distribution, respectively.

4.5.4 Congested link identification

As congested links we define those links in which the V/C ratio exceeds a specific threshold value h with a certain probability, i.e. $\Pr(V/C \geq h)$. Evaluation of congested links in terms of probability estimates is safer in comparison to point estimates and also reduces the margin of uncertainty. For instance, the expected value of V/C may be smaller than h, nevertheless the probability of exceeding h may be significantly greater than zero. This point is illustrated further in the proceeding analysis. In general, the value $h = 1$ is the common option since values greater than 1 imply that the traffic flow exceeds the capacity limit, thus resulting in congestion. In our case though, inference is limited in traffic flows for work and school trips made only by Flemish residents, which means that we would expect the V/C ratios to be higher if the proportion of traffic related to other trip-purposes and to non-Flemish residents was included in the analysis. From this perspective threshold values smaller than 1 may also be regarded as critical, but since the exact proportion of trips that is unaccounted for is not known exactly, h can only be selected on a heuristic basis. As a conservative choice and in order not to overestimate the number of congested links the value of 0.95 is adopted.

Eleven links are identified for $h = 0.95$, the V/C distributions of these links with the corresponding probabilities of exceeding 0.95 are presented in Figure 4.8. Two remarks can be made, based on the distributions of Figure 4.8, regarding V/C distributions and consequently link flow distributions from DUE assignment[43]. First, in contrast to the distributions presented in Figure 4.7, the V/C distributions of Figure 4.8 and consequently the corresponding link flow distributions resulting from DUE assignment are not necessarily close to normal distributions, for instance bimodality is observed. Second, the bimodalities appearing in some of the distributions may be attributed to the iterative user equilibrium procedure[44]. For instance, when the flows on a specific link and at a given iteration exceed a certain threshold – consequently leading to a high V/C ratio – and there exists an alternative link with has a cost which is close but lower, then in the following iteration a switch of flows will occur from the high-cost link to the low-cost link. This “switching” effect will eventually result to bimodal V/C distributions, as the ones presented in Figure 4.8. Results for the 11 congested links are also summarized in Table 4.3 with the corresponding link types, expected values and the probabilities of exceeding a V/C ratio of 0.95. Note that if the analysis was based on the expected V/C ratio instead of the probability of exceeding a V/C ratio of 0.95, then congestion on four out of the eleven links would not have been identified.

[43] The link flow distributions have exactly the same shape but are on a different scale, since they are proportional to the V/C distributions.

[44] Small sample size is excluded as being the cause of bimodalities. Initial results were based on execution of 100 assignments which were later increased to 200 assignments and finally to 500 assignments as presented here. The form of the distributions was the same under all sample sizes.

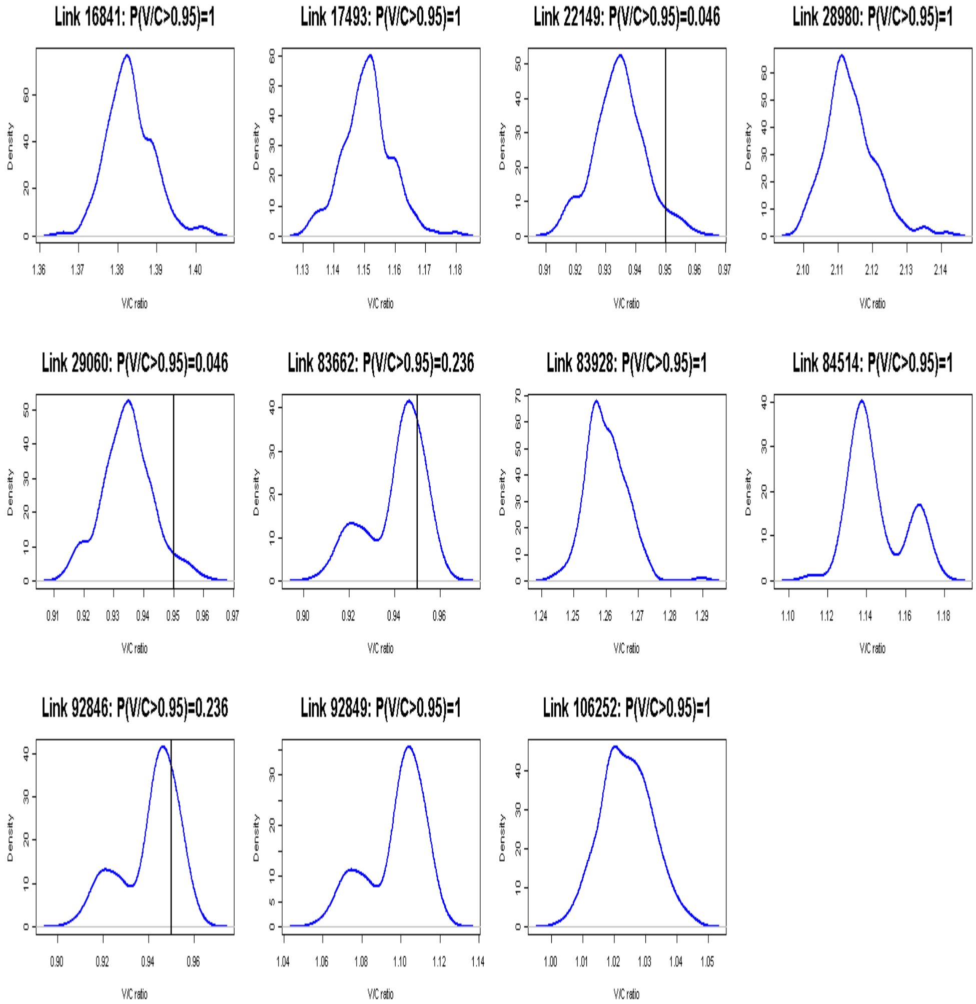

Figure 4.8 V/C kernel distribution estimates from method 2 for the 11 congested links which include or exceed the value of 0.95, highlighted by a vertical line in the distributions which include this value.

The results in Figure 4.8 and Table 4.3 show that seven out of the eleven congested links have a V/C value greater than 0.95 with probability 1. Visual examination of the distributions in Figure 4.8 additionally reveals that these seven links also exceed the value $h = 1$ with probability 1, except perhaps of link 106252 which has its minimum located near 1 and may therefore include smaller values than 1 with a low probability.

Congested link ID	**Link type**	E(V/C)	Pr(V/C ≥ h)
16841	Small regional road	1.384	1
17493	Local road	1.152	1
22149	Local road	0.935	0.046
28980	Highway	2.114	1
29060	Local road	0.935	0.046
83662	Local road	0.941	0.236
83928	Highway	1.260	1
84514	Main regional road	1.144	1
92846	Local road	0.941	0.236
92849	Local road	1.098	1
106252	Local road	1.024	1

TABLE 4.3. Congested links identified for $h = 0.95$ with the corresponding V/C expected values and the probabilities of exceeding a V/C of 0.95, based on 500 link flow vectors (method 2).

Conclusively, even without taking into consideration the proportion of traffic related to non-Flemish residents and to other trip-purposes than work or school trips, congestion on these 7 links is almost certain. The highest V/C ratio is observed in highway link 28980 with an expectation of 2.114, while regional road link 16841 and highway link 83928 follow with expected V/C ratios equal to 1.384 and 1.26, respectively. For the remaining 4 links, the probability of exceeding a V/C ratio of 0.95 is lower, equal to 23.6% for links 83662, 92846 and equal to 4.6% for links 22149 and 29060. Not surprisingly, the 11 congested links are situated near the major municipal centers of Antwerp, Ghent and Bruges; the exact locations are presented in Figure 4.9.

As illustrated in Figure 4.9, five of the congested links belong to the wider municipal area of Antwerp, including link 28980 which is a segment of R1 highway ring in the north of Antwerp and has the highest expected V/C ratio. Links 22149 and 29060 are local outgoing road segments near highway ring R1, whereas link 16841, which has the second highest V/C ratio, is a segment of N1 regional road directing right to the center of Antwerp. Finally, link 17493 is a local road segment outside Antwerp, nevertheless very close to Antwerp airport situated south-east of the city. Five congested links also appear in the municipal area of Ghent. Link 84514 to the north-east is a segment of N70 regional road very near the exit of the R4 highway ring with a direction to the center of Ghent, whereas links 83928, 83662, 92486 and 92849, near the center, are all road

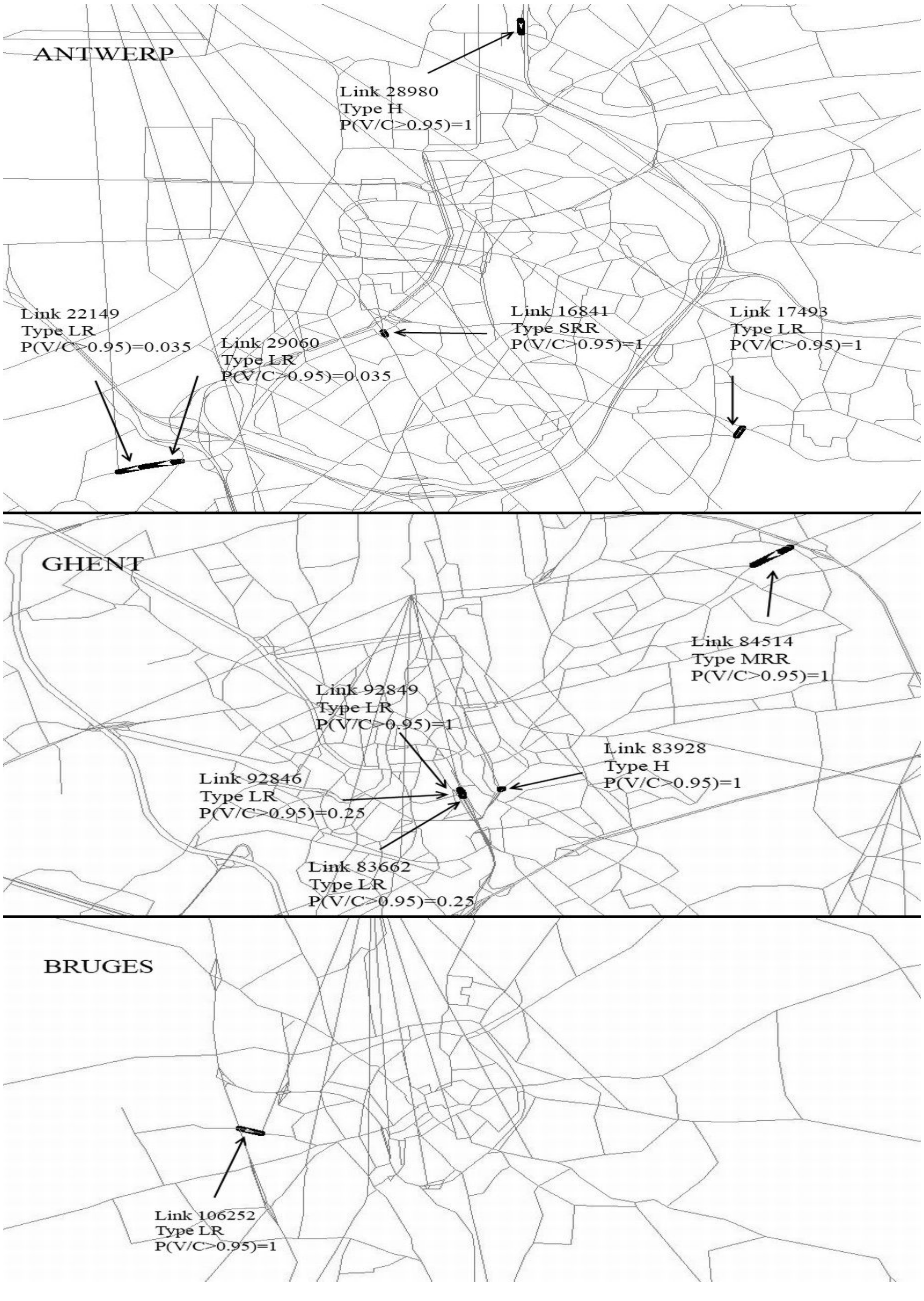

Figure 4.9 The congested links in the municipalities of Antwerp, Ghent and Bruges. The link type abbreviations H, MRR, SRR and LR correspond to Highways, Main Regional Roads, Small Regional Roads and Local Roads, respectively.

segments clustered around the end of the part of highway E17 that has a direction to the center of Ghent. Finally, one congested link appears near the municipality of Bruges that is link 106252. This link corresponds to a local road segment very near N31 regional road and with a direction towards the center of Bruges from the west.

The analysis of congested links is based on the relatively conservative choice $h = 0.95$. Naturally, for smaller values of h the number of congested links increases, e.g. for the values 0.9, 0.85, 0.8, 0.75, 0.7 the number of congested links rises to 16, 20, 25, 47 and 69, respectively. In general, for situations where the exact proportion of traffic is not known, as in the application presented in this study, the choice of h is under the control of the researcher or policy-planner. In such cases, inference may be based on more than one values of h. For cases in which there is certainty that all the potential traffic or – at least – most of the potential traffic of a network is included in the analysis, then the value $h = 1$ may be safely adopted.

4.6 Comparative network inference between PG and PIG predictions under DUE assignment

In this section the main results presented previously are compared to the results which are obtained when utilizing the PIG predictions as input for traffic assignment. The focus of interest is on potential differences between average network-states, TSTT, aggregated link flow distributions and V/C distributions from inputting method 2.

4.6.1 Differences in average states of the network

Maps of link volumes and V/C ratios for the average state of Flemish highways under method 2 are presented in Figure 4.10 for PG predictions and in Figure 4.11 for PIG predictions. It is difficult to distinguish any differences in the two maps, which implies that inference concerning the average state of highways is not influenced a lot, at least when visualizing this state on a global scale. In addition, as noted in section 4.5.3 when comparing Figure 4.10 with Figure 4.5 which is based on inputting method 1 it is also difficult to find any striking difference. This is in accordance with the mean estimates presented in Table 4.2

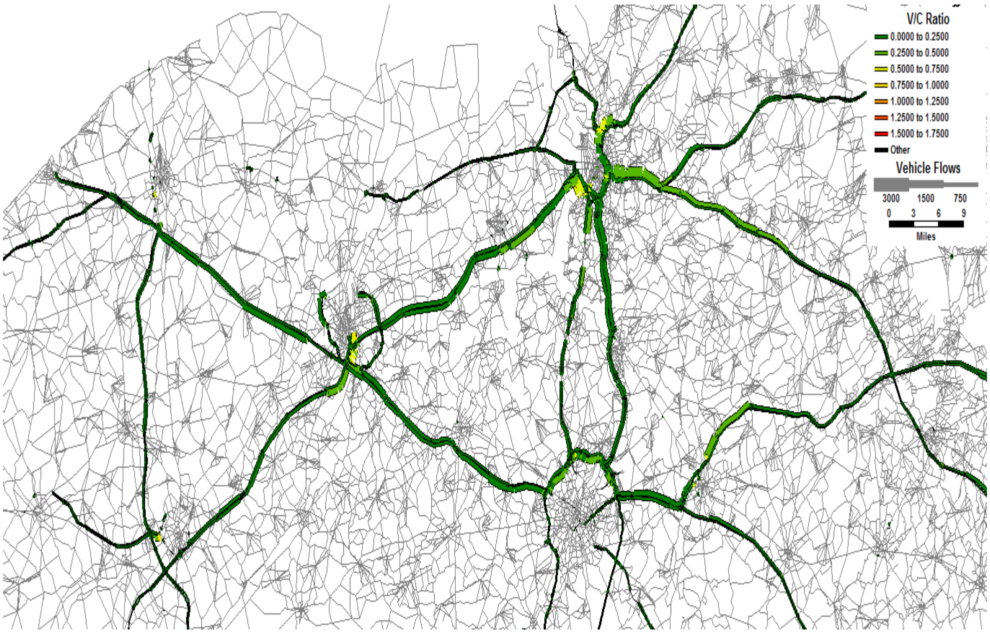

Figure 4.10 Mean link traffic flows in highways for work/school trips between 7 am and 8 am from DUE assignment and PG predictions based on the mean link flow vector from method 2.

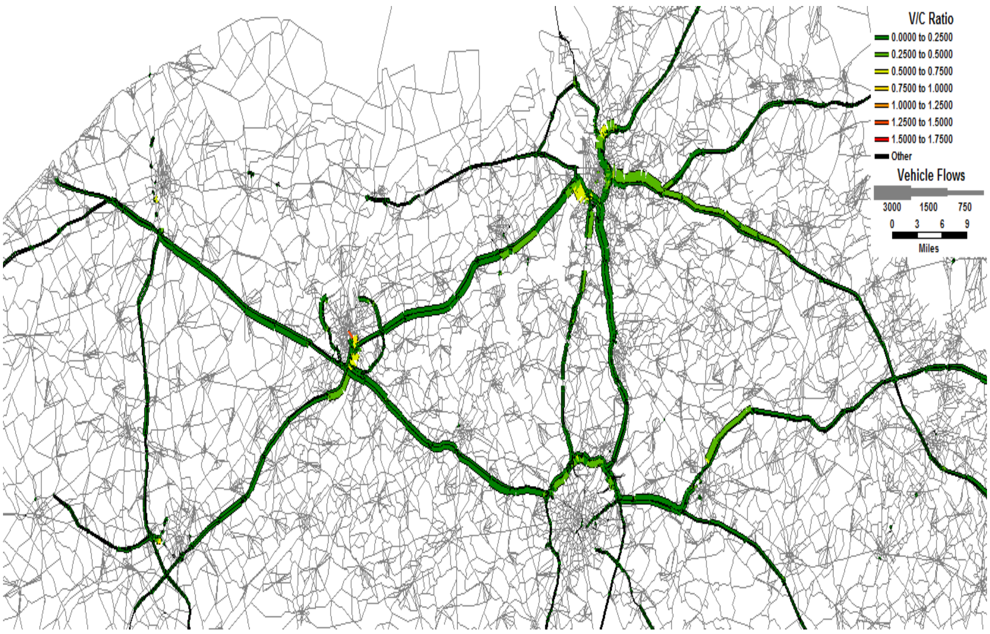

Figure 4.11 Mean link traffic flows in highways for work/school trips between 7 am and 8 am from DUE assignment and PIG predictions based on the mean link flow vector from method 2.

where it is shown that the total mean highway volume estimates are more or less the same, namely 1,565,723 from method 1 and 1,565,784 from method 2. Thus, according to Figures 4.10 and 4.11 the main findings which are discussed in section 4.5.3 regarding expected highway flows are the same under method for both the PG and PIG models.

4.6.2 Differences in TSTT and aggregated link flow distributions.

The resulting kernel estimated distributions of TSTT from the predictions of the two models are presented in Figure 4.12. The x-axes in Figure 4.12 are kept intentionally on the same range in order to highlight that the two TSTT distributions are not similar in terms of location, namely the distribution from PG predictions is located more to the right. The reason why this is happening is explained next.

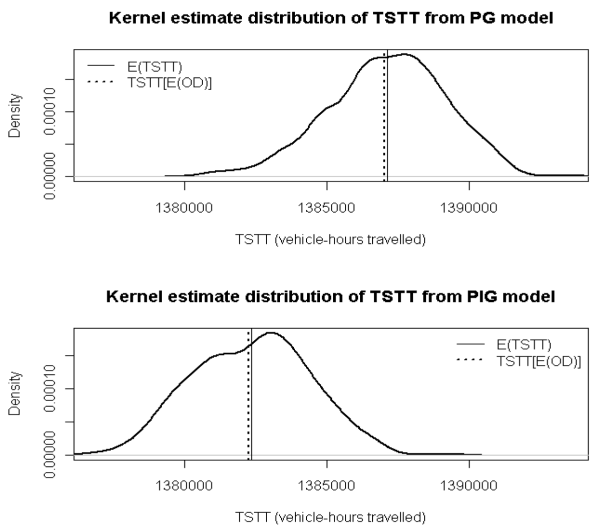

Figure 4.12 Kernel estimate of TSTT (vehicle-hours) distribution derived from 500 values of TSTT under PG and PIG predictions. The solid line represents the expected TSTT estimate (method 2), the dotted line represents TSTT derived from the expected demand (method 1).

As with PG predictions, the differences between TSTT under expected demand and the expected TSTT – conditional on PIG predictions – are practically not significant; TSTT[*E*(OD)] equals 1,382,246, *E*(TSTT) equals 1,382,352 and the 95% interval ranges from 1,383,048 to 1,390,812. The respective estimates from the PG models were 1,387,009, 1,387,129 and from 1,383,152 to 1,391,169. Thus, the location of the TSTT distribution under PG predictions is shifted about 5,000 vehicle-hours to the right.

The corresponding distribution estimates for link flows on highways, main regional roads, small regional roads and local roads which result from the predictions of the two models are presented in Figure 4.13. Once again the distributions based on PG predictions are shifted more to the right side of the x-axes. The differences observed in Figures 4.12 and 4.13 are explained by the findings presented in section 3.3.6. In that section, it is shown that the total travel-demand estimates of the two models are different, namely the total-travel demand from the PG model is greater than that of the PIG model. It is further shown that this is due to a greater degree of overestimation for the PG model regarding the total number of zero-valued cells. Evidently, this is also affecting the cumulative TSTT and link flow distributions presented in Figures 4.12 and 4.13. The explanation is simple as traffic assignment essentially distributes or allocates the total travel-demand on the available links of the network. Therefore, the greater the travel-demand is the greater will the resulting link volumes be.

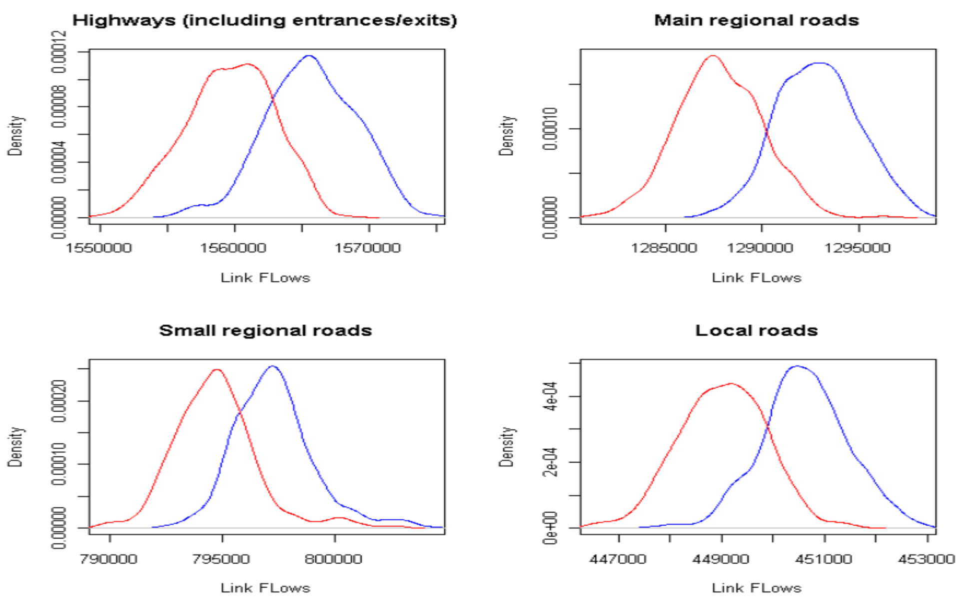

Figure 4.13 DUE link flow kernel estimated distributions for highways, main regional, small regional and local roads from 500 predictions of the PG (in blue) and PIG (in red) models.

It would be difficult at present to "choose" which link flow distributions are more appropriate. In general, the purpose of this section is to highlight the uncertainty originating from the choice of the statistical model and not choosing which predictions are "better". Given the fact that it is infeasible to compare these distributions with independent link traffic data, referring to the particular inferential problem at hand and for the whole Flemish network, choosing which predictions are closer to reality would be prone to extreme speculation. From a strictly statistical perspective that bounds inference to the specific OD data which are available, i.e. one-directional, going-to-work/school trips made by Flemish residents within Flanders, the distributions based on PIG predictions may be considered as more reliable due to the smaller degree of underestimation for the total number of zero-valued OD pairs.

Links types	Method 2				
	PG model				
	Mean	**Median**	**95% interval**	**Range**	**St.dev.**
Highways	1,565,784	1,565,614	(1,559,724 - 1,572,337)	(1,554,757 - 1,576,062)	3,312
Main regional roads	1,292,461	1,292,363	(1,288,519 - 1,296,716)	(1,285,404 - 1,298,399)	2,143
Small regional roads	797,313	797,061	(794,092 - 802,817)	(793,451 - 805,032)	1,991
Local roads	450,659	450,704	(448,947 - 452,236)	(447,668 - 452,842)	835
	PIG model				
	Mean	**Median**	**95% interval**	**Range**	**St.dev.**
Highways	1,559,602	1,559,782	(1,553,157 – 1,565,363)	(1,550,098 – 1,568,150)	3,259
Main regional roads	1,287,786	1,287,659	(1,283,352 – 1,292,018)	(1,281,312 – 1,296,229)	2,177
Small regional roads	794,718	794,625	(791,589 – 799,974)	(789,675 – 802,713)	1,893
Local roads	449,019	449,042	(447,364 – 450,574)	(446,512 – 451,519)	846

TABLE 4.4. Mean, median, 95% interval (2.5%-97.5%), range and standard deviation estimates for highway, main regional road, small regional road and local road traffic flows obtained from method 2 under PG and PIG predictions.

On the other hand, in practical terms the differences may not be as significant. Mean, median, interval and standard deviation estimates for the four types of link volumes are presented in Table 4.4. Given the large magnitudes of the estimates the differences are relatively not great. This is also depicted in the maps in Figures 4.10 and 4.11 from where the conclusions concerning highway volumes and V/C ratios are the same. It is also worth noting that the differences in Table 4.4 are larger for highway flows and are gradually decreasing for link types of smaller capacity.

4.6.3 Differences in congested link identification

Concerning identification of congested links the choice between PG and PIG predictions is not very influential. The same links presented in section 4.5.4 are also identified under PIG predictions. The PIG V/C distributions of these 11 links

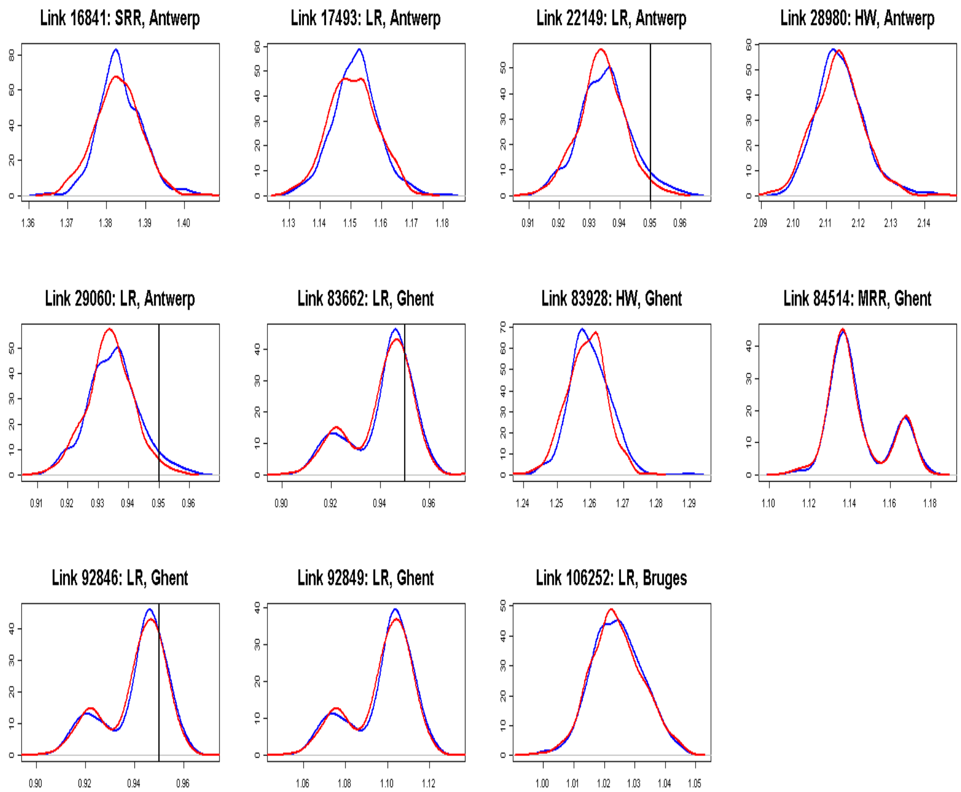

Figure 4.14 Kernel estimates of the PG (in blue) and PIG (in red) V/C distributions of the 11 congested links which either include or exceed the threshold value of 0.95 highlighted by a vertical black line in the distributions which include this value. The abbreviations HW, MRR, SRR and LR stand for highways, main regional roads, small regional roads and local roads.

are compared with the respective PG V/C distributions in Figure 4.14. As illustrated, the distributions under the two different sets of OD predictions almost concur. In contrast to aggregated link flow distributions as the ones presented in Figure 4.12, the disaggregated V/C distributions present no differences in location. This result is expected as the cumulative effect of underestimating zero-valued OD pairs is absent in disaggregated OD flows and disaggregated link flows, especially for links characterized by high V/C ratios which are commonly not related to OD pairs where zero-valued flows occur.

The expected values of the distributions presented in Figure 4.14 and the corresponding probabilities of exceeding a V/C ratio of 0.95 are presented in Table 4.5.

Congested link ID	**Link type**	**PG**		**PIG**	
		E(V/C)	Pr(V/C ≥ h)	E(V/C)	Pr(V/C ≥ h)
16841	Small regional road	1.384	1	1.383	1
17493	Local road	1.152	1	1.151	1
22149	Local road	0.935	0.046	0.934	0.022
28980	Highway	2.114	1	2.114	1
29060	Local road	0.935	0.046	0.934	0.022
83662	Local road	0.941	0.236	0.941	0.208
83928	Highway	1.260	1	1.259	1
84514	Main regional road	1.144	1	1.144	1
92846	Local road	0.941	0.236	0.941	0.208
92849	Local road	1.098	1	1.097	1
106252	Local road	1.024	1	1.024	1

TABLE 4.5. Congested links identified for $h = 0.95$ with the corresponding V/C expected values and the probabilities of exceeding a V/C of 0.95, based on 500 link flow vectors (method 2) derived from PG and PIG predictions.

The expected values under PG and PIG predictions are almost identical which is normal given that the V/C distributions are very similar in terms of location as well as shape. The probabilities of exceeding the threshold value of 0.95 are slightly higher under PG predictions which can be attributed to a small remaining effect of the higher total travel-demand figures predicted from the PG model.

4.7 Summary

In this chapter, a general approach for traffic-assignment and network-congestion inference based on Bayesian OD predictions was presented. In general, the material of this chapter relates closely with two relatively distinct research directions which receive increasing attention over the last years. The first is traffic assignment inference under travel-demand uncertainty. This research gives rise to interesting questions relating to the behavior of total travel time performance and to the reliability of link-volume point estimates under uncertain demand. The second is identification of congested links which is closely linked to identification of critical links, an important subject within the framework of vulnerability analysis.

Two methods of inputting OD predictions are discussed. In the first method an OD summary is calculated first and then assigned to the network, whereas in the second method all OD predictions are assigned to the network individually. Method 1 leads to approximate-network inference and despite the advantage of being computationally less demanding it is not as exact as method 2. In general, method 2 is promoted and advocated as it provides a suitable tool for full-network inference regarding point and interval estimates, link flow distributions and identification of congested links by means of probability estimates.

The methods were implemented on the Flemish road network for traffic concerning going-to-work/school trips made by Flemish residents between the peak hour from 7 am to 8 am. In general, traffic flows in Flanders were found to be denser around the major municipal centers of Antwerp, Ghent, Leuven and Bruges and on the highways which connect these cities with each other and also with Brussels. Eleven congested links were identified for a V/C threshold value of 0.95, the majority of which belonging to Antwerp and Ghent.

With respect to traffic assignment inference under demand uncertainty the comparison between the two methods provided some evidence that method 1 might be suitable when the sole goal is to have some point estimate of the expected state of a network, such as the expected total system travel time or central point estimates of link flows. Nevertheless, it is acknowledged that similar research must be conducted for networks of different topology with varying levels of congestion in order to properly support this statement. Regarding percentile estimates, it is found that estimates from method 1 result in interval estimates which are clearly narrower and thus fail to capture the true variability of link flows given travel-demand uncertainty.

With respect to congested/critical link identification, the use of method 2 has a potential to provide a robust probabilistic basis for network congestion inference and also for critical link identification and vulnerability analysis. Of course, implementing method 2 in conjunction with a full-network scan approach, as employed for instance in Jenelius et al. (2006), will be most probably prohibitive in terms of computational demand, for cases of large-scale, congested networks. Nevertheless, such a strategy might prove to be viable in combination with reduced-scan approaches as the one presented in Chen et al. (2012), for instance. From a simpler perspective, congested link identification by means of probability estimates may provide a reliable framework for pre-selecting candidate links (e.g. D'Este and Taylor, (2003), Taylor and D'Este (2004)) based on probability rankings. It should be noted that in the applications presented here inference was based on the simple V/C measure, nevertheless it is also straightforward to obtain probability estimates for other related measures, for example the various V/C variants discussed in Knoop et al. (2007).

Finally, uncertainty originating from the choice of the statistical model was also investigated up to a certain degree by utilizing predictions from both the PG and PIG models in separate assignment runs. Results revealed that the choice of the statistical model does have certain influence especially concerning inference for aggregated link flow distributions. That is due to the fact that the two models resulted in different total-demand distributions and since total-demand is influential to traffic assignment the resulting link flow distributions on aggregated levels were also different. Nevertheless, main inferences concerning V/C ratios and the behavior of TSTT given Jensen's inequality were not affected. Concerning disaggregated V/C distributions, inference was affected even to a lesser degree. The same congested links were identified under the predictions of both models and in addition estimates of expected values and probabilities were close.

5 Further insights

In this chapter, some further issues related to traffic assignment and OD modeling are investigated. In the first part of this chapter, the results from deterministic UE are now compared to results obtained under stochastic UE. In correspondence with the previous chapter, the comparison is implemented within the context of demand-uncertainty. The material of this section is summarized in Perrakis et al. (2012f). In the second part of this chapter, a recent and novel modeling approach known as the *radiation model* (Simini et al., 2012) is discussed. The purpose of this section is to find potential links between the radiation model and the statistical modeling approaches presented in this dissertation.

5.1 Uncertainty from the choice of assignment model

In this section a further source of uncertainty is investigated, namely the uncertainty originating from the choice of the assignment model. All of the results presented in the previous chapter were conditional on the DUE assignment operator. The results presented in this section are comparisons between the DUE and the stochastic user equilibrium (SUE) model based primarily on inputting method 2 and also on PG predictions.

5.1.1 SUE assignment as an alternative to DUE assignment

The SUE model is an extension of DUE where the fundamental principle of Wardorp (1952) which assumes that trip-makers have perfect knowledge of the expected travel costs on the network is relaxed. This is expressed by the inclusion of a random error term in the utility function of the route-choice model. In this way, expected travel costs in DUE become *perceived* travel costs in SUE under which the aim of trip-makers is to minimize the perceived travel costs. Thus, SUE flows are such that travelers cannot further reduce their perceived travel times by unilaterally changing routes. DUE assignment is in fact the limiting case of SUE; when the variance in the error distribution is zero, then there is no perception-error and the assignment model becomes the classical deterministic model of Wardrop.

The choice for the error distribution leads to two different formulations of SUE in the related literature. Daganzo and Sheffi (1977) assumed normally distributed errors which led to the probit based route-choice model. Dial (1971) proposed a Gumbel error distribution which in turn led to the logit based route-choice model. Each of the approaches has certain advantages and disadvantages. The logit route-choice model results in a closed form expression with an equivalent mathematical programming formulation (Fisk, 1980). The disadvantages are that one cannot introduce correlations among routes and also account for dependencies between link flows and travel times (Sheffi, 1985). Probit route-choice modeling allows for increased flexibility, since covariances between routes can be explicitly specified when using the normal distribution. Nevertheless, mathematically it is less consistent as it is not possible to formulate a corresponding mathematical program and thus prove convergence to a user equilibrium state. Therefore, tests of “convergence” for the various solution algorithms of probit route-choice modeling are based on heuristic stopping rules and numerical comparisons (e.g. LeBlanc et al., 1975; Sheffi and Powell, 1981). Finally, it should be noted that SUE assignment – regardless of the route-choice model – is stochastic in formulation, but has a deterministic solution which is based on a large-sample approximation (Hazelton, 1998). That means that repeated runs of SUE assignment will deliver an expected vector of link volumes which will always be the same for the same OD input[45].

The SUE model is routinely used as an alternative to DUE both in scientific research as well as in practice. In addition, DUE and SUE assignment are available as options in most of the transportation planning software packages. Given the broad acceptance and use of both models, it is rather odd that comparative studies are scarce in the related literature. A first comparative study between the two models is presented in Sheffi and Powell (1981) mainly within the context of introducing the method of *successive averages* under probit based route-choice modeling. As the interest in this study was focused primarily in comparing the successive average algorithm to other competing algorithms, the comparison was made on a small experimental network of 12 zones and 34 links in order to obtain reliable numerical results. Nevertheless,

[45] Interested readers are further referred to Hazelton (1998) who extends the probit-based SUE model of Daganzo and Sheffi (1977) to a purely stochastic model called conditional SUE (CSUE). Nevertheless, this model is not available in commercial software yet.

the comparison resulted in an interesting conclusion which is that SUE and DUE assignments yield similar results when congestion is high overall, i.e. for large values of total travel-demand with respect to network capacity. Intuitively the result is logical; since for high congestion levels one can argue that the equilibrium effects are stronger than the effects from inaccurate travel-time perception (for a detailed discussion see Sheffi, 1985, section 12.3).

A second, more recent study is that of Ji and Chen (2003). In this study, the comparison is not only between the traditional DUE and SUE models, discussed previously, but also with respect to modern variations of the two models which also take into account inherent network uncertainty concerning variability of link travel times[46]. SUE assignment in this paper is based on probit route-choice modeling. The network under consideration is once again relatively small, consisting of 24 zones and 76 links. The authors find significant differences between the traditional deterministic/stochastic UE models and the variations of these models which account for supply-side travel-time variability. With respect to differences between the traditional DUE and SUE models, the study simply verifies the finding of Sheffi and Powell (1981) and marginally adds the finding that total travel-time is greater in SUE assignment probably due to perception error. Although as the authors comment, the latter finding is not warranted for generalization.

Finally, Zhang (2011) adopts a more critical point of view from the perspective of behavioral realism and compares the traditional DUE and SUE models with emerging and developing variations of UE-based models, namely with bounded rational UE (BRUE) and behavioral UE (BUE) models[47]. This study concentrates on a real-world case study for about 600,000 travelers on a network of 7,976 zones with 20,194 links. In addition, it is the first study which includes a sensitivity analysis concerning the variance of the perception error which is assumed to be Gumbel distributed, i.e. for SUE assignment under logit route-choice modeling and for values of the Gumbel distribution scale parameter equal to 0.1, 0.2 and 1. With respect to this point, the author finds UE convergence being slower for larger values of perception-error variance which is

[46] Information on these variations of DUE and SUE can be found in Soroush (1984) and Mirchandani and Soroush (1987), respectively.

[47] Interested readers are referred to Mahmassani et al. (1986) and Mahmassani and Chang (1987) for BRUE, and to Zhang (2007) for BUE.

intuitively consistent, since a larger error-perception variance implies a wider range of perceived alternative routes. Concerning the comparison between DUE and SUE models, the study illustrates potential discrepancies between the resulting link volumes of each model. Nevertheless, according to the author similar research should be conducted on other networks in order to obtain concrete results. Finally, an interesting finding of this study is that SUE assignment in comparison to DUE assignment results in a V/C distribution which assigns less distributional mass to small and extremely high V/C ratios and more distributional mass to moderate-to-high V/C ratios.

In what follows, we present a comparative analysis between the traditional DUE and SUE models which takes into account both logit and probit route-choice modeling for three different values of perception-error variance. In addition the method of repeated assignments is used (method 2 – section 4.4.2) which provides full information – given demand uncertainty – in terms of point, interval and distribution estimates. The estimates are based on the 500 OD predictions from the PG model.

5.1.2 Formulation and performance of DUE and SUE assignment

As in section 4.5 the BPR formulation (Bureau of Public Roads, 1964) is used as link performance function. The assignments are implemented in TransCAD software version 4.7 (Caliper Corporation, 2008) which utilizes the adaptation of Frank and Wolfe algorithm (Frank and Wolfe, 1956) by LeBlanc et al. (1975) for DUE assignment and the method of successive averages by Sheffi and Powell (1981) for SUE assignment. The selected values for the variance under probit/logit route-choice modeling (normal/Gumbel errors) are 0.01, 0.05 and 0.1[48]. The values are chosen to be relatively small in purpose. As mentioned previously, DUE assignment is the limiting case of SUE assignment as the variance tends to zero. Therefore, we would expect results from probit/logit SUE assignment to be more similar to DUE assignment for the value of 0.01. In addition this value is also utilized by Ji and Chen (2003) for probit SUE modeling. Sheffi and Powell (1981) on the other hand uses a slightly larger

[48] It is useful to comment that the corresponding field for SUE assignment in TransCAD refers to standard deviation. This inquiry was clarified by Caliper Corporation. The corresponding standard deviations are 0.1, 0.2236 and 0.3162, respectively.

value for the variance of the normal distribution equal to 0.3, while Zhang (2011) uses the values of 0.1, 0.2 and 1 for the *scale* parameter of the Gumbel distribution[49]. Up to our best knowledge given the almost complete lack of guidelines in the relative literature concerning “appropriate” variance values, there is no particular evidence for objecting to the values which are adopted here.

Despite the size of the network, all seven assignment models – including the DUE model – converge relatively fast to an equilibrium state. That might be attributed to the fact that the network is in general not seriously congested, as shown in section 4.5.2 as well as section 4.6.1. Convergence checking is based on the commonly used relative gap criterion – recommended in the review of Rose et al. (1988) – for the default TransCAD threshold value of 0.01. According to this criterion convergence is achieved at iteration 4 for DUE and 6 for DUE assignment models. The relative gap, relative mean square error (RMSE) and the required running times for assignments based on the mean OD matrix are summarized in Table 5.1.

	Assignment information (at final iteration)			
Assignment model	**Iteration**	**Relative gap**	**RMSE**	**Runtime (sec’s)**
DUE	4	0.004621	8.37	13.178
SUE probit 0.01	6	0.004707	12.33	18.533
SUE probit 0.05	6	0.004722	12.38	18.620
SUE probit 0.1	6	0.004947	12.86	18.388
SUE logit 0.01	6	0.004693	12.47	18.530
SUE logit 0.05	6	0.004692	13.24	18.602
SUE logit 0.1	6	0.004702	13.43	19.544

TABLE 5.1. Relative gap, RMSE and runtime for the seven assignment models at the final iteration based on the mean OD matrix from the PG model.

In general, there are no major differences between the assignment models in terms of convergence performance. DUE assignment is in general faster which is to be expected. One thing that is odd is that runtime for the SUE probit model is slightly lower for an error variance equal to 0.1. In general, higher values of error variance should lead to slower convergence rates, since the greater the

[49] Zhang follows the Gumbel parameterization introduced in Scheffi (1985) where the variance is $(3.14)^2/(6\times(\text{scale})^2)$. Therefore, for scale parameters equal to 0.1, 0.2 and 1 the respective variances are 164.49, 41.12 and 1.65. Thus, for scale equal to 0.1 and 0.2 the corresponding variance becomes very large.

perception error is the more alternative routes are under consideration. This is properly reflected in the SUE logit.

5.1.3 Comparing TSTT and link volumes from DUE and SUE assignment

In section 4.5.3, it was found that TSTT (total vehicle-hours) under DUE assignment is indeed subjected to Jensen's inequality as argued by Waller et al. (2001). Nevertheless, in practical terms the difference between expected TSTT (*E*(TSTT)) and TSTT based on the expected or mean OD matrix (TSTT[*E*(OD)]) was found not to be significant for the Flemish network, at least for the overall low-to-medium estimated congestion for work/school flows between 7 am and 8 am. Now, the corresponding estimates under the various SUE models are presented in Table 5.2. According to results, Jensen's inequality seems to hold, since *E*(TSTT) is greater than TSTT[*E*(OD)] in all cases. But, given the order of magnitude of TSTT, the differences are once again not significant from a practical point of view.

Assignment model	Total System Travel Time (TSTT) in vehicle-hours			
	***E*(TSTT)-TSTT[*E*(OD)]**	**TSTT[*E*(OD)]**	***E*(TSTT)**	**95% TSTT interval**
DUE	120	1,387,009	1,387,129	(1,383,152, 1,391,169)
SUE probit 0.01	242	1,387,571	1,387,813	(1,383,864, 1,391,862)
SUE probit 0.05	497	1,387,449	1,387,946	(1,383,930, 1,391,977)
SUE probit 0.1	423	1,387,860	1,388,283	(1,384,414, 1,392,447)
SUE logit 0.01	187	1,387,766	1,387,953	(1,383,992, 1,392,118)
SUE logit 0.05	268	1,387,996	1,388,264	(1,384,320, 1,392,350)
SUE logit 0.1	341	1,388,239	1,388,580	(1,384,558, 1,392,608)

TABLE 5.2. Total system travel time comparison between inputting method 1 (TSTT[*E*(OD)]) and inputting method 2 (*E*(TSTT)) for seven assignment models.

Perhaps what is more interesting to observe is that the difference seems to increase as the variance of the perception-error increases. That is clearly observed for the case of logit route-choice modeling. For the case of probit route-choice modeling, there is a small decrease in this difference, for an error-variance of 0.1 in comparison to 0.05, nevertheless it is clearly higher to the difference for an error-variance of 0.01. In addition, TSTT – either from expected demand or expected TSTT – under SUE assignment is always greater than TSTT under DUE. That comes as an additional verification to the finding of Ji and Chen (2003) for which the authors are rather cautious in presenting. Furthermore, based on the mean and interval estimates of TSTT in Table 5.2, it is also obvious that an increase in variance leads to an increase of TSTT under both probit and logit route-choice models.

We proceed with a comparison of link flows under the categorization of link segments belonging to highways, main regional roads, small regional roads and local roads. In general, highway links may be considered as high-capacity links, main regional links as medium-to-high-capacity links, small regional links as medium-to-low-capacity links and finally local links as low-capacity links with respective average vehicle capacities approximately equal to 3785, 2498, 2002 and 1350. Kernel distributions and mean link flow estimates are presented in Figures 5.1 to 5.4. Numerical results in Tables 5.3 to 5.6

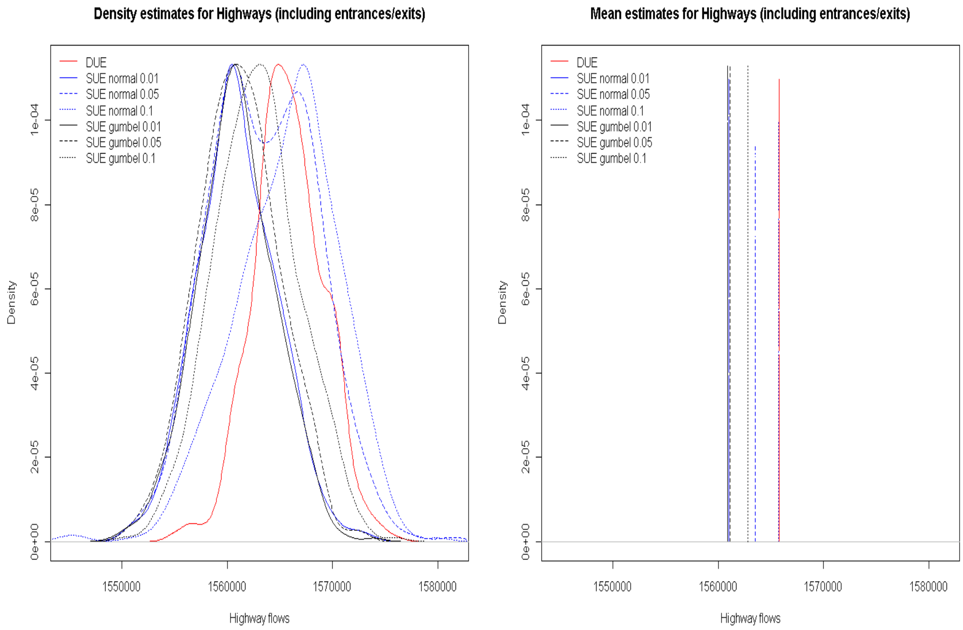

Figure 5.1 Kernel density estimates and mean point estimates for highway flows under DUE assignment (in solid red) and SUE probit/logit assignment (in blue/black) with error variances 0.01 (solid), 0.05 (dashed) and 0.1 (dotted). The estimates are based on repeated assignments of 500 PG predictions.

Figure 5.1 reveals that the highway flow distributions are quite different in location as well as shape. A first observation is that for increased values of variance both probit and logit route-choice models result in an increase of highway flows. This is verified by the point and interval estimates of Table 5.3.

Assignment model	Highway Flows		
	Mean	**Median**	**95% interval**
DUE	1,565,784	1,565,614	(1,559,724, 1,572,337)
SUE probit 0.01	1,560,977	1,560,791	(1,554,088, 1,568,489)
SUE probit 0.05	1,563,494	1,563,446	(1,554,691, 1,572,726)
SUE probit 0.1	1,565,713	1,566,430	(1,555,935, 1,574,225)
SUE logit 0.01	1,560,865	1,560,807	(1,553,539, 1,567,978)
SUE logit 0.05	1,561,096	1,561,088	(1,553,822, 1,568,436)
SUE logit 0.1	1,562,804	1,562,729	(1,555,104, 1,570,438)

TABLE 5.3. Means, medians and 95% intervals for highway flows for the seven assignment models.

The SUE probit and logit models for a variance of 0.01 are very similar, but as the variance increases to 0.05 and 0.1 the highway flows under probit and logit models begin to differ more. In general, the distributions from logit route-choice are closer than the distributions from probit route-choice for different values of variance. The behavior of the distribution under probit modeling is interesting to observe; as the variance increases from 0.01 to 0.05 the distribution forms a second smaller peak (bimodality) and as the variance increases even more to 0.1 the probability mass centers around the second peak forming once again a unimodal distribution. Interestingly, the distribution from SUE probit with variance 0.1 is the closest to the DUE distribution which is contrary to the notion that SUE assignment is similar to DUE assignment for small, close-to-zero values of error-perception variance.

Results are more balanced for main regional roads which have a medium-to-high traffic capacity. As illustrated in Figure 5.2, the distributions from DUE, SUE probit with variance 0.01, 0.05 and 0.1, and SUE logit with variance 0.01 are all relatively similar both in shape as well as in location. This time it is logit route-choice modeling which seems to be more sensitive to the value of variance as main regional road flows seem to increase for larger variance-values. This is also reflected in the estimates of Table 5.4. The most distant distribution from DUE assignment is the one under SUE logit with variance 0.1.

The corresponding estimates for small regional roads are presented in Figure 5.3 and Table 5.5. In this case, there are considerable differences between SUE probit/logit assignment and DUE assignment as it is obvious that SUE models result in more small regional link flows. Intuitively, this may be consistent, since in SUE assignment medium-to-low and low capacity links are

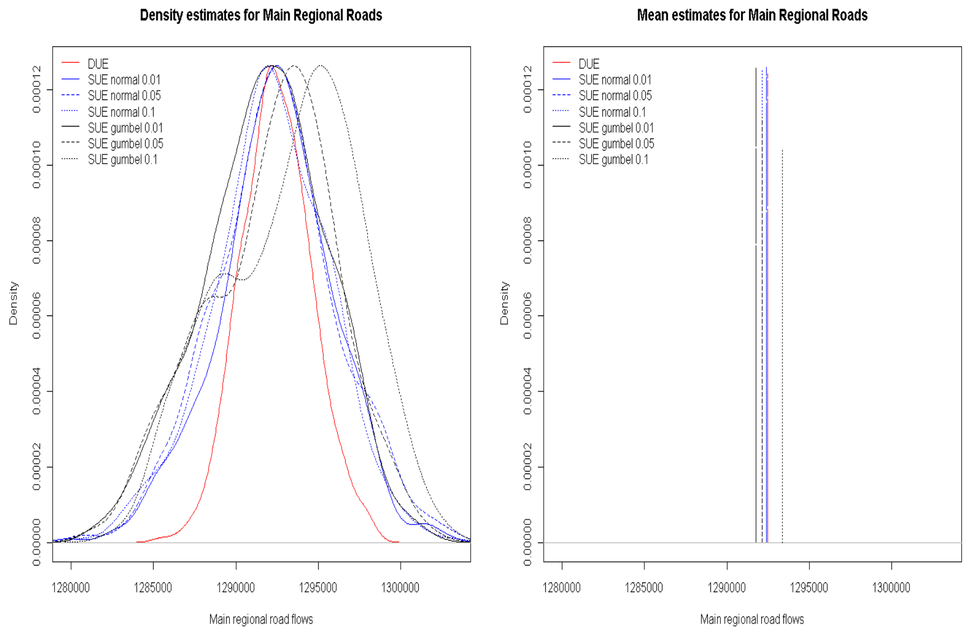

Figure 5.2 Kernel density estimates and mean point estimates for main regional flows under DUE assignment (in solid red) and SUE probit/logit assignment (in blue/black) with error variances 0.01 (solid), 0.05 (dashed) and 0.1 (dotted). The estimates are based on repeated assignments of 500 PG predictions.

Assignment model	**Main Regional Road Flows**		
	Mean	**Median**	**95% interval**
DUE	1,292,461	1,292,363	(1,288,519, 1,296,716)
SUE probit 0.01	1,292,397	1,292,468	(1,285,386, 1,298,633)
SUE probit 0.05	1,292,403	1,292,330	(1,285,260, 1,299,540)
SUE probit 0.1	1,292,135	1,292,085	(1,284,726, 1,298,708)
SUE logit 0.01	1,291,767	1,291,947	(1,284,383, 1,298,239)
SUE logit 0.05	1,292,142	1,292,652	(1,284,215, 1,299,179)
SUE logit 0.1	1,293,362	1,294,050	(1,285,586, 1,300,193)

TABLE 5.4. Means, medians and 95% intervals for main regional road flows for the seven assignment models.

more likely to be used in comparison to DUE assignment. In addition, as with the case of highway flows, SUE logit route-choice modeling seems to be less sensitive to the value of the variance. Although bimodality seems to start formulating for SUE logit with variance 0.05 and 0.1, the logit distributions are closer to each other in terms of location. Contrary, the probit distributions differ more in terms of location and strangely this time the increase of variance leads to a locational shift to the left of the x-axis.

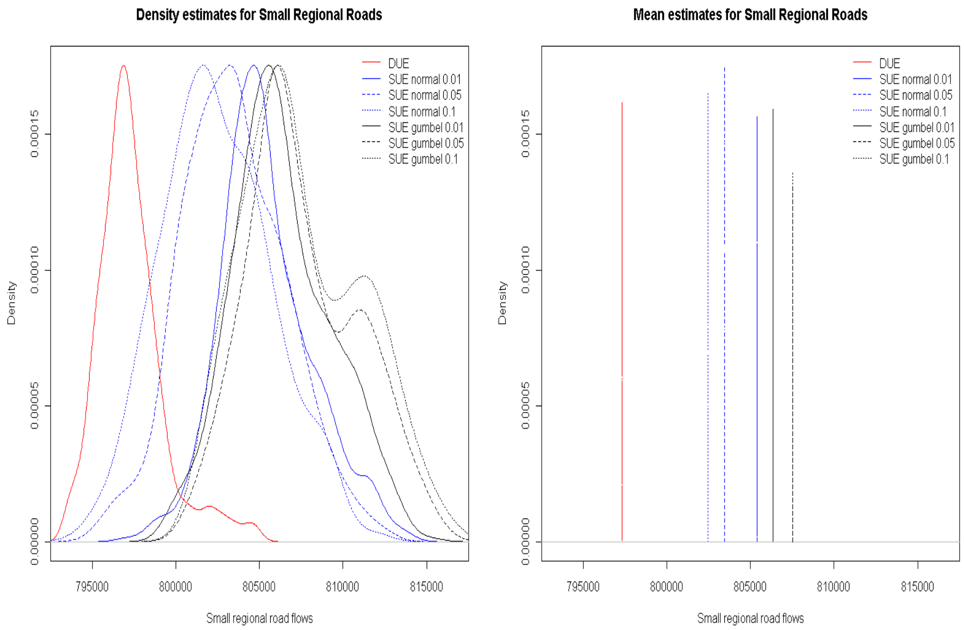

Figure 5.3 Kernel density estimates and mean point estimates for small regional flows under DUE assignment (in solid red) and SUE probit/logit assignment (in blue/black) with error variances 0.01 (solid), 0.05 (dashed) and 0.1 (dotted). The estimates are based on repeated assignments of 500 PG predictions.

Assignment model	**Small Regional Road Flows**		
	Mean	**Median**	**95% interval**
DUE	797,313	797,061	(794,092, 802,817)
SUE probit 0.01	805,387	805,032	(800,515, 811,613)
SUE probit 0.05	803,471	803,380	(797,405, 809,955)
SUE probit 0.1	802,460	802,164	(796,666, 809,158)
SUE logit 0.01	806,362	806,044	(800,911, 812,419)
SUE logit 0.05	807,539	806,855	(801,853, 814,199)
SUE logit 0.1	807,534	806,849	(801,751, 814,627)

TABLE 5.5. Means, medians and 95% intervals for small regional road flows for the seven assignment models.

Estimates for local roads are presented in Figure 5.4 and Table 5.6. In general, one would expect the same phenomenon which occurs on small regional roads to occur also on local roads which are of low capacity.

Nevertheless, this is not the case for local roads. In fact, SUE probit assignment results in less local road flows than DUE assignment, while SUE logit assignment is relatively close to DUE assignment.

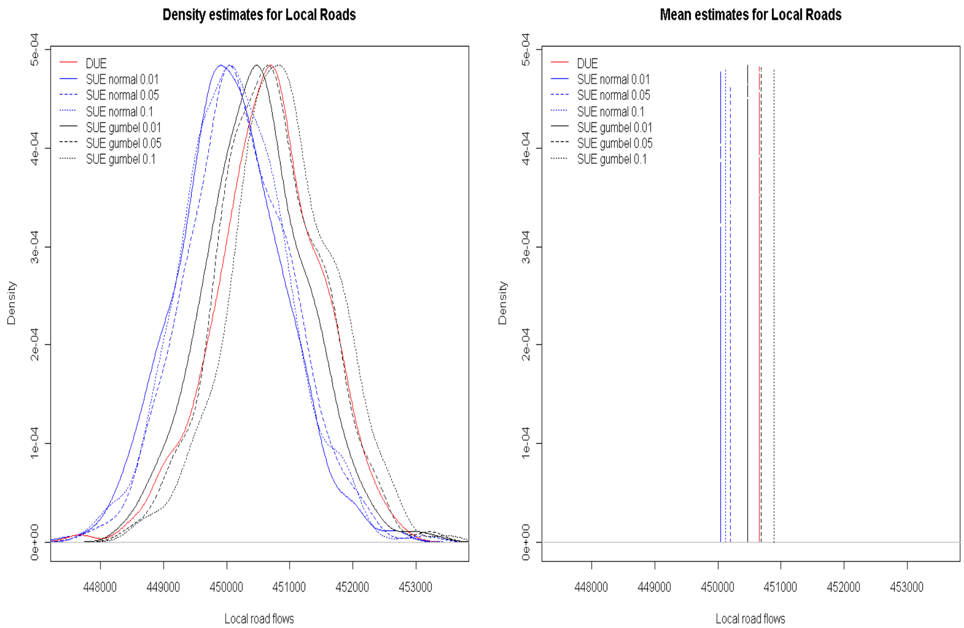

Figure 5.4 Kernel density estimates and mean point estimates for local flows under DUE assignment (in solid red) and SUE probit/logit assignment (in blue/black) with error variances 0.01 (solid), 0.05 (dashed) and 0.1 (dotted). The estimates are based on repeated assignments of 500 PG predictions.

Assignment model	**Local Road Flows**		
	Mean	**Median**	**95% interval**
DUE	450,659	450,704	(448,947, 452,236)
SUE probit 0.01	450,048	450,011	(448,452, 451,806)
SUE probit 0.05	450,197	450,131	(448,672, 451,977)
SUE probit 0.1	450,121	450,090	(448,585, 451,843)
SUE logit 0.01	450,474	450,458	(448,871, 452,118)
SUE logit 0.05	450,686	450,659	(449,108, 452,255)
SUE logit 0.1	450,894	450,872	(449,275, 452,492)

TABLE 5.6. Means, medians and 95% intervals for local road flows for the seven assignment models.

A possible explanation is that convergence is achieved relatively fast, without the need for re-routing on the level of low-capacity links such as local roads.

Possibly, for larger values of perception-error variance SUE assignment will lead to an increase of local road flows similar to what is observed for small regional road flows. Finally, both types of SUE route-choice models are not affected a lot by the different variance values.

In general, based on the numerical outputs presented in Tables 5.3 to 5.6 one can argue that the differences between the various models under consideration are practically not serious. Furthermore, differences occurring on different link types will probably balance out for total aggregated link flows. Results for all link flows, irrespective of link type, are presented in Figure 5.5 and Table 5.7. The distributions in Figure 5.5 under SUE assignment are all to the right of the DUE distribution which is to be expected as SUE assignment produces more traffic due to the perception-error component. As discussed in the beginning of this section, SUE assignment also results in more vehicle-hours travelled in comparison to DUE.

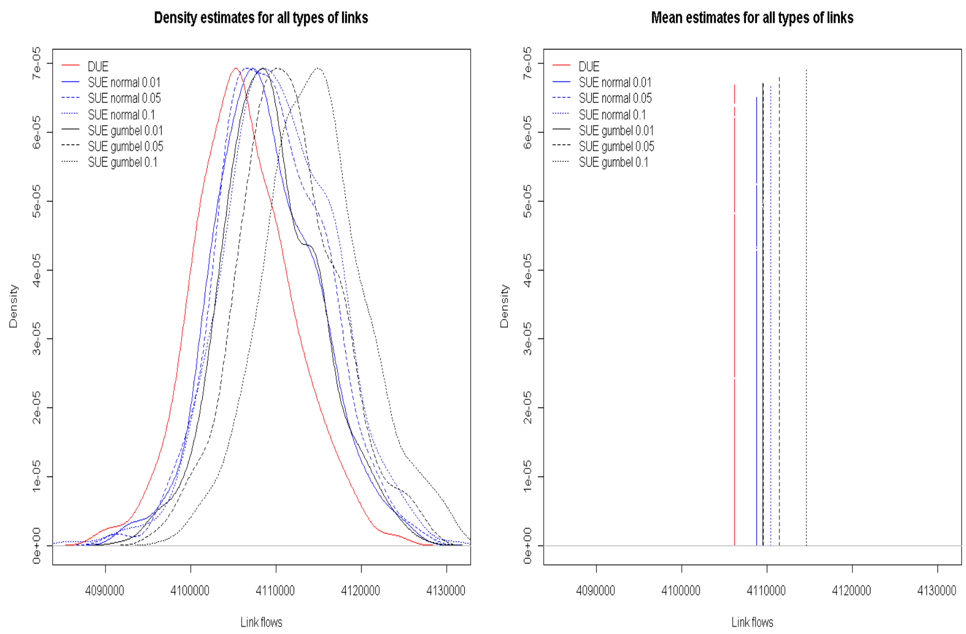

Figure 5.5 Kernel density estimates and mean point estimates for all types of link flows under DUE assignment (in solid red) and SUE probit/logit assignment (in blue/black) with error variances 0.01 (solid), 0.05 (dashed) and 0.1 (dotted). The estimates are based on repeated assignments of 500 PG predictions.

Assignment model	All Flows		
	Mean	Median	95% interval
DUE	4,106,217	4,105,728	(4,095,376, 4,118,223)
SUE probit 0.01	4,108,809	4,108,302	(4,097,431, 4,121,088)
SUE probit 0.05	4,109,564	4,109,286	(4,098,137, 4,121,583)
SUE probit 0.1	4,110,429	4,110,194	(4,098,368, 4,123,068)
SUE logit 0.01	4,109,468	4,109,069	(4,098,547, 4,121,553)
SUE logit 0.05	4,111,463	4,111,086	(4,100,248, 4,124,421)
SUE logit 0.1	4,114,593	4,114,384	(4,103,076, 4,127,616)

TABLE 5.7. Means, medians and 95% intervals for all types of link flows for the seven assignment models.

According to the numerical results of Table 5.7, main inferences are not seriously affected by the choice of the assignment model. The results on total aggregated link flows are also in agreement with theory, since SUE probit/logit assignment with variance equal to 0.01 is closer to DUE assignment than SUE probit/logit assignments with variances equal to 0.05 and 0.1. It is also interesting to observe that SUE probit assignment is generally closer to DUE assignment in comparison to SUE logit for the case of total link flows. That is probably due to the fact that SUE probit flows are more unstable in terms of location with respect to DUE flows for different link types (compare for instance Figures 5.3 and 5.4) and therefore there is a “balancing” effect when aggregating over all link flows.

Argumentatively, it is worth noting that results from DUE and SUE probit/logit assignment can be quite different depending on the link-type under consideration. The differences which were found in highway flows and – especially – in small regional flows are not to be ignored as easily. Finally, with respect to the sensitivity analysis for the perception-error variance, it is worth commenting that SUE assignment under logit route-choice modeling seems to yield more consistent results in comparison to probit route-choice modeling. That is, SUE logit flows tend to increase as the variance increases. That is evident – at least for the variances under consideration – in all the results of this section. On the other hand, such a correspondence is not evident for SUE probit flows.

A final note would include the effect of congestion-level. Nevertheless, given the computational requirements under demand uncertainty (repeated assignments) and also given the fact that all relative studies conclude with results in agreement (Sheffi and Powell, 1981; Ji and Chen, 2003; Zhang, 2011), this particular consideration is not pursued further here. In general, it would be strongly expected that experiments with a deflated/inflated OD demand for the morning peak hour between 7 am to 8 am would result in more/less serious differences between DUE and SUE assignment. That implies also that differences between DUE and SUE assignment would be more apparent for another less-congested, off-peak time interval.

5.1.4 Comparing network congestion from DUE and SUE assignment

A first interesting application concerning network congestion under the different assignment models is to compare the overall distribution of links with respect to the corresponding V/C ratios. Results based on averaged V/C ratios from the 500 V/C predictions are illustrated in Figure 5.6.

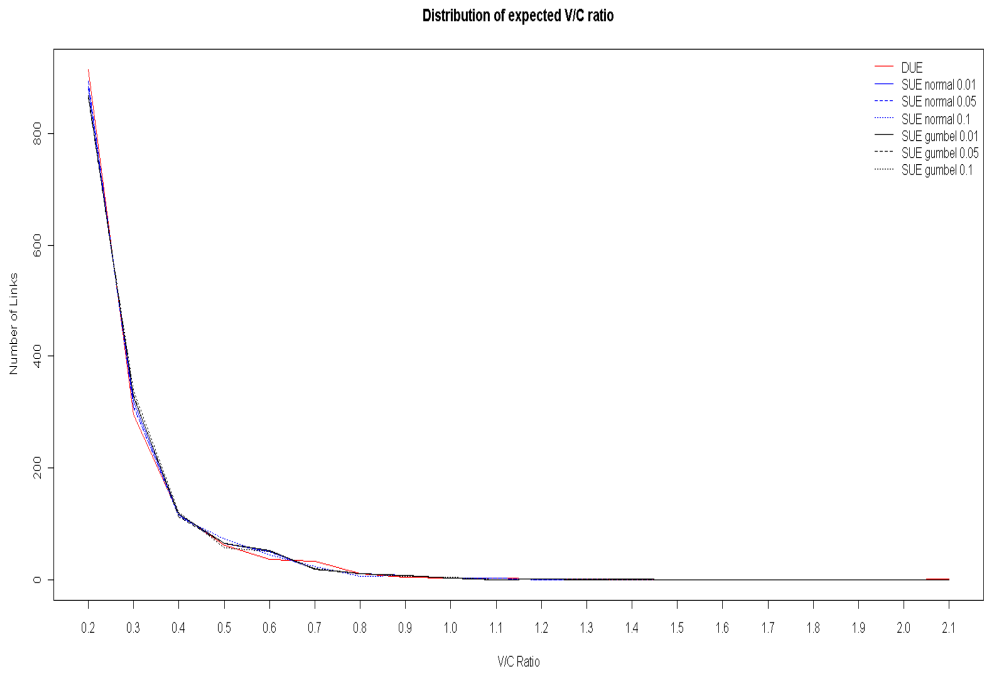

Figure 5.6 Mean V/C ratio distribution for the approximately 5% percent of the links which exceed an expected V/C ratio of 0.2 under DUE and SUE probit/logit assignment and 500 repeated assignments based on PG predictions.

In order to keep the y-axis on a reasonable level, Figure 5.6 shows the expected V/C distribution for the links which exceed an expected V/C ratio of 0.2. These links correspond to only 5% of the total number of links. Nevertheless, for the rest of 95% of the links all assignment models are more or less in agreement, predicting expected V/C ratios between 0 and 0.2. As illustrated in Figure 5.6, the overall congestion level is low and the various assignment models are generally in agreement regarding the distribution of links with respect to their V/C ratio. The exact allocations of links under the different assignment models are presented in Table 5.8.

V/C ratio	Number of links						
	DUE	**SUE probit 0.01**	**SUE probit 0.05**	**SUE probit 0.1**	**SUE logit 0.01**	**SUE logit 0.05**	**SUE logit 0.1**
0.0-0.1	92945	92940	92940	92928	92933	92941	92938
0.1-0.2	3026	3033	3030	3058	3045	3040	3039
0.2-0.3	915	885	894	881	869	867	864
0.3-0.4	295	319	310	311	327	331	338
0.4-0.5	117	115	118	113	117	112	121
0.5-0.6	61	65	65	72	65	65	56
0.6-0.7	36	51	51	44	52	52	51
0.7-0.8	33	19	19	24	19	19	20
0.8-0.9	11	10	10	6	10	10	10
0.9-1.0	4	7	7	7	7	7	6
1.0-1.1	2	3	2	2	3	3	4
1.1-1.2	2	0	2	2	0	0	0
1.2-1.3	1	1	0	0	1	1	1
1.3-1.4	1	0	0	1	0	1	1
1.4-1.5	0	1	1	0	1	0	0
2.1-2.2	1	0	0	0	0	0	0
2.2-2.3	0	1	1	1	1	1	1

TABLE 5.8. Distribution of links with respect to expected V/C ratios for the seven assignment models.

Table 5.8 reveals small alterations in the allocation of links between DUE and SUE assignments for V/C ratios greater than 0.2. In SUE assignment less links have V/C ratios between 0.2 and 0.3 and more links have V/C ratios between 0.3 and 0.4, in comparison to DUE assignment. Also, it is worth noting that this is more obvious for logit route-choice modeling. For V/C ratios ranging from 0.4 to 0.6 there are no major differences, but for V/C ratios between 0.6 and 0.8 differences occur once again; SUE – compared to DUE – results in more links with V/C ratios between 0.6 and 0.7 and in less links with V/C ratios between 0.7 and 0.8. Finally, for V/C ratios greater than 0.8, the allocation of links is more or less the same. V/C ratios between 1.5 and 2.1 are not displayed, since there is no link with a V/C ratio within that interval.

In general, some results are in partial agreement with the results in Zhang (2011), e.g. for the V/C intervals 0.3 – 0.4 and 0.7 – 0.8. Nevertheless, the differences between DUE and SUE assignment presented here are considerably smaller. This might be attributed to the fact that Zhang (2011) experiments with much larger values of perception-error variance (logit route-choice) which probably affect much more the overall distribution of V/C ratio. Also, it should be noted that the distributions in Figure 5.6 and the results in Table 5.8 are derived from expected V/C ratios given demand uncertainty, i.e. averaged over 500 V/C predictions. Given that, it seems that the expected state of the V/C distribution is quite robust with respect to the selection of UE assignment method for relatively small SUE variance. At least, for the particular network and time-interval which are under consideration here.

We proceed now to the issue of congested link identification in terms of V/C probabilities. The resulting V/C distributions from all assignment models for links which exceed a V/C ratio of 0.95 are presented in Figure 5.7. In this case the choice of assignment model seems to have an increased impact on the results. First of all DUE and SUE assignment do not identify congestion on exactly the same links. Two links which are identified through DUE assignment (links 83662, 92846) are not identified through SUE assignment. Vice versa, four links identified through SUE assignment (links 16503, 16607, 18086 and 29809) are not identified th-rough DUE assignment. A reassuring finding is that the same links are identified by SUE models irrespective of the choice of route-choice model and also of the variance value. Thus, SUE assignment leads to 13 congested links while DUE assignment leads to 11. In general, it would be expected that SUE assignment would probably lead to a higher number of congested links, since more links are taken under consideration when allocating traffic through SUE assignment due to perception-error. Given this expectation, one can say that results are not tremendously different in comparison to DUE assignment.

Continuing with similarities between DUE and SUE assignment, it is worth observing that distribution estimates for 3 links are relatively similar, namely for links 22149, 29060 and 106252. In these three cases DUE and SUE assignment models seem to be in agreement. In addition, variations of SUE modeling are not influential at all, since it is hardly possible to distinguish any difference on the corresponding SUE distributions for different route-choice models and for different variances of perception-error. At present, it is difficult to answer with certainty the question why these distributions are similar while other V/C distributions are dissimilar. A first indication is that the three links belong to the same link-type, i.e. they are local road links. As discussed in the previous section, local road flows are perhaps the least affected flows by the choice of assignment model, which is probably due to the low values of perception-error variance used in the SUE models. On the other hand, based on Table 5.9, link 17493 is also a local road segment which is identified by all 7 assignment models, but in this case the V/C distributions are very dissimilar. Therefore, the

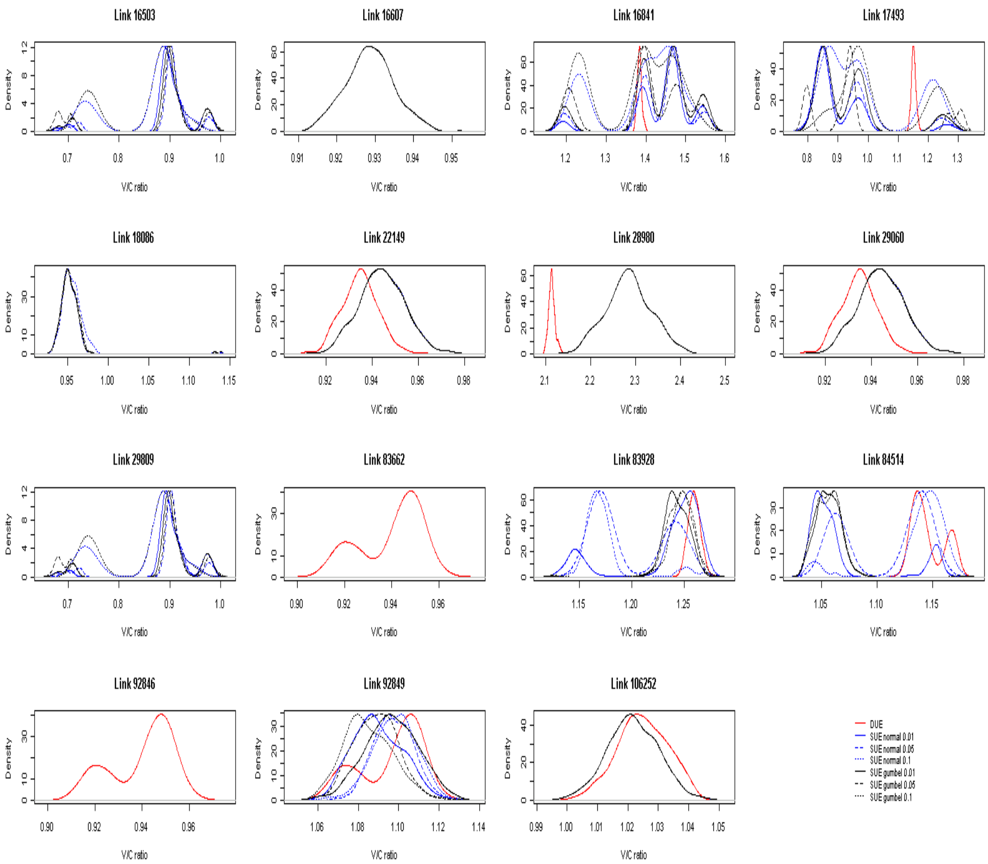

Figure 5.7 Kernel estimates of V/C distributions of congested links which have a non-zero probability of exceeding a V/C ration of 0.95. The distributions are derived based on repeated assignment of 500 PG predictions and conditional on DUE (red), probit SUE with variance 0.01 (solid blue), 0.05 (dashed blue), 0.1 (dotted blue) and logit SUE with variance 0.01 (solid black), 0.05 (dashed black) and 0.1 (dotted black).

link type by itself does not provide a full explanation regarding the similarities for links 22149, 29060 and 106252. The only intuitive supplementary explanation is that these three links are close to significant network centers with respect to traffic allocation. According to the brief locational analysis of section 4.5.4, this seems to be valid. Links 22149 and 29060 are both local outgoing road segments to highway ring R1, to the west side of Antwerp, and also very near highway E17 which connects Antwerp with Ghent. As for link 106252, it seems to be one of the very few local road links which provides a connection between N31 regional road, to the west of Bruges, and Bruges city center.

Despite the similarities, interesting observations relate more to dissimilarities between DUE and SUE assignment. In particular, the “switching effect” – discussed in section 4.5.4 – is evidently much stronger under SUE assignment as this time bimodalities as well as multimodalities are observed, for instance in the links with id’s 16503, 16841, 17493, 29809, 83928 and 84514. Once again this is intuitively consistent, since the inclusion of the perception error is expected to increase re-routing to links which are not considered under DUE. To this respect, in section 4.5.4 it was argued that the expected V/C value is not an optimum criterion for identifying congested links as it fails to identify links on which congestion is quite probable, but on which the V/C expectation can be lower than the selected threshold V/C value. Now, under SUE assignment it is further argued that the expected V/C ratio is a poor and inadequate criterion for identifying congested links. Almost one third of the congested links under SUE assignment have multimodal V/C distributions and in these cases the expected V/C ratio will not correspond to a likely value at all. The same holds for other point estimates as the median. In fact, the only truly informative point estimates would be the modes and the corresponding mode-probabilities, but such estimates are almost impossible to estimate a-priori, i.e. without knowing the form of the V/C distributions given demand uncertainty. From this perspective probability estimates seem to provide a safe and robust solution for evaluating network congestion given demand uncertainty under different assignment models.

The probability estimates for exceeding a V/C ratio of 0.95 under DUE and probit/logit SUE are presented in Table 5.9. For the 9 links which are commonly identified under both DUE and SUE the corresponding probabilities are generally in agreement. One exception where the probabilities vary more is link 17493. In addition, for the remaining 4 links which are identified only through SUE assignment, the differences in the probabilities between probit/logit route-choice modeling and for different variances are again not great.

Congested Link ID	Link type	Probability of exceeding a V/C ratio of 0.95						
		DUE	SUE probit 0.01	SUE probit 0.05	SUE probit 0.1	SUE logit 0.01	SUE logit 0.05	SUE logit 0.1
16503	HW	-	0.010	0.008	0.006	0.010	0.006	0.002
16607	SRR	-	0.002	0.002	0.002	0.002	0.002	0.002
16841	SRR	1.000	1.000	1.000	1.000	1.000	1.000	1.000
17493	LR	1.000	0.268	0.368	0.574	0.426	0.690	0.810
18086	LR	-	0.592	0.588	0.650	0.590	0.586	0.588
22149	LR	0.046	0.080	0.080	0.080	0.080	0.080	0.080
28980	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
29060	LR	0.046	0.080	0.080	0.080	0.080	0.080	0.080
29809	HW	-	0.010	0.008	0.006	0.010	0.006	0.002
83662	LR	0.236	-	-	-	-	-	-
83928	HW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
84514	MRR	1.000	1.000	1.000	1.000	1.000	1.000	1.000
92846	LR	0.236	-	-	-	-	-	-
92849	LR	1.000	1.000	1.000	1.000	1.000	1.000	1.000
106252	LR	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 5.9. Estimates of probabilities of exceeding a V/C ratio of 0.95 for 11 congested links under DUE assignment and 13 congested links under probit/logit SUE assignment with differing perception-error variance. The estimates are based on repeated assignments of 500 OD predictions from the PG model. The abbreviations HW, MRR, SRR and LR stand for highway, main regional roads, small regional roads and local roads, respectively.

5.2 The radiation model

In the recent and very interesting study of Simini et al. (2012) a new approach is suggested for modeling of mobility and migration patterns. The proposed model is in fact a generalization of the traditional gravity model which provides remedies for certain disadvantages of the latter. Namely, i) lack of rigorous derivation, ii) uncertainty for the use of proper deterrence functions and parameter values, iii) the need for existing traffic data, iv) obvious predictive discrepancies, v) asymptotic inconsistency for large destination population and vi) deterministic limitations. The proposed model is based on radiation emission and absorption processes used in physical sciences and it is thus referred to as the *radiation model*.

An analytical description of the mathematical derivation of the radiation model would be out of the scope of this section, interested readers are referred to the original paper by Simini et al. (2012). In what follows attention is restricted to the concluding assumptions of the model. In order to remind the matrix notation, the OD flows are denoted by T_{od} where $o, d = 1, 2, \ldots m$ and m is the number of zones. The radiation model results in a binomial distribution for the OD flows, that is $T_{od} \sim Bin(T_o, p_{od})$ for $o, d = 1, 2, \ldots m$ and $o \neq d$. According to the radiation model the probabilities p_{od} of the binomial distribution are given by

$$p_{od} = \frac{m_o n_d}{(m_o + s_{od})(m_o + n_d + s_{od})},$$

where m_o is the population of origin *o*, n_d is the population of destination *d* and variable s_{od} is the population within the circle defined by the radius of distance between origin *o* and destination *d*, excluding the populations of *o* and *d*. The binomial probabilities are defined so that $\sum_{d \neq o} p_{od} = 1$. It should also be noted that the radiation model applies to inter-zonal flows, i.e. when $o \neq d$. The variable T_o is essentially the total number of travelers leaving from origin *o*, i.e. $T_o = \sum_{d \neq o} T_{od}$, which is also proportional to the population of origin zone and thus $T_o = m_o(N_T / N)$, where N_T is the total number of travelers and N is the total population of the geographical area under consideration. Given the above, the radiation model provides estimates for the expectation and variance of the OD flows given by $E(T_{od}) = T_o p_{od}$ and $\mathrm{var}(T_{od}) = T_o p_{od}(1 - p_{od})$, respectively.

The significant advantages of the radiation model in comparison to the gravity model or even the statistical models presented in this dissertation are; first it does not require estimation as it is parameter-free and second it does not require traffic data of any form. It should also be noted that it has a natural interpretation in terms of the binomial distribution which implies that on disaggregated levels the probability of traveling from *o* to *d* for each individual traveler is given by the Bernoulli distribution with success probability p_{od}. The following sections explore possibilities of borrowing strength from the radiation model and even adapting it in order to assimilate it into the modeling framework which is under consideration in this dissertation.

5.2.1 Circular area population as explanatory variable

The circular area population variable s_{od} is one of the main differentiating elements of the radiation model with respect to the gravity model. Given the strong impact of the "radiation law" and within the context of our modeling approach, it is interesting to assess the result of including this variable in the negative binomial model[50]. One would expect that this variable would be highly-significant and that inclusion of it will potentially constitute other explanatory variables redundant, thus resulting in a simpler model with fewer variables and consequently fewer parameters.

In order to remove potential influence from other explanatory variables and compare distance – the fundamental variable of the gravity model – with circular area population, two simple negative models with logarithmic link functions are initially fitted; one with intercept and the logarithm of distance and one with intercept and the logarithm of circular area population. Results under ML estimation are presented in Table 5.10[51]. According to these results it seems that as a stand-alone variable distance is more influential, since the model with distance has a lower value of AIC. The same models were also fitted with the explanatory variables on natural scale and/or with the identity function instead of the logarithmic function as link. In all cases both AIC values were greater than those of Table 5.10, but consistently the model with distance always yielded a lower value of AIC than the model with circular area population. In addition, the log-link model which includes both the logarithms of distance and circular area populations as explanatory variables results in a value of AIC equal to 346932. The value is lower of course but not promising any practical improvement in terms of predictive inference.

[50] This experiment is not necessarily constrained to the negative binomial/PG model. Nevertheless, since this model has been used more or less as a benchmarking model with respect to the PLN and PIG models, we utilize this model for these first results.

[51] ML estimation is adequate for this simple comparison. Based on previous experience from the models presented in chapter 3, ML estimates and posterior means from Bayesian implementation are almost identical. In addition, AIC under ML estimation is usually very close if not exactly equal to the marginal DIC. These results are expected due to large sample size ($n = 94,864$).

Further experiments indicate that including the circular area population variable in the full model of section 3.3 does not contribute significantly in model improvement. The full model with log of circular area population instead of log of distance yields an AIC of 294378 which is substantially larger than 281492 which is the value of AIC for the full model with distance. The extended model with distance and circular area population results in a decreased AIC equal to 281301. Thus, including the log of circular area population in the full negative binomial model presented in section 3.3 decreases AIC just slightly, from a value of 281492 to a value of 281301.

Parameter	**Model**							
	intercept + log*d*				**intercept + log*s***			
	Estimate	**Std.Error**	**z-value**	**p-value**	**Estimate**	**Std.Error**	**z-value**	**p-value**
β_0	11.478	0.04005	286.6	2×10^{-16}	26.543	0.06986	379.9	2×10^{-16}
β_1	-2.887	0.01099	-262.6	2×10^{-16}	-1.756	0.00481	-364.9	2×10^{-16}
AIC	347166				353704			

TABLE 5.10. ML estimates and values of AIC for two simple negative binomial models. In the first model log of distance (log*d*) is used as an explanatory variable, in the second model the log of circular area population (logs) – the variable of the radiation model – is considered.

These first results do not meet with initial expectations regarding the inclusion of the radiation model variable into a regression modeling approach. Of course, the current experiments may be deemed as premature and one can argue that there are many other possible transformations and combinations of the explanatory variables as well as different distributional assumptions inspired by the assumption of the radiation model (e.g. a binomial or beta-binomial regression model) which would potentially provide more encouraging results. On the other hand, as discussed in section 3.2.6, current research is suggestive of the fact that the regression models presented in chapter 3 are practically saturated from explanatory variables. That is, explanatory power in the current models cannot be significantly increased by including additional explanatory variables, or to put it another way it is difficult to find explanatory variables which will explain the variability in OD flows from an aspect which is not already covered by the choice of the existing socio/geo-economical explanatory variables.

Possibly, variables related to psychological characteristics would contribute more in explaining the unaccounted variability. Recent transportation studies focus on psychological aspects of traveling and provide interesting findings regarding two-way interactions between such aspects and the act of travelling, e.g. in Ory (2007). Nevertheless, variables related to psychological aspects are difficult to quantify and even when that is possible such variables have no meaning for OD modeling on aggregated levels.

5.2.2 Extra stochasticity in the radiation model

As described previously the radiation model is a stochastic model which assumes that the OD flows are binomially distributed. Nevertheless, the radiation model is not a statistical model in the sense that there is no estimation procedure involved as the parameters of the binomial distribution are given by known quantities. Of course, as mentioned in section 5.2 that is also the major advantage of the radiation model, i.e. it is a parameter-free model.

On the other hand, it might be worth considering a statistical-extension of the radiation model. An initial advantage from such an approach is that one could benefit from the well-established statistical tradition of goodness-of-fit testing[52]. Another, perhaps more important aspect, is that of prediction. As illustrated throughout this dissertation one of the main advantages of the proposed methodology is the well-defined predictive framework. Assimilating the radiation model into a similar Bayesian framework would allow for predictive testing on disaggregated and aggregated levels. For instance, say that one would like to use the radiation model for inferring on the aggregation levels which are used as examples in sections 3.2.5 and 3.3.6. With the radiation model in its current form it is only possible to calculate the expected values on aggregated levels as summations of expected values on disaggregated levels.

[52] Arguably, since the radiation model is essentially a stochastic model which provides an expected value and a variance value for each OD pair one can still attempt to calculate common goodness-of-fit measures, such as the chi-square and the deviance (likelihood-ratio) measures. Nevertheless, a problem of the chi-square statistic is that it is increasing in analogy with sample size whenever the fitted model is not the “true” model, and thus the null hypothesis tends to be rejected for cases of large datasets. The deviance measure, although slightly more reliable, shares the large-sample problems encountered by the chi-square statistic, which is in fact an approximating statistic to the deviance, while in addition the concepts of a *full* and of a *null* model are difficult to determine in the case of the parameter-free radiation model.

The corresponding observed quantities – without taking into account intra-zonal flows – and the expected quantities from the radiation model are presented in Figure 5.8.

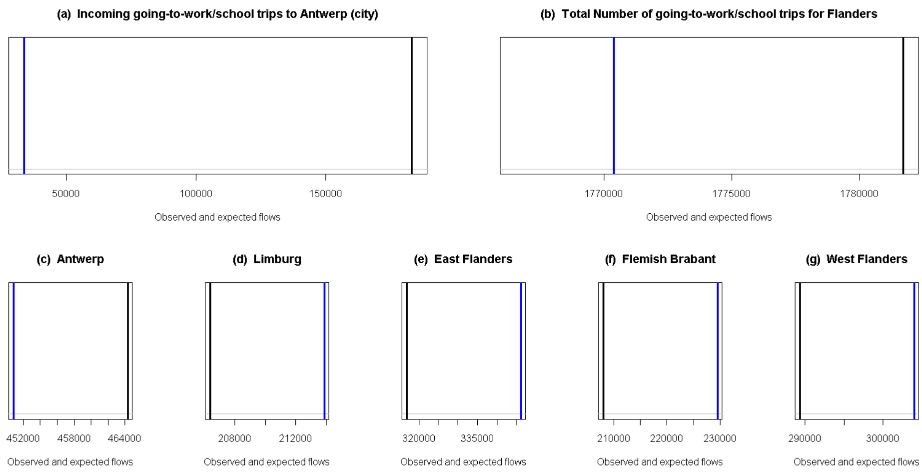

Figure 5.8 Observed quantities and expected quantities of inter-zonal going-to-work/school trips from the radiation model for incoming trips to Antwerp (a), total number of trips in Flanders (b) and internal trips within each of the five Flemish provinces; Antwerp (c), Limburg (d), East Flanders (e), Flemish Brabant (f) and West Flanders (g). The vertical lines correspond to observed quantities, the vertical blue lines to the expected values from the radiation model.

The differences between observations and expectations in graphs (b) to (g) of Figure 5.8 *seem* relatively small, but one can say little regarding the extremeness of the expectations in terms of statistical significance. Contrary, under a statistical, in particular a Bayesian, framework the vertical blue lines would be replaced by predictive distributions allowing for calculation of p-values.

The fact that the radiation model is applicable only to inter-zonal flows – and consequently a statistical extension of it would be also limited to inter-zonal flows – seems to be a minor disadvantage in comparison to the potential benefits. Note that traffic assignment does not take into account intra-zonal flows and therefore a Bayesian version of the radiation model could also be utilized for all the applications described in chapter 4 and section 5.1 of this chapter. Finally, if intra-zonal flows are really the focus of interest, then one can still obtain rough estimates of intra-zonal flows calculated as $E(T_{oo}) = m_o - \sum_{d \neq o} E(T_{od})$ for $d = 1, 2, \ldots, m$.

Given these considerations, a statistical extension of the radiation model might be worth investigating. In what follows an initial Bayesian representation of the radiation model is proposed. It is argued that in order to keep the initial form of the radiation model unchanged, an informative Bayesian approach must be adopted.

5.2.3 An informative Bayesian representation of the radiation model

Let us start from the concluding assumption of the radiation model which is $T_{od} \sim Bin(T_o, p_{od})$, so that $E(T_{od}) = T_o p_{od}$ and $\mathrm{var}(T_{od}) = T_o p_{od}(1 - p_{od})$, for $o, d = 1, 2, \ldots m$. From a statistical perspective the binomial distribution is the likelihood distribution, i.e.

$$L(T_{od} \mid p_{od}) = \binom{T_o}{T_{od}} p_{od}{}^{T_{od}} (1 - p_{od})^{T_o - T_{od}} .$$

Under the radiation model we have a deterministic relation of the form

$$p_{od} = \frac{m_o n_d}{(m_o + s_{od})(m_o + n_d + s_{od})} . \quad (5.1)$$

Now let us relax the assumption above and assume a stochastic representation of equation (5.1). In particular, a Bayesian representation which will introduce a prior distribution $\pi(\cdot)$ for p_{od}. A convenient distribution in general is the beta distribution, i.e. $p_{od} \sim Beta(\alpha_{od}, \beta_{od})$. As presented next, the beta distribution is convenient since it is conjugate to the binomial distribution. Furthermore, it is also an appropriate distribution with support in (0,1) and with p.d.f. – momentarily ignoring subscripts – given by $\pi(p) = \mathrm{B}(\alpha, \beta)^{-1} p^{\alpha-1}(1 - p)^{\beta-1}$, where $\alpha, \beta \in \mathbb{R}^+$ and $\mathrm{B}(\cdot)$ is the beta function. Note that the expectation and variance are given by $E(p) = \alpha / (\alpha + \beta)$ and $\mathrm{var}(p) = \alpha\beta / [(\alpha + \beta)^2 (\alpha + \beta + 1)]$, respectively.

Given the strong impact of the “radiation law” one would like to adopt an informative beta prior which will reflect belief in equation (5.1). This can be achieved by assuming that

$$E(p_{od}) = \frac{m_o n_d}{(m_o + s_{od})(m_o + n_d + s_{od})}$$

and that $\text{var}(p_{od})$ is relatively small in comparison to the mean. By expanding the denominator on the right side of equation (5.1) we have that $(m_o + s_{od})(m_o + n_d + s_{od}) = m_o n_d + [(s_{od} + m_o)^2 + s_{od} n_d]$. Thus, by setting $\alpha_{od} = m_o n_d$ and $\beta_{od} = [(s_{od} + m_o)^2 + s_{od} n_d]$ we obtain exactly

$$E(p_{od}) = \frac{\alpha_{od}}{\alpha_{od} + \beta_{od}} = \frac{m_o n_d}{(m_o + s_{od})(m_o + n_d + s_{od})}.$$

Note that in real-world applications where $m_o, n_d, s_{od} >> 1$ the variance $\text{var}(p_{od}) = \alpha_{od}\beta_{od} / [(\alpha_{od} + \beta_{od})^2(\alpha_{od} + \beta_{od} + 1)]$ will be small since the denominator will be greater than the numerator. Thus, for the proposed values of α_{od} and β_{od} we have an informative prior which is highly peaked around equation (5.1) with small variance. In summary, the "Bayesian-radiation" model is of the form

$$T_{od} \mid p_{od} \sim Bin(T_o, p_{od}) \text{ (likelihood)},$$
$$p_{od} \sim Beta(\alpha_{od}, \beta_{od}) \text{ (prior)},$$

with $\alpha_{od} = m_o n_d$ and $\beta_{od} = [(s_{od} + m_o)^2 + s_{od} n_d]$. It is interesting to note that integrating over p_{od} results in a marginal likelihood or prior predictive distribution which is a beta-binomial distribution with parameters T_o, α_{od} and β_{od}, that is

$$\begin{aligned} L(T_{od}) &= \int_0^1 L(T_{od} \mid p_{od}) \pi(p_{od}) dp_{od} \\ &= \binom{T_o}{T_{od}} [\mathrm{B}(\alpha_{od}, \beta_{od})]^{-1} \int_0^1 p_{od}^{T_{od} + \alpha_{od} - 1} (1 - p_{od})^{T_o - T_{od} + \beta_{od} - 1} dp_{od} \\ &= \binom{T_o}{T_{od}} \frac{\mathrm{B}(T_{od} + \alpha_{od}, T_o - T_{od} + \beta_{od})}{\mathrm{B}(\alpha_{od}, \beta_{od})}. \end{aligned}$$

As discussed in section 2.1 the prior predictive distribution has a rather awkward interpretation in Bayesian statistics as it gives the probability of witnessing the data that were actually manifested, calculated before any data became available. Nevertheless, it is very useful for formal model comparison based on Bayes factors. What is more interesting is that due to the conjugate beta prior the posterior distribution resulting from $p(p_{od} \mid T_{od}) = L(T_{od} \mid p_{od}) \pi(p_{od}) / L(T_{od})$ is also of known form, namely $p_{od} \mid T_{od} \sim Beta(a_{od} + T_{od}, T_o - T_{od} + \beta_{od})$ with posterior expectation and variance given by

$$E(p_{od} \mid T_{od}) = \frac{a_{od} + T_{od}}{T_o + a_{od} + \beta_{od}} \quad (5.2) \text{ and}$$

$$\text{var}(p_{od} \mid T_{od}) = \frac{(a_{od} + T_{od})(T_o - T_{od} + \beta_{od})}{(T_o + a_{od} + \beta_{od})^2 (T_o + a_{od} + \beta_{od} + 1)} \quad (5.3),$$

respectively. At present, issues related to the added value or even to the validity of the aforementioned Bayesian “version” of the radiation model should be discussed and contemplated.

Starting from the disadvantages of this approach, it must be mentioned first that the radiation model loses its desired generality as it becomes a statistical model which depends on observed data for the estimation of parameters p_{od} for $o, d = 1, 2, \ldots m$. This argument can be partially bypassed, since under specific conditions – as the ones that prevail in transportation modeling – OD modeling is not the main inferential purpose *per se*. Under such conditions where inference for disaggregated as well as aggregated levels is of importance, accurate predictions on all levels are needed and therefore loss in generality can be partially allowed in exchange for predictive power. On the other hand there also is an issue with the statistical aspect of this model as the number of parameters is equal to the number of data points. Given this, the proposed model is meaningful only under a truly *informative* Bayesian approach. Otherwise, the adoption of a vague prior would probably result in posterior results that fit the data perfectly, but the related posterior inference would have a doubtful scientific value.

The potential added-value of this approach is that one borrows strength from prior beliefs (radiation assumption) as well as observed data. In cases where the observed OD flows are close to the radiation expectation (equation 5.1) results from the radiation model and the Bayesian formulation will not differ a lot. On the other hand, in cases where observed flows are far from the radiation expectation then the posterior expectation in equation 5.2 will be somewhere in-between observed data and prior assumptions, thus closer to the former. Another advantage is that when new data become available, then the Bayesian model provides a natural updating framework, i.e. the posterior distribution can be utilized as a prior distribution for analyzing data from a new census study. Finally, it is very easy to generate properly defined short-term predictions of OD flows and compare these predictions with the observed OD flows.

6 Discussion

In this final chapter conclusions, considerations and future directions of research are discussed. Initially, a summary of the main results is provided followed by a section in which results are viewed and interpreted from the perspective of input and model-uncertainty. Considerations concerning the issue of applicability of this current research as well as certain econometrical considerations are discussed next. The chapter concludes with a summary of possible future research directions.

6.1 Summary of main results and contributions

In chapter 3 statistical OD modeling with covariates is introduced. The proposed methodology is applied on matrices derived from the 2001 Belgian census study and inference is focused on OD movements between the 308 Flemish municipalities. In general, the approach may be seen as a framework which incorporates the trip-generation and trip-distribution steps of the four-step model into statistical models that provide richer parametrical and predictive inference. At the same time, the particular experimental design allows for an alternative interpretation of the models from a direct-demand, gravity modeling perspective.

The starting assumption was a Poisson model where expected OD flows are related to a set of 25 explanatory variables through a log-linear link function. A first finding was that the simple Poisson model – a modeling assumption which is commonly adopted for OD flows – is clearly not appropriate when modeling large matrices with a great degree of overdisperion like the ones under consideration here. Contrary, model comparison based on information criteria provided much support in favor of the negative binomial or PG model indicating that Poisson-mixture modeling is a more suitable starting assumption. The PG model was further compared to the PLN model and also to the PIG model. In this first Bayesian application of the latter, it is shown that the model has desired and distributional properties which constitute its use particularly convenient, especially in comparison to the PLN model. Furthermore, the PIG model provided the best marginal fit among all three models.

With respect to parameter-significance, all parameters proved to be statistically significant within the Poisson-mixture framework. The parameter estimates from the PIG and PLN models – which both provided a better marginal fit than the PG model - are in general more similar. Concerning parameter-interpretation, the related inference led to consistent results, which are mainly in accordance to traditional trip-production and trip-attraction studies as well as recent transportation studies. In addition, predictions generated from the hierarchical structure of the PG and PIG models were compared to observed data according to several predictive overall goodness-of-fit and case-specific tests. The overall fit is found to be satisfactory for both models. Nonetheless, one important finding is that both models tend to underestimate the number of zero-valued cells and as shown through further analysis zero-valued cells eventually have a strong cumulative influence on the figure of total travel-demand.

In chapter 4, two methods of inputting OD predictions in traffic assignment models are demonstrated. Method 1 is based on individual assignments of OD summaries and leads to approximate-network inference, whereas method 2 is based on repeated assignment of OD predictions and results in full-network inference given demand uncertainty. The methods are tested on the Flemish road network for the morning peak hour between 7 and 8 am. Implementation under DUE assignment is based initially on PG predictions and subsequently on PIG predictions.

The first comparison between methods indicated that centrality estimates derived from the two methods such as means and medians are relatively close. Regarding TSTT and Jensen's inequality, the TSTT based on mean demand (method 1) was found indeed to be smaller than the mean TSTT (method 2). Nevertheless, in practical terms the two estimates are relatively close. Method 1 as an approximate method is computationally less demanding and therefore these initial findings provide some support in favor of its use when the goal is to have an estimate of the expected state of a network. Of course, as already stated, it is important to conduct similar research for networks of different topology and with varying levels of congestion. Nevertheless, method 2 is in general advocated, despite being computationally demanding. Results based on method 2 provided more reliable percentile estimates and also dispersion estimates as well as probability estimates which cannot be calculated through method 1.

The second comparison focusing on traffic assignment results from PG and PIG predictions revealed that the choice of the statistical model can have a significant impact on aggregated link flows. In our case, the underlying reason is that the two models resulted in different predictions of total travel-demand which are highly influential to traffic assignment results on aggregated levels. On the other hand, main inferences concerning disaggregated link volumes and V/C ratios as well as TSTT were not affected by the selection between PG and PIG Gaussian predictions.

In general, traffic flows in Flanders were found to be denser around the major municipal centers of Antwerp, Ghent, Leuven and Bruges and on the highways which connect these cities with each other and also with Brussels metropolitan area. Congestion analysis indicated that eleven links have a non-zero probability of exceeding a V/C threshold ratio of 0.95. As shown, identification of congested links under demand-uncertainty is safer through probability estimates in comparison to mean estimates. A further interesting finding is that V/C distributions are not necessarily normally distributed as bimodalities may often appear.

In the first part of chapter 5 further traffic assignment experiments were implemented. In particular, DUE assignment was compared to SUE assignment with both probit and logit route-choice modeling and for different values of perception-error variance, specifically for variances equal to 0.01, 0.05 and 0.1. This time all traffic assignments were based on PG predictions. With respect to TSTT, results once again indicated that under SUE assignment mean TSTT is greater than TSTT based on mean demand, in accordance with Jensen's inequality. In addition, the differences are increasing as perception-error variance is increasing. Nevertheless, in practical terms the differences are not particularly alarming.

Results for aggregated link volumes are less straightforward to interpret. Certainly the choice of assignment model induced some variability which is apparent when examining link volume distributions as well as point and interval estimates. Nevertheless, the degree of this induced variability seems to depend on the link-type category under consideration. Based on numerical results differences for local roads (low capacity) are almost negligible, differences for main regional roads (medium-to-high capacity) and especially highways (high-capacity) are more noticeable, while differences for small regional roads (medium-to-low capacity) are more than obvious. Despite these diverse results some general conclusions are in agreement with theoretical expectations. First, DUE assignment assigned more traffic to high-capacity links, while SUE assignment tended to assign more traffic to medium-capacity links. Second, when considering the total amount of traffic, then SUE assignment always produced more traffic than DUE and in addition traffic under SUE increased with increases in error-perception variance.

Concerning congestion analysis, results revealed that the selection of assignment model does not seriously affect the overall allocation of links with respect to expected V/C ratios. On the other hand, variability is induced once again when examining individual V/C distributions of specific links. In particular, examination of V/C distribution for links which exceed a V/C ratio of 0.95 led to the conclusion that V/C distributions can differ significantly. In addition, under SUE assignment bimodalities as well as multimodalities were observed. These results provide additional support to the use of probability estimates based on V/C distributions as opposed to centrality estimates as measures of identifying congestion. In general, congestion was identified on 13 links under SUE assignment with corresponding probability estimates which did not vary a lot for 12 out of the 13 links. The same holds for 8 out of the 9 links which were mutually identified under DUE and SUE assignment.

The second part of chapter 5 dealt with the recently developed radiation model of Simini et al. (2012). In particular, some initial attempts of assimilating the radiation model within our modeling framework were demonstrated. Initially, the variable of circular area population – introduced in the radiation model – was used as an explanatory variable in a statistical framework under negative binomial likelihood assumptions. These results did not meet initial expectations, nevertheless further future experiments based on different assumptions might provide more promising results. This section concluded with a possible Bayesian extension or variation of the radiation model. Results from such a modeling approach remain to be evaluated.

6.2 Relations to input and model uncertainty

Although not always explicitly mentioned throughout the text, the material of this dissertation is constantly related to modeling and incorporating various sources of uncertainty. Therefore, it is useful to interpret results from the perspective of uncertainty modeling. As discussed in section 1.3, two important review papers for uncertainty modeling in transportation studies are those of de Jong et al. (2006) and Rasouli and Timmermans (2012). In both papers two major sources of uncertainty are recognized; input-uncertainty which is due to future unobserved input, sampling bias, survey design and so forth, and also model-uncertainty which is due to model-specification and parameter-estimation error.

Chapter 3 can be essentially viewed as a study of model-uncertainty in OD modeling. In particular, the structure of the chapter is actually shaped by model-specification error due to distributional assumptions. Initially, a comparison is performed between the simple Poisson model and the negative binomial model accounting for overdisperion. Subsequently, further distributional assumptions are taken into account by changing the distribution of random effects from gamma to lognormal and also to inverse Gaussian, and results are compared under both hierarchical and marginal modeling structures.

On the other hand, less weight is placed on the deterministic assumption of the log-link function which is another component of model-uncertainty. Nevertheless, this is justifiable based on the following reasons. Firstly, the log-link function is commonly the prevailing option in Poisson or Poisson-mixture modeling as it is in theory a more consistent option in comparison, for instance, to the identity function. Secondly, initial experiments also revealed in practice that the Poisson and negative binomial models with a log-link function performed better than the models with an identity-link function in terms of AIC. Thirdly, as described in section 3.4 the use of the log-link function leads to an interesting alternative direct-demand gravity modeling interpretation.

The issue of omitted explanatory variables – still with respect to model-uncertainty – is in general more difficult to approach. Nevertheless, as discussed in section 3.2.6 and also as shown in section 5.2.1 experiments with additional explanatory variables have been conducted and the gains in model improvement were not found to be significant. The overall feeling with respect to the use of explanatory variables is that it is difficult to find other variables which will explain OD variability from an aspect which is not explained already by the choice of the existing variables.

The methodology of chapter 3 is purely statistical and therefore estimation-error – the second main source of model-uncertainty – for the parameters of scientific interest is constantly considered and quantified in terms of point, interval and standard deviation estimates. One might argue that estimation-error in the frequentist sense can be considered as sampling-error under the Bayesian framework. Nevertheless, sampling-error is also accounted for by monitoring MCMC convergence and the MC simulation error of each parameter (Appendix B). In addition, with Bayesian methodology uncertainty for parameters – or indeed for any function of the parameters – can be explicitly expressed in terms of posterior distributions. Another important advantage is that model uncertainty can be evaluated by means of short-term predictive testing.

The aspect of input-uncertainty is not pursued a lot in chapter 3. The reason actually relates to one of the fundamental arguments in favor of this research, namely that direct statistical modeling is justified given the fact that the main input-uncertainty sources of error (sampling and non representation-error) are not present in the census OD. Nevertheless, input-uncertainty is still considered to a certain degree. As discussed in section 3.1.3, due to a certain non-response rate two slightly different OD matrices are used in sections 3.2 and 3.3. Comparison based on the estimates from the negative binomial model indicated that there are no major differences when it comes to parameter interpretation and also statistical significance.

Input-uncertainty is the main theme of chapter 4. As demonstrated throughout chapter 4 and given the methodological framework of chapter 3, incorporating and quantifying input-uncertainty is relatively easy to implement. Once again, one of the advantages of the overall approach is that quantification of uncertainty is not restricted to traditional point/interval estimation, but is extended to distributional estimates of link volumes and V/C ratios given input-uncertainty.

Under this flexible methodological framework, aspects of model-uncertainty related to traffic assignment are also considered. In section 4.6, the initial results of section 4.5 based on PG predictions are compared to results based on PIG predictions. One can argue that it is rather unclear whether this comparison should be considered as an input-uncertainty or as a model-uncertainty comparison. From one point of view, the choice of the statistical model is directly related only to the OD predictions used as input to traffic assignment and therefore this comparison can be validly considered to be an input-uncertainty experiment. On the other hand, if one views the proposed methodology as a two-stage modeling approach, i.e. statistical OD modeling followed by traffic assignment modeling, then the comparison can be also validly viewed as a model-uncertainty experiment. Irrespective of point of view, the results discussed in section 4.6 highlight interesting differences.

Model-uncertainty concerning traffic assignment modeling per se is further explored in the first part of chapter 5, specifically in section 5.1. In this part model uncertainty originating from the choice of statistical model is kept fixed, i.e. only PG predictions are considered, and model-uncertainty originating from the choice of assignment procedure is investigated given input-uncertainty. The comparison takes into account several sources or levels of model-uncertainty. On a first level deterministic versus stochastic UE is considered, on a second level functional assumption of SUE are considered (probit versus logit route-choice modeling) while on a third level parametrical assumptions of SUE are let to vary (perception-error variance) in a sensitivity-analysis manner.

It should be noted that alternatives to the BPR (Bureau of Public Roads, 1964) link performance function, which is used for all traffic assignment applications of chapters 4 and 5, are not considered. Although that the link performance function is another functional assumption which would be possible to test, the BPR formulation seems to be widely accepted within the field and it is also used in most scientific UE traffic assignment applications, for example the BPR formulation is present in the studies of Sheffi and Powell (1981), Ji and Chen (2003) and Zhang (2011), among many others.

6.3 Applicability considerations

At this point it would be fair to contemplate on the applicability of the proposed methodology. One can argue that a serious limitation is that the overall approach is data-dependent, in particular with respect to the OD matrix. Direct statistical modeling requires that the OD matrix must be “reliable” to start with, i.e. the various sources of input-uncertainty error should have minimal effects on the final OD estimate used for the analysis.

The OD matrix derived from the 2001 Belgian census can be described as a reliable OD estimate, since at least the two main sources of input-uncertainty, sampling-error and non representation-error, are not present. Of course, OD matrices from census studies are extremely expensive data items and in addition not collected in frequent time-intervals. Even more, travel census data do not exist for the majority of countries at least not in the real meaning of “census” which strictly speaking is a study for the whole population of a country. The country of Belgium might be regarded as an exception, since the relatively small surface area and also population facilitates census tracking. On the other hand, one can easily imagine enormous difficulties for implementation of travel census studies in countries like Germany, France or USA, for instance.

The question then is whether it is safe to apply statistical modeling on OD matrices derived from travel-surveys which do not account for the population as a whole. Obviously this question is not straightforward to answer and most probably an absolute “no” as an answer would be equally wrong with a definite “yes”. Basically, it would all depend on the quality of the travel-survey in terms of taking into account propagation of error from various sources during the design and implementation of the survey.

Many developed countries support ongoing travel-surveys which are well-organized and are implemented on a nationwide level. Such surveys are frequently referred to as *micro-census* studies, since the aim is to project travel-survey information to the level of a nation. Examples can be found in McKenzie and Rapino (2011) for USA and in Marconi et al. (2005) who provide a comparative study for Switzerland, 8 EU countries and USA.

For these cases, the answer to the aforementioned question would probably be affirmative. In fact, since micro-census studies are updated in regular intervals, Bayesian modeling approaches might prove to be particularly suitable due to the natural updating prior/posterior framework. In addition, statistical models can always be extended so that additional sources of uncertainty are included in a given analysis. That is, additional error structures can be introduced accounting for sampling or non representation-errors, for instance.

6.4 Econometrical considerations

As with any scientific research there are further issues which arise during the course of experiments which are not initially under consideration. One of these issues relates to the fact that the research presented here started from a purely statistical point of view and not from an econometrical perspective. Therefore, attention did not capture specific modeling details which are probably important within an econometrical framework. Namely, not all explanatory variables correspond to the base-year of the 2001 Belgian census.

Specifically, the explanatory variables of employment ratio, population density and car ownership ratio correspond to those of year 2006. From a strictly statistical perspective such negligence is probably not significant in statistical and practical terms. In general, all three variables are related to population figures and one would not expect a sudden steep increase of such figures between the years of 2001 and 2006. In addition all three variables are actually transformed to ratios and therefore the corresponding differences between 2001 and 2006 are expected to be even less significant than those of untransformed variables. In practice, one would expect that the corresponding parameter estimates might differ up to a certain decimal point if the explanatory variables of the base year would have been used.

Nevertheless, according to the discussion of section 3.4, especially with respect to similarities with econometrical trade-flow studies, an econometrical implementation would be of potential interest for long-term forecasting applications. That is, if the models presented here were fitted with all explanatory variables measured in the base year, then it would be feasible to forecast long-term predictions by keeping parameters estimates fixed and by substituting the 2001 explanatory variables with future estimates of the explanatory variables. Of course, that would be possible given that future estimates of explanatory variables exist.

6.5 Future research

This dissertation concludes with a brief discussion concerning possible future research directions. Some are purely statistical considerations while others relate more to transportation modeling issues which remain to be addressed.

From a statistical perspective, additional models will be worthwhile the consideration. In particular, zero-inflated versions of the models presented here are a first clear option. As illustrated in section 3.3.6, both the PG and PIG models fail to replicate successfully the total number of zeros in the OD matrix. Although replicating the number of zeros was initially not one of the basic goals of the analysis, this particular failure of both models affects total demand predictions and consequently aggregated link flow predictions. This line of research can start similarly from simple Bayesian versions of the zero-inflated Poisson (Lambert, 1992) and negative binomial (Heibron, 1994) models and build up to new and novel modeling applications based on other distributions.

Including a wider range of models may also lead to an interesting application of Bayesian *model averaging*. According to Kass and Raftery (1995) this technique yields consistently and substantially better prediction than methods based on individual models (see also Hoeting et al., 1999). Within our working framework this approach might prove to be useful for the traffic assignment stage. At present, the link flow predictions are subjected to model-uncertainty both from statistical modeling as well as traffic assignment modeling. Bayesian model averaging for the predictions generated by numerous models will at least reduce the uncertainty originating from statistical modeling.

Another more ambitious statistical research direction involves changing completely the inferential framework to dynamic modeling. That is, simultaneous modeling of OD matrices on discrete time-intervals, e.g. on daily intervals. At present, it would be difficult to express anything more than conceptual suggestions starting from the covariate-based Poisson modeling framework introduced here only this time from a time-series perspective which can potentially build up to Poisson-mixture time-series models and to analogous extensions of these according to the future directions discussed previously, i.e. zero-inflated versions.

Future research directions concerning transportation modeling issues are many. First of all, with respect to the two inputting methods introduced in section 4.4 research focusing on method 1 (approximate network-inference) might deserve some further attention. Method 1 is approximate and as such it requires less computational time which makes it suitable for fast policy-planning decisions. To this end, the comparison between method 1 and method 2 (full network-inference) provided some evidence that method 1 might be suitable when point estimates of expected network states are of importance. Nevertheless, as discussed in section 4.7 this statement can be properly supported only if similar research is conducted for networks with different characteristics. A further topic of interest would be to investigate alternative methods for deriving percentile estimates from method 1, methods which will potentially provide percentile estimates closer to the corresponding estimates of method 2. In fact, this research field is free for experimentation. Although research focusing on improvement of centrality measures under approximate network-inference does exist (e.g. Ukkusuri and Waller, 2006), relevant research on percentile estimates seems to be non-existing in the relative literature, as yet.

A lot could be said with respect to congested link identification in terms of probability estimates which was introduced in section 4.5.4 and further analyzed in sections 4.6.3 and 5.1.4. Nevertheless, the most challenging and perhaps also most interesting direction for this type of research is a bridging attempt with critical link identification studies within the framework of vulnerability analysis. The challenging, but also promising, part for this type of research is that both research approaches – method 2 introduced here and also critical link identification algorithms – are relatively new and untested. Of course, as mentioned in section 4.7 a combination of full network input-uncertainty inference with the full-network scan approach of Jenelius et al. (2006) will prove to be, most probably, prohibitive in terms of computational expense, at least for cases of large-scale congested networks. On the other hand, combining full network input-uncertainty inference with the reduced-scan algorithm proposed in Chen et al. (2012) might be feasible.

A further issue which was not included in this current analysis and can be incorporated in future research is that of modal-split modeling, the third phase of the sequential four-step model. Modal-spit is a well established aspect of transportation modeling and therefore several methodologies are available (see e.g. chapter 6 in Ortúzar and Willumsen, 2001). Nevertheless, despite the choice of modal-split technique, an interesting question arising in this case would be whether to implement modal-split first and then use statistical modeling for the disaggregated OD matrices – either separately or simultaneously – or implement modal-split directly upon the OD predictions generated by a statistical model.

Another research category relates to traffic assignment experiments concerning theoretical as well as practical aspects of traffic assignment models. One of the advantages of the proposed methodology is that it is not restrained by the selection of the traffic assignment model. Therefore, future research can focus on comparative studies concerning point, interval and distribution estimates resulting by different traffic assignment models. A first comparison concerning differences between DUE and SUE estimates was presented in section 5.1. This line of research can potentially extend to truly stochastic traffic assignment approaches such as the models proposed for instance by Cascetta (1989) and Hazelton (1998).

Finally, the radiation model of Simini et al. (2012) is a new and novel methodological development which will potentially impact transportation modeling in general. A first conceptual Bayesian representation of this model has been presented in section 5.2.3. This modeling approach remains to be implemented, evaluated and most probably improved and refined.

Appendix A

A.1 Generating a MH sample of size M from the Poisson model.

1. Set starting value $\boldsymbol{\beta}^0$.
2. For iterations $t = 1, 2, \ldots, M$:
    a. Generate $\boldsymbol{\beta}^*$ from the proposal distribution $q(\boldsymbol{\beta})$.
    b. Calculate the transition probability $a_{MH} = \min\left[\dfrac{p(\boldsymbol{\beta}^* \mid \mathbf{y})q(\boldsymbol{\beta}^{t-1})}{p(\boldsymbol{\beta}^{t-1} \mid \mathbf{y})q(\boldsymbol{\beta}^*)}, 1\right]$.
    c. Generate a uniform random number u from $U(0,1)$.
    d. Set $\boldsymbol{\beta}^t = \begin{cases} \boldsymbol{\beta}^*, & \text{if } u \le a_{MH}, \\ \boldsymbol{\beta}^{t-1}, & \text{if } u > a_{MH}. \end{cases}$

The un-normalized posterior density $p(\boldsymbol{\beta} \mid \mathbf{y})$ is given in expression 3.3.

A.2 Generating a MH sample of size M from the PG/PLN/PIG models.

1. Set starting values $\boldsymbol{\beta}^0, \omega^0$.
2. For iterations $t = 1, 2, \ldots, M$:
    a. Generate $\boldsymbol{\beta}^*$ from the proposal distribution $q(\boldsymbol{\beta})$ and ω^* from the proposal distribution $q(\omega)$.
    b. Calculate the transition probability

$$a_{MH} = \min\left[\frac{p(\boldsymbol{\beta}^*, \omega^* \mid \mathbf{y})q(\boldsymbol{\beta}^{t-1})q(\omega^{t-1})}{p(\boldsymbol{\beta}^{t-1}, \omega^{t-1} \mid \mathbf{y})q(\boldsymbol{\beta}^*)q(\omega^*)}, 1\right].$$

    c. Generate a uniform random number u from $U(0,1)$.
    d. Set $(\boldsymbol{\beta}^t, \omega^t) = \begin{cases} (\boldsymbol{\beta}^*, \omega^*), & \text{if } u \le a_{MH}, \\ (\boldsymbol{\beta}^{t-1}, \omega^{t-1}), & \text{if } u > a_{MH}. \end{cases}$

Parameter ω corresponds to the over-dispersion parameter; for the PG model $\omega = \theta$, for the PLN model $\omega = \sigma_{\mathbf{u}}^2$ and for the PIG model $\omega = \zeta$. The corresponding un-normalized posterior densities $p(\boldsymbol{\beta}, \omega \mid \mathbf{y})$ of the marginal models are given in expressions 3.6 (PG model), 3.9 (PLN model) and 3.11 (PIG model).

Appendix B

B.1 Convergence diagnostics for the Poisson model (section 3.2.3).

Parameter	Convergence diagnostics				
	Heidelberger & Welch		Raftery & Lewis	Geweke	MC error
	Stationarity test p-value*	Half-width test p-value**	Dependence factor (I)	Equality of means test absolute z-value***	Spectral density estimate
β_0	0.4777	9.94e-04	0.970	0.290	5.070102e-04
β_1	0.2798	9.77e-07	0.970	1.091	4.986712e-07
β_2	0.2902	8.58e-07	0.991	0.473	4.377202e-07
β_3	0.6494	1.40e-06	1.030	0.701	7.119120e-07
β_4	0.7805	1.05e-06	0.991	0.888	5.356828e-07
β_5	0.9490	2.60e-06	1.010	0.296	1.324426e-06
β_6	0.5705	5.73e-05	0.991	0.904	2.923395e-05
β_7	0.7609	5.95e-05	0.970	0.566	3.035946e-05
β_8	0.6383	4.44e-05	1.100	1.740	2.265771e-05
β_9	0.7294	1.78e-04	1.010	0.476	9.068293e-05
β_{10}	0.9702	2.69e-04	0.991	0.549	1.374306e-04
β_{11}	0.6663	2.10e-04	1.010	0.929	1.072890e-04
β_{12}	0.6895	5.27e-05	0.991	0.272	2.686587e-05
β_{13}	0.8246	6.37e-05	0.991	0.361	3.250084e-05
β_{14}	0.7521	7.80e-05	1.010	0.904	3.978662e-05
β_{15}	0.2602	1.04e-04	1.020	0.711	5.281253e-05
β_{16}	0.6369	7.01e-05	1.030	0.069	3.574182e-05
β_{17}	0.5064	1.07e-04	0.991	0.263	5.450614e-05
β_{18}	0.0638	5.93e-04	0.970	1.111	3.024667e-04
β_{19}	0.3556	6.06e-04	0.951	0.607	3.089782e-04
β_{20}	0.2060	7.45e-06	0.991	0.653	3.798645e-06
β_{21}	0.2383	9.72e-06	1.080	1.411	4.957593e-06
β_{22}	0.7432	3.78e-05	1.010	0.324	1.927498e-05
β_{23}	0.7806	5.40e-05	1.010	0.079	2.755243e-05
β_{24}	0.9343	4.67e-05	1.050	0.057	2.383729e-05

* Stationarity is rejected for p-values smaller than 0.05.

** The test is passed for p-values greater then 0.05, i.e. when rejecting the null hypothesis.

*** Equality of means is rejected for z-values greater than 1.96 in absolute terms.

B.2 Convergence diagnostics for the negative binomial model (section 3.2.3).

Parameter	Convergence diagnostics				
	Heidelberger & Welch		Raftery & Lewis	Geweke	MC error
	Stationarity test p-value*	Half-width test p-value**	Dependence factor (I)	Equality of means test absolute z-value***	Spectral density estimate
β_0	0.1645	1.39e-02	1.050	2.265	6.646236e-03
β_1	0.8842	4.55e-06	1.010	1.249	2.319640e-06
β_2	0.5270	6.79e-06	1.120	0.623	3.464543e-06
β_3	0.2343	1.12e-05	1.010	0.714	5.366391e-06
β_4	0.1296	1.08e-05	1.020	0.787	5.529487e-06
β_5	0.7534	2.52e-05	1.070	1.253	1.284598e-03
β_6	0.2562	5.13e-04	1.140	2.413	2.616511e-04
β_7	0.8058	7.03e-04	0.970	0.344	3.587888e-04
β_8	0.0666	5.98e-04	1.010	1.335	3.049852e-04
β_9	0.6047	2.25e-03	1.120	0.884	1.146667e-03
β_{10}	0.0942	2.51e-03	1.170	0.489	1.278680e-03
β_{11}	0.4000	2.57e-03	1.050	0.824	1.313721e-03
β_{12}	0.2273	4.20e-04	0.991	1.840	2.143684e-04
β_{13}	0.2760	4.55e-04	1.010	1.522	2.320403e-04
β_{14}	0.9060	7.20e-04	1.010	0.539	3.671063e-04
β_{15}	0.8341	7.67e-04	1.080	0.883	3.913551e-04
β_{16}	0.7531	7.20e-04	1.120	0.572	3.673767e-04
β_{17}	0.9580	6.43e-04	1.030	0.107	3.281314e-04
β_{18}	0.6247	5.47e-03	1.010	0.853	2.789266e-03
β_{19}	0.3821	5.33e-03	0.991	1.385	2.853554e-03
β_{20}	0.4186	6.50e-05	1.170	0.284	3.317650e-05
β_{21}	0.3241	6.88e-05	1.050	0.950	3.510134e-05
β_{22}	0.1668	3.34e-04	1.270	1.638	1.701795e-04
β_{23}	0.3336	3.16e-04	1.030	0.412	1.609898e-04
β_{24}	0.8220	3.43e-04	1.110	0.584	1.751428e-04
θ	0.0620	3.02e-04	1.030	1.157	1.543145e-04

* Stationarity is rejected for p-values smaller than 0.05.

** The test is passed for p-values greater then 0.05, i.e. when rejecting the null hypothesis.

*** Equality of means is rejected for z-values greater than 1.96 in absolute terms.

B.3 Convergence diagnostics for the PG model (section 3.3.4).

Parameter	Convergence diagnostics				
	Heidelberger & Welch		Raftery & Lewis	Geweke	MC error
	Stationarity test p-value*	Half-width test p-value**	Dependence factor (I)	Equality of means test absolute z-value***	Spectral density estimate
β_0	0.7889	1.36e-02	1.030	1.491	6.915429e-03
β_1	0.2841	4.70e-06	0.991	0.890	2.398124e-06
β_2	0.4956	7.52e-06	1.020	0.774	3.838305e-06
β_3	0.0736	1.25e-05	1.010	1.182	6.364534e-06
β_4	0.7396	1.11e-05	1.050	0.183	5.662365e-06
β_5	0.2376	2.78e-05	1.050	2.275	1.410724e-05
β_6	0.4366	6.61e-04	1.100	0.559	3.374072e-04
β_7	0.9112	5.83e-04	0.970	0.027	2.976950e-04
β_8	0.2266	7.58e-04	1.220	0.271	3.868395e-04
β_9	0.5413	6.23e-04	1.010	1.521	3.179306e-04
β_{10}	0.2373	2.29e-03	0.991	1.354	1.168372e-03
β_{11}	0.4305	2.36e-03	1.100	0.317	1.205982e-03
β_{12}	0.6138	1.88e-03	0.970	0.555	9.591993e-04
β_{13}	0.2162	4.54e-04	0.991	0.874	2.316420e-04
β_{14}	0.3770	5.10e-04	1.010	1.924	2.600273e-04
β_{15}	0.0529	8.35e-04	1.010	0.998	3.821023e-04
β_{16}	0.2691	7.49e-04	1.050	1.845	3.823767e-04
β_{17}	0.0883	7.64e-04	1.080	1.254	3.899247e-04
β_{18}	0.6361	6.72e-04	0.951	0.363	3.426730e-04
β_{19}	0.5192	4.56e-03	1.100	0.378	2.326434e-03
β_{20}	0.4427	5.02e-03	1.030	0.408	2.560980e-03
β_{21}	0.9868	6.10e-05	1.080	0.589	3.114013e-05
β_{22}	0.8197	5.09e-05	1.100	0.849	2.598050e-05
β_{23}	0.0611	3.43e-04	1.050	1.023	1.751166e-04
β_{24}	0.0539	2.96e-04	0.991	0.686	1.511616e-04
β_{25}	0.1399	3.53e-04	0.991	0.845	1.801585e-04
θ	0.1417	2.82e-04	1.010	1.475	1.439187e-04

* Stationarity is rejected for p-values smaller than 0.05.

** The test is passed for p-values greater then 0.05, i.e. when rejecting the null hypothesis.

*** Equality of means is rejected for z-values greater than 1.96 in absolute terms.

B.4 Convergence diagnostics for the PLN model (section 3.3.4).

Parameter	Convergence diagnostics				
	Heidelberger & Welch		Raftery & Lewis	Geweke	MC error
	Stationarity test p-value*	Half-width test p-value**	Dependence factor (I)	Equality of means test absolute z-value***	Spectral density estimate
β_0	0.6368	1.52e-02	0.991	0.773	7.771081e-03
β_1	0.2690	5.68e-06	0.991	0.481	2.896595e-06
β_2	0.9831	6.83e-06	0.951	0.747	3.484824e-06
β_3	0.2264	1.19e-05	1.010	1.342	6.056341e-06
β_4	0.4814	1.43e-05	0.991	0.695	7.309160e-06
β_5	0.4549	2.53e-05	1.030	0.072	1.289004e-05
β_6	0.2473	6.49e-04	0.999	1.825	3.309656e-04
β_7	0.9291	6.20e-04	1.030	1.120	3.160934e-04
β_8	0.0640	7.27e-04	1.050	0.824	3.711024e-04
β_9	0.0947	6.82e-04	0.970	1.919	3.477838e-04
β_{10}	0.2870	2.71e-03	0.970	0.959	1.382304e-03
β_{11}	0.6166	2.13e-03	0.970	0.151	1.087384e-03
β_{12}	0.9858	2.38e-03	0.991	1.140	1.213049e-03
β_{13}	0.0566	5.13e-04	0.942	0.280	2.618366e-04
β_{14}	0.0618	4.69e-04	0.970	1.933	2.392921e-04
β_{15}	0.3057	8.10e-04	1.010	0.542	4.134017e-04
β_{16}	0.3500	7.04e-04	0.991	1.226	3.591040e-04
β_{17}	0.4447	7.82e-04	1.030	0.173	3.991502e-04
β_{18}	0.1124	6.96e-04	1.010	1.349	3.550184e-04
β_{19}	0.6309	5.33e-03	1.050	0.349	2.718799e-03
β_{20}	0.8395	4.81e-03	1.030	1.373	2.454120e-03
β_{21}	0.9281	5.10e-05	1.050	1.525	2.603442e-05
β_{22}	0.9033	4.44e-05	1.010	0.423	2.263907e-05
β_{23}	0.1721	2.88e-04	1.030	0.353	1.467539e-04
β_{24}	0.2299	3.23e-04	1.080	1.896	1.646602e-04
β_{25}	0.3316	3.05e-04	1.010	0.324	1.558590e-04
$\sigma_{\mathbf{u}}^2$	0.4935	3.14e-04	1.040	0.604	1.600820e-04

* Stationarity is rejected for p-values smaller than 0.05.

** The test is passed for p-values greater then 0.05, i.e. when rejecting the null hypothesis.

*** Equality of means is rejected for z-values greater than 1.96 in absolute terms.

B.5 Convergence diagnostics for the PIG model (section 3.3.4).

Parameter	Convergence diagnostics				
	Heidelberger & Welch		Raftery & Lewis	Geweke	MC error
	Stationarity test p-value*	Half-width test p-value**	Dependence factor (I)	Equality of means test absolute z-value***	Spectral density estimate
β_0	0.5788	1.50e-02	1.190	0.477	7.678064e-03
β_1	0.5341	4.74e-06	1.050	0.246	2.416379e-06
β_2	0.6168	7.52e-06	0.979	0.617	3.835573e-06
β_3	0.2348	1.09e-05	0.970	1.218	5.544257e-06
β_4	0.2671	1.20e-05	1.010	1.216	6.135603e-06
β_5	0.3034	2.89e-05	0.970	0.772	1.475210e-05
β_6	0.0893	8.09e-04	0.991	1.994	3.188259e-04
β_7	0.9351	5.64e-04	1.530	0.837	2.876201e-04
β_8	0.2437	7.77e-04	0.970	1.211	3.962487e-04
β_9	0.9977	6.10e-04	1.080	0.456	3.114344e-04
β_{10}	0.1428	2.41e-03	0.962	0.779	1.230607e-03
β_{11}	0.2589	2.26e-03	0.979	0.613	1.151560e-03
β_{12}	0.1298	2.31e-03	0.991	0.173	1.178499e-03
β_{13}	0.0571	4.98e-04	0.991	1.887	2.540307e-04
β_{14}	0.6332	5.15e-04	1.100	0.020	2.628275e-04
β_{15}	0.3166	6.26e-04	0.991	1.456	3.194033e-04
β_{16}	0.8687	8.09e-04	1.070	0.036	4.126415e-04
β_{17}	0.1072	7.68e-04	1.080	2.242	3.916494e-04
β_{18}	0.5121	8.75e-04	0.991	0.117	4.463998e-04
β_{19}	0.9373	5.31e-03	0.970	0.146	2.708647e-03
β_{20}	0.0954	5.33e-03	0.991	0.505	2.719050e-03
β_{21}	0.3577	5.43e-05	1.130	1.041	2.771298e-05
β_{22}	0.2668	6.36e-05	0.979	0.066	3.245703e-05
β_{23}	0.1153	2.94e-04	1.020	2.377	1.500488e-04
β_{24}	0.6832	3.19e-04	1.010	0.100	1.626946e-04
β_{25}	0.7769	3.32e-04	0.979	0.271	1.694052e-04
ζ	0.1185	3.28e-04	1.020	1.809	1.556842e-04

* Stationarity is rejected for p-values smaller than 0.05.

** The test is passed for p-values greater then 0.05, i.e. when rejecting the null hypothesis.

*** Equality of means is rejected for z-values greater than 1.96 in absolute terms.

Appendix C

Routines and external libraries in R (when needed) utilized for ML estimation, density calculations, sampling and MCMC convergence diagnostics.

Purpose	Model / Distribution	Routine	R library
ML estimation	Poisson	glm	(default)
	Negative binomial	glm.nb	MASS
	Poisson-lognormal (GLMM form)	lmer glmmML	lme4 glmmML
	Poisson-inverse Gaussian	gamlss	gamlss
Density calculations and sampling	Multivariate normal	rmvtnorm dmvtnorm	mvtnorrm
	Poisson-lognormal	dpoilog	poilog
	Poisson-inverse Gaussian	dPIG	gamlss
	Generalized inverse Gaussian	rgig	HyperbolicDist
MCMC convergence diagnostics		heidel.diag raftery.diag geweke.diag	coda

References

Abrahamsson, T. (1998). Estimation of origin-destination matrices using traffic counts – a literature survey, Working papers, International Institute for Applied Systems Analysis.

Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed., John Wiley and Sons, New York.

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models, *Statistics and Computing*, 6, 251-262.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.

Anderson, D.A. and Hinde, J.P. (1988). Random effects in generalized linear models and the EM algorithm, *Communications in Statistics - Theory and Methods*, 17, 3847-3856.

Anderson, J.E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review*, 69(1), 106-116.

Arentze, T.A. and Timmermans, H.J.P. (2004). A learning-based transportation oriented simulation system. *Transportation Research Part B*, 38(7), 613-633.

Arentze, T.A. and Timmermans, H.J.P. (2011). A dynamic model of time-budget and activity generation: development and empirical validation. *Transportation Research Part C*, 19(2), 242-253.

Ashley, D.L. (1980). Uncertainty in the context of highway appraisal. *Transportation*, 9(3), 249-267.

Athreya, K.B. (1986). Another conjugate family for the normal distribution. *Statistics and Probability Letters*, 4(2), 61-64.

Atkinson, A.C. (1982). The simulation of generalized inverse Gaussian and hyperbolic random variables. *SIAM Journal on Scientific and Statistical Computing*, 3, 502-515.

Ayyub, B.M. and Gupta, M.M. (1998). *Uncertainty Analysis in Engineering and Sciences*, 1st ed., Springer, New York.

Bain, R. and Polakovic, L. (2005). Traffic forecasting risk study update 2005: through ramp-up and beyond. In *Proceedings of European Transport Conference*, Strasbourg, France.

Bell, M.G.H. (1991). The estimation of origin-destination matrices by constrained generalized least squares. *Transportation Research Part B*, 25(1), 13-22.

Bellemans, T., Kochan, B., Janssens, D., Wets, G., Arentze, T.A. and Timmermans, H.J.P. (2010). Implementation framework and development trajectory of FEATHERS activity-based simulation platform. *Transportation Research Record*, 2175, 111-119.

Ben Akiva, M. (1974). Structure of passenger travel demand models. *Transportation Research Record*, 526, 26-41.

Ben Akiva, M. and Bierlaire, M. (1999). Discrete choice models and their application to short term travel decisions. In R.W. Hall, ed., *Handbook of Transportation Science*, Kluwer, Massachusetts, pp. 5-34.

Berdica, K. (2002). An introduction to road vulnerability: what has been done, is done and should be done. *Transport Policy*, 9(2), 117-127.

Bergstrand, J.H. (1985). The gravity equation in international trade: some microeconomic foundations and empirical evidence. *The Review of Economics and Statistics*, 67(3), 474-481.

Bernardo, J. and Smith, A. (1994), *Bayesian Theory,* John Wiley and Sons, Chichester, UK.

Bhat, C.R., Guo, J.Y., Srinivasan, S. and Sivakumar, A. (2004). A comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record*, 1894, 57-66.

Bhat, C.R. and Koppelman, F.C. (1999). Activity-based modeling of travel demand. In R.W. Hall, ed., *Handbook of Transportation Science*, Kluwer, Massachusetts, pp. 35-62.

Bickel, D.R. and Früwirth, R. (2006).On a fast, robust estimator of the mode: comparisons to other robust estimators with applications, *Computational Statistics and Data Analysis,* 50(12), 3500-3530.

Bierlaire, M. and Toint, P.L. (1995). Meuse: an origin-destination matrix estimator that exploits structure. *Transportation Research Part B*, 29(1), 47-60.

Bonsall, P.W., Champerowne, A.F., Mason, A.C. and Wilson, A.G. (1977). Transport modeling: sensitivity analysis and policy testing. *Progress in Planning*, 7(3), 153-237.

Boucher, J.-P. and Denuit, M. (2006). Fixed versus random effects in Poisson regression models for claim counts: a case study with motor insurance. *Astin Bulletin*, 36(1), 285-301.

Boyce, D. (2002). Is the sequential travel forecasting paradigm counterproductive? *Journal of Urban Planning and Development*, 128(4), 169-183.

Boyles, S.D, Kockelman, K.M. and Waller, S.T. (2010). Congestion pricing under operational, supply-side uncertainty. *Transportation Research Part C*, 18(4), 519-535.

Brach, R.M. and Dunn, P.F. (2004). *Uncertainty Analysis for Forensic Science*, Lawyers and Judges Pub Co, Tucson, Arizona.

Brooks, S.P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin and van der Linde. *Journal of the Royal Statistical Society B*, 64(4), 616-618.

Bureau of Public Roads (1964). *Traffic assignment manual for application with a large, high speed computer*, U.S. Dept. of Commerce Bureau of Public Roads Office of Planning Urban Planning Division, Washington.

Breslow, N.E. (1984). Extra-Poisson variation in log-linear models. *Journal of the Royal Statistical Society Series C,* 33(1), 38-44.

Brög, W. and Ampt, E. (1982). State of the art in the collection of travel behaviour data. In *Travel Behaviour for the 1980's*, Special Report 201, National Research Council, Washington, DC.

Brög, W. and Erl, E. (1982). Application of correction and weighting factors to obtain a representative data base. In *Proceedings of 10th PTRC Summer Annual Meeting*, University of Warwick, Warwick, UK.

Brooks, S.P. and Roberts, G.O. (1998). Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing*, 8(4), 319-335.

Browne, W.J. (2003). *MCMC estimation in MLwiN*, Centre for Multilevel Modelling, University of Bristol.

Caliper Corporation (2008). *TransCAD: Transportation Planning Software*, Newton, USA.

Cantarella, G.E. and Cascetta, E. (1998). Stochastic assignment to transportation networks: models and algorithms. In P. Marcotte and S. Nguyen, eds., *Equilibrium and advanced transportation modeling*, Kluwer Academic, Boston, pp. 87-107.

Carlin, B.P. and Louis, T. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*, 1st ed., Chapman and Hall/CRC, London.

Casella, G. and George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46(3), 167-174.

Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research Part B*, 18(4-5), 289-299.

Cascetta, E. (1989). A stochastic process approach to the analysis of temporal dynamics in transportation networks. *Transportation Research Part B*, 23(1), 1-17.

Cascetta, E. (2009). *Transportation Systems Analysis: Models and Applications*, 2nd ed., Springer, New York.

Cascetta, E. and Nguyen, S. (1988). A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B*, 22(6), 437-455.

Casey, H.J. (1955). Applications to traffic engineering of the law of retail gravitation. *Traffic Quarterly*, IX, 23-35.

Castillo, E., Menéndez, J.M. and Sánchez-Cambronero, S. (2008). Predicting traffic flow using Bayesian networks. *Transportation Research Part B*, 42(5), 482-509.

Celeux, G., Forbes, F., Robert, C.P and Titterington, D.M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4), 651-674.

Chen, B.Y., Lam, W.H.K., Sumalee, A., Li, Q. and Li, Z.-C. (2012). Vulnerability analysis for large-scale and congested road networks with demand uncertainty. *Transportation Research A*, 46(3), 501-516.

Chen, J.J. and Ahn, H. (1996). Fitting mixed Poisson regression models using quasi-likelihood methods. *Biometrical Journal*, 38(1), 81-96.

Chen, Y. (1994). *Bilevel programming problems: analysis, algorithms and applications*, Report 984, University of Montreal, Centré de recherché sur les transports (CRT), Montreal, Canada.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4), 327-335.

Congdon, P.P. (2001). *Bayesian Statistical Modelling*, 1st ed., John Wiley and Sons, Chichester, UK.

Cools, M. (2009). *Inter- and intraday variability of Flemish travel behavior*. PhD dissertation, Hasselt University, Transportation Research Institute, Diepenbeek, Belgium.

Cools, M., Moons, E. and Wets, G. (2010). Assessing quality of origin-destination matrices derived from activity and travel surveys: results from a Monte Carlo experiment. *Transportation Research Record*, 2183, 49-59.

Cowles, M.K. and Carlin, B.P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434), 883-904.

Daganzo, C.F. (1979). *Multinomial Probit: The Theory and its Applications to Demand Forecasting*, Academic Press, New York, USA.

Daganzo, C.F. and Sheffi, Y. (1977). On stochastic models of traffic assignment. *Transportation Science*, 11(3), 253-274.

Dagpunar, J. (1988). *Principles of Random Variate Generation*, Oxford University Press, USA.

Davidson, W., Donnelly, R., Vovsha, P., Freedman, J., Ruegg, S., Hicks, J., Castiglione, J. and Picado, R. (2007). Synthesis of first practices and operational research approaches in activity-based travel demand modeling. *Transportation Research Part A*, 41(5), 464-488.

Dean, C., Lawless, J. F. and Willmot, G.E. (1989). A mixed Poisson-inverse-Gaussian regression model. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 17(2), 171-181.

Dean, C.B. (1994). Modified pseudo-likelihood estimator of the overdispersion parameter in Poisson mixture models. *Journal of Applied Statistics*, 21, 523-532.

de Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R. and Hofman, F. (2006). Uncertainty in traffic forecasts: literature review and new results for the Netherlands. *Transportation* 34(4), 375-395.

D'Este, G.M. and Taylor, M.A.P. (2003). Network vulnerability: an approach to reliability analysis at the level of national strategic transport networks. In Y. Iida and M.G.H. Bell, eds., *The Network Reliability of Transport*, Pergamon-Elsevier, Oxford, pp. 23–44.

Dial, R. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5(2), 83-111.

Domencich, T.A. and McFadden, D. (1975). *Urban Travel Demand: A behavioural analysis*, North-Holland Publishing Co., Amsterdam, The Netherlands.

Domencich, T.A., Kraft, G. and Valette, J. (1968). Estimation of urban passenger travel behavior: an economic demand model. *Highway Research Record*, 238, 64-278.

Drissi-Kaïtouni, O. and Lundgren, J. (1992). *Bilevel origin-destination matrix estimation using a descent approach*, Report, Linköping University, Department of Mathematics, Institute of Technology, Linköping, Sweden.

Duthie, J. C., Unnikrishnan, A. and Waller, S. T. (2011). Influence of demand uncertainty and correlations on traffic predictions and decisions. *Computer Aided Civil and Infrastructure Engineering*, 26(1), 16-29.

Eaves, D. and Chang, T. (1992). Posterior mode estimation for the generalized linear model. *Annals of the institute of statistical mathematics*, 44(3), 417-434.

Ettema, D. and Timmermans, H. (2006). Costs of travel time uncertainty and benefits of travel time information: conceptual model and numerical examples. *Transportation Research Part C*, 14(5), 335-350.

Fahrmeir, L. and Osuna, L. (2003). Structured count data regression. Collaborative Research Project 386, Discussion Paper 334.

Fernández, C., Ley, E. and Steel, M.F.J. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 100(2), 381-427.

Fisk, C. (1980). Some developments in equilibrium traffic assignment. *Transportation Research Part B*, 14(3), 243-255.

Fisk, C.S. (1989). Trip matrix estimation from link traffic counts: the congested network case. *Transportation Research Part B*, 23(5), 331-336.

Florian, M. and Hearn, D. (1995). Network equilibrium models and algorithms. In M.O. Ball, T.L. Magnanti, C.L. Monma and G.L. Nemhauser, eds., *Handbooks in OR and MS*, Elsevier, Amsterdam, pp. 485-550.

Flyvbjerk, B. (2005). Measuring inaccuracy in travel demand forecasting: methodological considerations regarding ramp up and sampling. *Transportation Research Part A*, 39(6), 522-530.

Flyvbjerk, B., Holm, M. and Buhl, S. (2005). How (in) accurate are travel demand forecasts in public work projects? *Journal of the American Planning Association*, 71(2), 131-146.

Flyvbjerk, B., Holm, M. and Buhl, S. (2006). Inaccuracy in traffic forecasts. *Transport Reviews*, 26(1), 1-24.

FOD Economie, (2010). FOD Economie - SPF Economie - Statbel. <http://www.statbel.fgov.be>.

Frank, M. and Wolfe, P. (1956). An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2), 95-110.

Frantzeskakis, J.M. and Iordanis, D.I. (1987). Volume-to-capacity ratio and traffic accidents on interurban four-lane highways in Greece. *Transportation Research Record*, 1112, 29-38.

Furness, K.P. (1965). Time function iteration. *Traffic Engineering and Control*, 7, 458-460.

Gamerman, D. and Lopes, H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed., Chapman and Hall/CRC, New York/London.

Gardner, L. M., Unnikrishnan, A. and Waller, S.T. (2008). Robust pricing of transportation networks under uncertain demand. *Transportation Research Record*, 2085(1), 21-30.

Gardner, L. M., Unnikrishnan, A. and Waller, S. T. (2010). Solution methods for robust pricing of transportation networks under uncertain demand. *Transportation Research Part C*, 18(5), 656-667.

Gardner, L. M., Boyles, S.D. and Waller, S. T. (2011). Quantifying the benefit of responsive pricing and travel information in the stochastic congestion pricing problem. *Transportation Research Part A*, 45(3), 204-218.

Geisser. S. and Eddy, W.F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365), 153-160.

Gelfand, A.E. (2000). Gibbs sampling. *Journal of the American Statistical Association*, 95(452), 1300-1304.

Gelfand, A.E. and Ghosh, M. (2000). Generalized linear models: a Bayesian view. In D.K. Dey, S.K. Ghosh, and B.K. Mallick, eds., *Generalized Linear Models: A Bayesian Perspective*, Marcel Dekker, New York, pp. 3-14.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003). *Bayesian Data*

Analysis, 2nd ed., Chapman and Hall/CRC, London.

Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, 1st ed., Cambridge University Press, New York.

Gelman, A., Meng, X.-L. and Stern, H. (1993). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, 6, 733-807.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 721-741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. Bernardo, J. Berger, A. Dawid, and A. Smith, eds., *Bayesian Statistics*, Vol. 4, Claredon Press, Oxford, pp. 169-193.

Ghosh, J.K. and Mukerjee, R. (1995). On perturbed ellipsoidal and highest posterior density regions with approximate frequentist validity, *Journal of the Royal Statistical Society B*, 57(4), 761-769.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC, New York.

Habib, K.M.N. and Miller, E.J. (2009). Modeling activity generation: a utility-based model of activity agenda formation. *Transportmetrica*, 5(1), 3-23.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97-109.

Hazelton, M.L. (1998). Some remarks on stochastic user equilibrium. *Transportation Research Part B*, 32(2), 101-108.

Hazelton, M.L. (2000). Estimation of origin-destination matrices from link flows on uncongested networks. *Transportation Research Part B*, 34(7), 549-566.

Hazelton, M.L. (2008). Statistical inference for time varying origin-destination matrices. *Transportation Research Part B*, 42(6), 442-452.

Hazelton, M.L. (2010). Bayesian inference for network-based models with a linear inverse structure. *Transportation Research Part B*, 44(5), 674-685.

Hedeker, D. and Gibbons, R.D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 50(4), 933-944.

Heibron, D.C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, 36(5), 531-547.

Heidelberger, P. and Welch, P.D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6), 1109-1144.

Hensher, D.A. and Button, K.J. (2000). *Handbook of Transport Modelling*, Elsevier Ltd, London.

Henson, K., Goulias, K. and Golledge, R. (2009). An assessment of activity-based modeling and simulation for applications in operational studies, disaster preparedness, and homeland security. *Transportation Letters*, 1(1), 19-39.

Hess, S. (2005). *Advanced discrete choice models with applications to transport demand.* PhD dissertation. Imperial College London, Centre for Transport Studies, London.

Hinde, J. (1982). Compound Poisson regression models. In R. Gilchrist, ed., *Lecture Notes in Statistics*, Vol. 14, Spinger-Verlag, New York, pp. 109-121.

Hinde, J. and Demétrio, C.G.B. (1998). Overdispersion: models and estimation. *Computational Statistics and Data Analysis*, 27, 151-170.

Highway Capacity Manual (HCM) Update. (1994). Transportation Research Board, Washington, DC.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). Bayesian model averaging: a tutorial. *Statistical Science*, 14(4), 382-401.

Holla, M.S. (1967). On a Poisson-inverse Gaussian distribution. *Metrika*, 11, 115-121.

Hyndman, R.J. (1996). Computing and graphing highest posterior density regions. *The American Statistician*, 50(2), 120-126.

Jenelius, E., Petersen, T. and Mattsson, L.G. (2006). Importance and exposure in road network vulnerability analysis. *Transportation Research Part A*, 40(7), 537-560.

Jones, P.M., Dix, M.C., Clarke, M.I. and Heggie, I.G. (1983). *Understanding Travel Behavior*, Gower, Aldershot, UK.

Ji, E.Z. and Chen, A. (2003). Numerical comparisons of deterministic and stochastic route choice models. *Journal of the Eastern Asia Society for Transportation Studies*, 5, 1183-1193.

Jorgensen, B. (1982). Statistical properties of the generalized inverse Gaussian distribution. In D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, J. Kiefer and K. Krickeberg, eds., *Lecture Notes in Statistics*, Vol. 9, Springer-Verlag, New York.

Jörnsten, K. and Nguyen, S. (1983). *Estimation of an OD trip matrix from network data: dual approaches*, Monograph, Linköping University, Department of Mathematics, Linköping, Sweden.

Karlis, D. (2001). A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Statistical Modelling*, 1(4), 305-318.

Kass, R.E. and Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773-795.

Kass, R.E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435), 1343-1370.

Kawakami, S., Lu, H. and Hirobata, Y. (1992). Estimation of origin-destination matrices from link traffic counts considering the interaction of the traffic modes. *Papers in Regional Science*, 71(2), 139-151.

Knoop, V.L., Snelder, M. and Van Zuylen, H.J. (2007). Comparison of link-level robustness indicators. In *Proceedings of the Third International Symposium on Transportation Network Reliability*, Hague, the Netherlands.

Kraft, G. (1968). *Demand for intercity passenger travel in the Washington-Boston corridor*, Project Report, Systems Analysis and Research Corporation, Boston, Massachusetts.

Kuhn, K.D. and Madanat, S.M. (2005). Model uncertainty and the management of a system of infrastructure facilities. *Transportation Research Part C*, 13(5-6), 391-404.

Kutner, M., Nacthsheim, C. and Neter, J. (2004). *Applied Linear Regression Models*, 4th ed., McGraw Hill/Irwin, New York.

Lambert, D. (1992). Zero-inflated Poisson regression, with applications to defects in manufacturing. *Technometrics*, 34(1), 1-24.

Laud, P. and Ibrahim, J. (1995). Predictive model selection. *Journal of the Royal Statistical Society B*, 57(1), 247-262.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 15(3), 209-225.

Lawson, A.B., Browne, W.J. and Rodeiro, C.L.V. (2003). *Disease mapping with WinBUGS and MLwiN*, John Wiley and Sons, Chichester, UK.

LeBlanc, L., Morlok, E. and Pierskalla, W. (1975). An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research*, 9(5), 309-318.

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B*, 58(4), 619-678, (with discussion).

Lee, Y. and Nelder, J.A. (2004). Conditional and marginal models: another view. *Statistical Science*, 19(2), 219-228.

Li, B. (2005). Bayesian Inference for origin-destination matrices of transport networks using the EM algorithm. *Technometrics*, 47(4), 399-408.

Li, B. (2009). Markov models for Bayesian analysis about transit route origin-destination matrices. *Transportation Research Part B*, 43(3), 301-310.

Li, R. and Rose, G. (2011). Incorporating uncertainty into short-term travel time predictions. *Transportation Research Part C*, 19(6), 1006-1018.

Li, H. and Wu, J. (2006). Uncertainty analysis in ecological studies: an overview. In J. Wu, K.B. Jones, H. Li, and O. Loucks, eds., *Scaling and uncertainty analysis in ecology: methods and applications*. Springer, Dordrect, Netherlands, pp. 43-64.

Lin, D., Eluru, N., Waller, S.T. and Bhat, C.R. (2008). Integration of activity-based modeling and dynamic traffic assignment. *Transportation Research Record*, 2076, 52-61.

Lo, H.P., Zhang, N. and Lam, W.H.K. (1996). Estimation of an origin-destination matrix with random link choice proportions: a statistical approach. *Transportation Research Part B*, 30(4),309-324.

Lord, D., Manar, A. and Vizioli, A. (2005). Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. *Accident Analysis and Prevention*, 37(1), 185-199.

Maher, M.J. (1983). Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. *Transportation Research Part B*, 17(6), 435-447.

Manheim, M.L. (1979). *Fundamentals of Transportation Systems Analysis*, MIT Press Cambridge, Massachusetts.

Mahmassani, H.S., Chang, G.L. and Herman, R. (1986). Individual decisions and collective effects in a simulated traffic system. *Transportation Science*, 20(4), 258-271.

Mahmassani, H.S. and Chang, G.L. (1987). On boundedly rational user equilibrium in transportation systems. *Transportation Science*, 21(2), 89-99.

Marconi, D., Simma, A. and Gindraux, M. (2004). The Swiss microcensus 2005: an international comparison on travel behaviour. In *4th Swiss Transport Research Conference*, Monte Verità, Ascona, Switzerland.

Marzano, V., Papola, A. and Simonelli, F. (2009). Limits and perspectives of effective O-D matrix correction using traffic counts. *Transportation Research Part C*, 17(2), 120-132.

Matas, A., Raymond, J.L. and Ruiz, A. (2012). Traffic forecasts under uncertainty and capacity constraints. *Transportation*, 39(1), 1-17.

Mathew, T.V. and Rao, K.V.K. (2007). *Introduction to Transportation Engineering*, NPTEL Program, India.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall/CRC, London.

McKenzie, B. and Rapino, M. (2011). Commuting in the United States: 2009. In *American Community Survey Reports*, ACS-15, U.S. Census Bureau, Washington, DC.

McNeney, B. and Petkau, J. (1994). Overdispersed Poisson regression models for studies of air pollution and human health. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 22(4), 421-440.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. and Teller, A.H. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087-1092.

Mirchandani, P. and Soroush, H. (1987). Generalized traffic equilibrium with probabilistic travel times and perceptions. *Transportation Science*, 21(3), 133-152.

Mollié, A. (1996). Bayesian mapping of disease. In W. Gilks, S. Richardson and D. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC, London, pp. 359-376.

Morris, C.N. (1983). Natural exponential families with quadratic variance functions: statistical theory. *The Annals of Statistics*, 11, 515-529.

Nagelkerke, N.J.D. (1991). A Note on a general definition of the coefficient of determination. *Biometrika*, 78(3), 691-692.

Ng, M.W., Zhang, Z. and Waller, S.T. (2011). The price of uncertainty in pavement infrastructure management planning: an integer programming approach. *Transportation Research Part C*, 19(6), 1326-1338.

Nikoloulopoulos, A.K. and Karlis, D. (2008). On modeling count data: a comparison of some well-known discrete distributions. *Journal of Statistical Computation and Simulation*, 78, 437-457.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*, 1st ed., John Wiley and Sons, New York.

Oakley, J.E. and O'Hagan, A. (2007). Uncertainty in prior elicitations: a nonparametric approach. *Biometrika*, 94(2), 427-441.

O'Brien, R.M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41(5), 673-690.

Odoki, J.B., Kerali, H.R. and Santorini, F. (2001). An integrated model for quantifying accessibility-benefits in developing countries. *Transportation Research Part A*, 35(7), 601-623.

Ortúzar, J. de D. and Willumsen, L.G. (2001). *Modelling Transport*, 3rd ed., John Wiley and Sons, Chichester, UK.

Ory, D.T. (2007). *Structural equation modeling of relative desired travel amounts*. PhD dissertation. University of California, Davis, Department of Civil and Environmental Engineering, Davis, CA.

Parthasarathi, P. and Levinson, D. (2010). Post-construction evaluation of traffic forecast accuracy. *Transport Policy*, 17(6), 428-443.

Patriksson, M. (1994). *Traffic Assignment Problems: Models and Methods. Topics in transportation*, V.N.U. Science Press, Utrecht.

Pendyala, R.M., Kitamura, R., Kikuchi, A., Yamamoto, T. and Fujji, S. (2005). FAMOS: Florida activity mobility simulator. In *84th Annual Meeting of Transportation Research Board*, Washington, DC.

Perrakis, K., Karlis, D., Cools, M., Janssens, D. and Wets, G. (2011). A Bayesian approach for modeling origin-destination matrices. In *90th Annual Meeting of Transportation Research Board*, Washington, DC.

Perrakis, K., Karlis, D., Cools, M., Janssens, D., Vanhoof, K. and Wets, G. (2012a). A Bayesian approach for modeling origin-destination matrices. *Transportation Research Part A*, 46(1), 200-212.

Perrakis, K., Cools, M., Karlis, D., Janssens, D., Kochan, B., Bellemans, T. and Wets, G. (2012b). Quantifying input-uncertainty in traffic assignment models. In *91st Annual Meeting of Transportation Research Board*, Washington, DC.

Perrakis, K., Karlis, D., Cools, M., Janssens, D. and Wets, G. (2012c). Poisson mixture regression for Bayesian inference on large over-dispersed transportation origin-destination matrices. In *27th International Workshop on Statistical Modelling*, Prague, Czech Republic.

Perrakis, K., Cools, M., Karlis, D., Janssens, D. and Wets, G. (2012d). Bayesian inference on traffic assignment and network congestion given demand uncertainty. Submitted.

Perrakis, K., Karlis, D., Cools, M. and Janssens, D. (2012e). The Poisson-inverse Gaussian for Bayesian inference on transportation origin-destination matrices. Submitted.

Perrakis, K., Cools, M., Karlis, D. and Janssens, D. (2012f). A comparative analysis between SUE and DUE assignment given demand uncertainty. Submitted.

Poole, D. and Raftery, A.E. (2000). Inference for deterministic simulation models: the Bayesian melding approach. *Journal of the American Statistical Association*, 95(452), 1244-1255.

R Development Core Team, (2008). *R: A Language and Environment for Statistical Computing*, The R Foundation for Statistical Computing, Vienna, Austria, <http://www.r-project.org/>.

Rabe-Hesketh, S., Toulopoulou, T. and Murray, R. (2001). Multilevel modeling of cognitive function in schizophrenic patients and their first degree relatives. *Multivariate Behavioral Research*, 36(2), 279-298.

Raftery, A.E. and Lewis, S. (1992). How many iterations in the Gibbs sampler? In J. Bernando, J. Berger, A. Dawid and A. Smith, eds., *Bayesian Statistics*, Vol. 4, pp. 763-773.

Rasouli, S. and Timmermans, H.J.P. (2012). Uncertainty in travel demand forecasting models: literature review and research agenda. *Transportation Letters*, 4(1), 55-73.

Richardson, A.J., Ampt, E.S. and Meyburg, A.H. (1995). *Survey Methods for Transport Planning*, Eucalyptus Press, Melbourne.

Richmond, J. (2001). A whole-system approach to evaluating urban transit investments. *Transport Reviews*, 21(2), 141-179.

Rubin, D.B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistic*, 12(4), 1151-1172.

Roorda, M.J., Miller, E.J. and Habib, K. (2008). Validation of TASHA: a 24-hour activity scheduling microsimulation model. *Transportation Research Part A*, 42(2), 360-375.

Rose, G., Daskin, M. and Koppelman, F. (1988). An examination of convergence error in equilibrium traffic assignment models. *Transportation Research Part B*, 22(4), 261-274.

Sapatinas, T. (1995). Identifiability of mixtures of power-series distributions and related characterizations. *Annals of the Institute of Statistical Mathematics*, 47(3), 447-459.

Schneider, M. (1959). Gravity models and trip distribution theory. *Papers in Regional Science*, 5(1), 51-56.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.

Ševčíková, H., Raftery, A.E. and Waddell, P.A. (2007). Assessing uncertainty in urban simulation using Bayesian melding. *Transportation Research Part B*, 41(6), 652-669.

Shaban, S.A. (1981). On the discrete Poisson-inverse Gaussian distribution. *Biometrical Journal*, 23(3), 297-303.

Shaban, S.A. (1988). Poisson-lognormal distributions. In E.L. Crow and K. Shimizu, eds., *Log-normal distributions: theory and applications*, Marcel Dekker, New York, pp. 195-210.

Sheffi, Y. (1985). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice Hall, England Cliffs, New Jersey.

Sheffi, Y. and Powell, W. (1981). A comparison of stochastic and deterministic traffic assignment over congested networks. *Transportation Research Part B*, 15(1), 53-64.

Simini, F., González, M.C., Maritan, A. and Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature*, 484(7392), 96-100.

Siu, B.W.Y. and Lo, H.K. (2008). Doubly uncertain transportation network: degradable capacity and stochastic demand. *European Journal of Operational Research*, 191(1), 166-181.

Sohn, K. and Kim, D. (2010). Zonal centrality measures and the neighborhood effect. *Transportation Research Part A*, 44(9), 733-743.

Soroush, H. (1984). *Routes and flows through stochastic networks*. PhD dissertation. Rensselaer Polytechnic Institute, Troy, New York.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B*, 64(4), 583-639.

Spiess, H. (1987). A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B*, 21(5), 395–412.

Spiess, H. (1990). *A gradient based approach for the OD matrix adjustment problem*, Report 693, University of Montreal, Centré de recherché sur les transports (CRT), Montreal, Canada.

Stasinopoulos, D. and Rigby, R. (2007). Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, 23(7).

Stein, G.Z. and Juritz, J.M. (1988). Linear models with an inverse Gaussian Poisson error distribution. *Communications in Statistics - Theory and Methods*, 17, 557-571.

Stein, G.Z., Zucchini, W. and Juritz, J.M. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, 82(399), 938-944.

Stopher, P.R. and Greaves, S.P. (2007). Household travel surveys: where are we going? *Transportation Research Part A*, 41(5), 367-381.

Stouffer, S.A. (1940). Intervening opportunities: a theory relating mobility and distance. *American Sociological Review*, 5(6), 845-867.

Taylor, M.A.P. (2008). Critical transport infrastructure in urban areas: impacts of traffic incidents assessed using accessibility-based network vulnerability analysis. *Growth and Change*, 39(4), 593-616.

Taylor, M.A.P. and D'Este, G.M. (2004). Critical infrastructure and transport network vulnerability: developing a method for diagnosis and assessment. In *Proceedings of Second International Symposium on Transportation Network Reliability*, Christchurch and Queenstown, New Zealand.

Taylor, M.A.P., Sekhar, S.V.C. and D'Este, G.M. (2006). Application of accessibility based methods for vulnerability analysis of strategic road networks. *Networks & Spatial Economics*, 6(3), 267–291.

Tebaldi, C. and West, M. (1996). Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association*, 93(442), 557-576, (with discussion).

Thall, P.F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics*, 44(1), 197-209.

Thomas, R. (1991). *Traffic Assignment Techniques*, Avebury Technical, Aldershot.

Timms, P. (2001). A philosophical context for methods to estimate origin destination trip matrices using link counts. *Transport Reviews*, 21(3), 269-301.

TRB Travel Survey Manual, (2012). *The On-Line Travel Survey Manual: A Dynamic Document for Transportation Professionals*, Transportation Research Board's Travel Survey Methods Committee, <http://www.travelsurveymanual.org/>.

Ukkusuri, S. V. and Waller, S. T. (2006). Single-point approximations for traffic equilibrium problem under uncertain demand. *Transportation Research Record*, 1964, 169-175.

Ukkusuri, S. V., Mathew, T. V. and Waller, S. T. (2007). Robust transportation network design under demand uncertainty. *Computer Aided Civil and Infrastructure Engineering*, 22(1), 6-18.

Ukkusuri, S. V. and Patil, G. (2009). Multi-period transportation network design under demand uncertainty. *Transportation Research Part B*, 43(6), 625-642.

Vaida, F. and Blanchard, S. (2005). Conditional Akaike criterion for mixed-effects models. *Biometrika*, 92(2), 351-370.

Van de Ven, R. and Weber, N.C. (1995). Log-linear models for mean and dispersion in mixed Poisson regression models. *Australian Journal of Statistics*, 37(2), 205-216.

Van Zuylen, H.J. and Willumsen, L.G. (1980). The most likely trip matrix estimated from traffic counts. *Transportation Research Part B*, 14(3), 281-293.

Vardi, Y. (1996). Network tomography: estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association*, 91(433), 365-377.

Vosha, P., Petersen, E. and Donnelly R. (2002). Microsimulation in travel demand modeling: lessons learned from the New York best practice model. *Transportation Research Record*, 1805, 68-77.

Walker, J. (2005). Making household microsimulation of travel and activities accessible to planners. *Transportation Research Record*, 1931, 38-48.

Waller, S.T., Schofer, J. and Ziliaskopoulos, A. (2001). Evaluation with traffic assignment under demand uncertainty. *Transportation Research Record*, 1771, 69-74.

Wardrop, J. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II*, 1(36), 352-362.

Wermuth, M.J. (1981). Effects of survey methods and measurement techniques on the accuracy of household travel-behaviour surveys. In P.R. Stopher, A.H. Meyburg and W. Brög, eds., *New Horizons in Travel Behaviour Research*, D.C. Health and Company, Lexington, Massachusetts, pp. 523-541.

West, M. (1994). Statistical inference for gravity models in transportation flow forecasting. Discussion Paper 94-40, ISDS, Duke University.

Williams, H.C.W.L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9(3), 285-344.

Willmot, G.E. (1990). Asymptotic tail behaviour of Poisson mixtures with applications. *Advances in Applied Probability*, 22(1), 147-159.

Wilson, A.G. (1970). *Entropy in Urban and Regional Modelling*, Pion Ltd, London.

Wilson, A.G. (1974). *Urban and Regional Models in Geography and Planning*, John Wiley and Sons, London.

Wright, D.E. (1986). A Note on the construction of highest posterior density intervals. *Applied Statistics*, 35(1), 49-53.

Xue, D. and Deddens, J.A. (1992). Overdispersed negative binomial regression models. *Communications in Statistics - Theory and Methods*, 21, 2215-2226.

Yang, H., Iida, Y. and Sasaki, T. (1994). The equilibrium-based origin-destination matrix estimation problem. *Transportation Research Part B*, 28(1), 23-33.

Yao, E. and Morikawa, T. (2005). A study of an integrated intercity travel demand model. *Transportation Research Part A*, 39(4), 367-381.

Zeger, S.L. and Karim, M.R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, 86(413), 79-86.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In P. Goel and A. Zellner, eds., *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*, North-Holland, Amsterdam, pp. 233-243.

Zhang, L. (2007). Developing a positive approach to travel demand forecasting: theory and model. *Transportation and Traffic Theory*, 17, 136-150.

Zhang, L. (2011). Behavioral foundation of route choice and traffic assignment. *Transportation Research Record*, 2254, 1-10.

Zhou, M. and Sisiopiku, V.P. (1997). Relationship between volume-to-capacity ratios and accident rates. *Transportation Research Record*, 1581, 47-52.