Modeling office firm dynamics in an agent-based micro simulation framework : methods and empirical analysis

DOI:
10.6100/IR733478

Document status and date:
Published: 01/01/2012

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Modeling Office Firm Dynamics in an Agent-Based Micro Simulation Framework

Methods and Empirical Analyses

Gustavo Garcia Manzato

/ Department of the Built Environment

bouwstenen 157

Modeling Office Firm Dynamics in an Agent-Based Micro Simulation Framework

Methods and Empirical Analyses

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. C.J. van Duijn, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op dinsdag 12 juni 2012 om 16.00 uur

door

Gustavo Garcia Manzato

geboren te Bauru, Brazilië

Dit proefschrift is goedgekeurd door de promotor:

prof.dr. H.J.P. Timmermans

Copromotoren:
dr. T.A. Arentze
en
dr. D.F. Ettema

Modeling Office Firm Dynamics in an Agent-Based Micro Simulation Framework

Methods and Empirical Analyses

Samenstelling promotiecommissie:

prof.ir. E.S.M. Nelissen, voorzitter
prof.dr. H.J.P. Timmermans, Technische Universiteit Eindhoven, promotor
dr. T.A. Arentze, Technische Universiteit Eindhoven, copromotor
dr. D.F. Ettema, Universiteit Utrecht, copromotor
prof.dr. K. Frenken, Technische Universiteit Eindhoven
prof.dr. F.G. van Oort, Universiteit Utrecht
prof.dr. F. Witlox, Universiteit Gent

A catalogue record is available from the Eindhoven University of Technology Library

ISBN 978-90-6814-640-0
NUR 955

Cover design by Ton van Gennip, Tekenstudio, Bouwkunde
Printed by the Eindhoven University Press, Eindhoven, The Netherlands
Published as issue 157 in de Bouwstenen series of the faculty of the Built Environment of the Eindhoven University of Technology

Acknowledgements

This thesis is the result of almost five years that I spent on my PhD research project carried out at the Urban Planning Group of the Eindhoven University of Technology. While all the details involving the theoretical approaches, modeling techniques and findings obtained will be described in the following pages of this book, this section is dedicated to acknowledge the "backstage" of this research project.

One of the most frequent questions that people used to ask me was what I was doing by leaving Brazil and moving to the Netherlands. I guess that for most of these people this question was related to the differences in culture and climate (especially) between these two countries and how I was coping with that. However, these issues did not bother me at all, as I always had a determination to live abroad during some period of my life, experiencing something different. When I first heard about the possibility to move to the Netherlands to do a PhD, I felt that it was not only my chance to have my determination fulfilled, but also to work on something that would make me grow professionally. So I took this opportunity and the result of this decision was amazing. To be living in the Netherlands allowed me to get in touch with people from different nationalities and cultures, which expanded my perception of the world. Also, I could explore parts of the "Old World" by visiting most of the highlights that I had seen, until then, only in books of History and Geography, for example. Obviously this was a very enriching experience.

If this experience was beyond my expectations from a personal point of view, so was it professionally. I could acquire some knowledge in the most recent developments in the area of urban and transportation planning, having the opportunity to work in one of the leading groups of the world. Therefore, I would like to thank Professor Harry Timmermans, my supervisor, for having given me this chance and for investing in my potential. It was a great honor to work with him and to be part of his group. I would like to thank Dr. Theo Arentze as well, my first co-supervisor, for his endless support and patience. His motivation and moments of inspiration were crucial to my path toward the successful conclusion of this research project. I also would like to thank my second co-supervisor, Dr. Dick Ettema, for his ideas, suggestions and contributions.

My professional achievements should be also attributed to some special people. First, I would like to thank Dr. Antônio Nélson Rodrigues da Silva for also investing in my potential and conducting me in the research "world" since I was his student in

Brazil. I am also grateful for his support in this process of going abroad for my PhD after I finished my Master's degree under his supervision. Yet, I am glad for our continuous research collaborations and for his support in my future professional plans. Second, I would like to thank Dr. Eliane Viviani. If I follow the path of my academic career backward, she is there at the starting point, when I was still an undergraduate student. I am glad for having her as the supervisor of my undergraduate scientific initiation projects, triggering my interest for research activities. Finally, I would like to thank Dr. Rui António Rodrigues Ramos for his support in several ways and periods, and Dr. Carlos Felipe Grangeiro Loureiro for his interest in my research activities.

My stay in the Urban Planning group was very nice, thanks to all who have warmly welcomed me since my first day and have kept a friendly work environment every day since then. I am very grateful to Mandy van de Sande - van Kasteren, Peter van der Waerden and Freek van der Heijden, who were the first people that I met and gave me all the support for setting me up in a smooth way. Mandy was always there whenever I needed some help with the daily administrative issues involving the department, and Peter was always of assistance regarding technical issues involving my research project. I also would like to thank Linda van de Ven, Annemiek Engelen and Ingrid Dekkers - de Bruijn for their endless support regarding secretarial issues, making all the necessary efforts for an effective assistance. A special thanks to Linda van de Ven for helping me out with the procedures involving the preparation of my defense. In addition, I would like to thank Dr. Astrid Kemperman, Aloys Borgers, Elle Abzach and Sophie Rousseau for the friendly interaction within the department, Mark Wolffe for the assistance with some of my bibliographic searches, and Carol van Iperen for the HR support whenever I needed it.

During this period I could obviously get in touch with new colleagues who were/are also aboard this PhD experience. I am glad that we had numerous opportunities to know and help each other, building a very nice friendship. It is impossible to name everybody here, but for sure they all had an important participation in my life. This applies especially to my colleagues from the Urban Planning Group. However, I would like to thank a few people with whom I had some good times together: Pauline van den Berg (and Joep), Anastasia Moiseeva (and Dominique), Oliver Horeni, Nicole Ronald, Linda Nijland (and Pieter), Marloes Verhoeven, Brano Glumac, Dave Havermans, and Gamze Dane (and Erik).

Even though I was far away from home, I am glad that I could have the company of Brazilian friends around, so that I felt less homesick. I would like to thank

Anna Beatriz Grigolon and Ivan Elias Oliveira, Ana Paula Melo, Daniel Cóstola and Carolina Castanheira, and Aida Pontes and Nazareno Andrade for all the moments that we spent together, trying to ease my (our) stay in the Netherlands in relation to cultural and (mainly) gastronomic differences. A special thanks to Anna and Ivan who "adopted" me and offered me a place to stay in their house during my last weeks in the Netherlands before returning to Brazil.

If on the one hand this experience abroad was very fruitful for me at so many levels, on the other hand it was not obviously easy to my loved ones who missed me when I left home in 2007. However, they always gave me great support. I would like to thank my parents Lucia and Valdomiro for their unconditional love and for always making the possible (and sometimes even the impossible) in relation to my fulfillments; my brothers Daniel and Marcelo for our true friendship; my little sisters Maira, Maísa and Aparecida for our joy at all times that we are together; my aunt and godmother Josefa for always motivating me to see the world from above; and my grandparents Maria and Henrique, and Gema and José for always wishing me all the best in my life. A very special thanks to Eliana, who I married in 2010, for her everlasting support, understanding, patience and love during this journey. I also would like to thank my parents-in-law Verginia and Nelson for always wishing me all the best as well.

Last but not least, I would like to thank the AlBan program (the European Union Program of High Level Scholarships for Latin America, scholarship E07D400519BR) and the RGI program (Ruimte voor Geo-Informatie) for the financial support. Also, the DTZ company and all other agencies, organizations and institutions that provided the numerous datasets for this research project.

Contents

List of figures

List of tables

1

Introduction

1.1 Motivation

Office firms represent a large share of economic activities nowadays, especially in the sector of professional services. The crescent urbanization process along with the specialization of employment, which is sometimes caused by higher levels of people's education, have contributed to this effect. Although the manufacturing labor and the industry sector are still important bases of economy, a shift to the service sector can be observed, especially in cities. These contain many interactions among people, jobs and services usually associated with urban areas, such as: schools, labor locations, culture and recreation infrastructures, shopping facilities, among others. These interactions consequently result in travel patterns that influence the transportation infrastructures. Traditionally, the specific relationships between transportation and land use have been examined in the framework of LUTI (integrated land use-transportation) models (Mackett, 1985). These were originally based on aggregated frameworks, but the field is moving toward a more disaggregated approach, using micro simulation and agent-based models.

In line with the above, the specific interest in firm development has been examined by traditional location theories, starting with principles of profit maximization in isotropic space, which were gradually replaced with psychological theories, focusing on behavioral models for business activity. However, these theories did not fully account for the fact that location preferences are also related to the firm characteristics and their lifecycle. An approach claiming the relevance of the firm's structural factors emerged. It led to the development of a firm demographic approach, conceptualizing the changes of firm dynamics by an evolutionary cycle of starting-up, finding a location to establish their business, growing or declining, probably relocating and going out of business. The firm demographic approach has found strong linkages with multi-agent systems and micro simulation, which are able to accommodate the behaviorally richer concepts of firm dynamics. Various applications can be found in the literature, e.g., van

Wissen (2000); Khan et al. (2002a and 2002b); Moeckel (2005 and 2009); Maoh and Kanaroglou (2005); de Bok and Bliemer (2006a and 2006b); de Bok (2007); Bodenmann (2011a and 2011b); and Schirmer et al. (2011).

Regardless this progress, avenues for further investigations are open. Especially considering office firms, the literature is relatively scarce, needing studies that bring advances in the field. While existing applications consider firms in general, a first question addressed here regards the focus on office firms and their related dynamics. The examination of these dynamics using the firm demography concept applied to office firms constitutes one of the contributions of this research project. As for the developments of LUTI models, the underlying approach relies on the idea that office firm dynamics are influenced by the urban environment. This is not new compared to previous studies, but a contribution aimed here is to extend the set of location attributes and test them in such a LUTI framework. For example, the inclusion of other transportation infrastructures, such as: airports, train station, public transportation, among others, as opposed to the simple accessibility indicator based on the proximity to highways. Also, the consideration of more continuous measures to calculate accessibility, as opposed to some qualitative or discrete proximity measures, which are usually applied, constitutes the objectives of this research project. These considerations apply for other location attributes defined here, as will be detailed in upcoming chapters.

Along with the above, another specific contribution is to empirically examine alternative methods to compose an agent-based modeling approach to simulate the evolution of office firms in time and space. To this end, a set of statistical/econometric models is used to investigate the relationships between specific firm demographic processes and the urban environment, using very detailed nationwide data from The Netherlands. This is also not new compared to previous studies, but the need for constant further improvements reported in such existing models motivated this research project. Although the final product of this research does not result in a fully operational simulation model for office firm demography, the keystones for its implementation will be structured in detail. Hence, this thesis is far from bringing an end to the topic, but it is expected to provide better understandings about office firm dynamics and their relationships in a LUTI framework through the findings obtained here.

1.2 Thesis outline

The thesis is comprised of 11 chapters. This introductory chapter starts with the discussion on the motivations that resulted in this research project and presents its objectives, as detailed in the previous section. Chapter 2 brings a discussion on studies about firm dynamics, starting with the evolution of LUTI models and derived investigations on firm development. A further motivation for studying office firms is provided. It should be observed that this chapter does not comprise a comprehensive and detailed review of LUTI models and its various derived applications. Instead, it will focus on basic concepts and important developments in firm studies that have a direct relevance to the object of study in this research project.

Chapter 3 then introduces a conceptual framework based on the firm demography approach to model office firm dynamics. Given the urban and transportation planning perspective of this research, this chapter addresses the use of the most recent LUTI modeling approaches, based on multi-agent systems and micro simulation for examining office firm demographic processes. However, it should be observed that the focus is on developing the various related components, whereas a full simulation framework is not implemented. In order to test the proposed framework, the data obtained is presented in chapter 4. These regard both firm micro data and several datasets about the urban environment. The chapter also discusses the source of these various data and the several derived attributes.

From chapter 5 onwards, the details of the various components related to firm demographic processes are depicted. These chapters follow a similar structure, based on an introductory section that discusses the basics of the demographic process addressed. A modeling approach is then presented, based on a specific statistical/econometric model. Next, an empirical application is delineated followed by the results and analyses obtained. At the end, some concluding remarks are drawn. In sum, Chapter 5 presents the start-up model. Chapter 6 introduces the carrying capacity model. It is not a demographic process *per se*, but regards a component to account for the capacity of a given zone in terms of its infrastructure capabilities in relation to possible firm-related activities (start-up, growth, etc). The location choice model is presented in chapter 7, followed by the growth model introduced in chapter 8. In chapter 9 a model that examines the process of firms going out of business is presented, here referred to as the closure model. Finally, chapter 10 investigates the relocation process. Ideally, the introduction of the relocation model should come before chapter 9 (closure model), following an expected firm lifecycle. However, given that one of the techniques adopted to analyze the relocation patterns involves duration models and these are better presented when discussing the closure process, this decision to invert the chapter order was taken.

At the end of this thesis, chapter 11 brings the main conclusions obtained in light of the empirical findings of this research, reflects on the limitations observed during its development and discusses some recommendations for further investigations. The references cited in this thesis are listed after chapter 11, followed by an appendix, author and subject indexes, a summary, and ending with some information about the author of this thesis.

2

Studies on firm dynamics

2.1 Introduction

This chapter presents a discussion about modeling firm dynamics and related empirical findings reported in the literature. For decades, there has been an overwhelming interest in the development of economic activities in disciplines such as (economic) geography and regional science. Recently, in urban and transportation research, the interest has also increased given the strong interaction between firms and mobility in terms of congestion, accessibility and traffic flows. The issue has been examined in the context of Land Use-Transport Interaction (LUTI) models, aiming at trying to explain the observed spatial distribution of economic activities in a plausible way.

The objective of this chapter is to outline the evolution of the investigations in this field. It starts with a discussion about the development of LUTI models, from the initial generation based on aggregate data and principles of gravitation and entropy-maximization, to the most recent stages built on agent-based modeling and micro simulation concepts. Given that the focus of this research is on office firm dynamics, the development of the investigations in this specific field is also addressed. In line with the aim of this research project, the chapter brings together the use of agent-based models and micro simulation on the investigation of office firm dynamics.

2.2 LUTI models

In urban and transportation planning, the interaction between land use and transportation is an important concern. Land use patterns resulting from the spatial allocation of urban functions (e.g., residences, shopping and entertainment facilities, job locations, health and education needs, etc.) generate trips that people need to make in order to fulfill their needs. This generates traffic flows that will influence mobility and accessibility. However, this interaction also occurs the other way around, when properties of the transportation infrastructure determine the location of households, firms, schools, hospitals, parks, shops, etc. The literature shows a long tradition in the development of integrated land use-transport (LUTI) models addressing this interaction and attempting to explain it in a plausible way. Timmermans (2003) presents an extensive discussion on the development of LUTI models. Further updates can also be found in Koomen and Stillwell (2007); Iacono et al. (2008); and Wu and Silva (2010) for a review of approaches using artificial intelligence. Basically, three generations regarding the development of LUTI models can be identified. The key concepts in each of them are outlined hereafter.

The first generation of models started in the 1960s and was based on aggregate data, concepts of equilibrium, and gravitation principles, namely spatial-interaction based models. Economic and demographic development was described as the outcome of allocation processes at the zonal level and the specification of travel patterns as interactions between zones. The Lowry-Garin model initiated this framework in the research community. It was further elaborated resulting in quantitative improvements that motivated the emergence of several applications. Some examples comprise: TOMM (Time Oriented Metropolitan Model – Crecine, 1964 and 1968); PLUM (Projective Land Use Model – Goldner, 1971; Goldner et al., 1972); IRPUD (Wegener, 1982a, 1982b and 1983); LILT (Leeds Integrated Land-Use Model – Mackett, 1983); ITLUP (Integrated Transportation Land Use Package); DRAM (Disaggregated Residential Allocation Model); EMPAL (Employment Allocation Model); and METROPILUS (Metropolitan Integrated Land Use System), these last four models developed by Putman (1983).

Gradually, these aggregate spatial interaction models were replaced by principles of choice behavior, based on the utility theory. This second generation comprised, namely, the utility maximizing multinomial logit (MNL) models. These have their foundations in the micro economic consumer theory, which assumes that individuals derive a certain utility for each product in a choice set, depending on their preferences, related prices and available budget. Translating this notion to the urban and transportation context, the probability of an individual (household, firm, etc.) selecting a particular location is seen as the result of maximizing the utility derived from the various characteristics of a number of locations. Examples include CATLAS (Anas, 1982, 1983a and 1983b); TRANUS (de la Barra et al., 1984); IMREL (Anderstig and Mattson, 1991, 1992 and 1998; Boyce and Mattsson, 1999); MEPLAN (Echenique, 1994; Echenique et al., 1990; Hunt 1994); METROSIM (Anas, 1994); BASS/CUF (Landis, 1994; Landis and Zhang, 1998a and 1998b); MUSSA (Martinez, 1997); UrbanSim (Waddell, 2000 and 2002; Waddell et al., 2003); TILT (Eliasson, 2000; Eliasson and Mattsson, 2001); Uplan (Johnston et al., 2003); DELTA (Simmonds, 2005); and PECAS (Abraham et al., 2005).

The developments in the field of modeling land use-transport interaction has had contributions from cellular automata models, motivated by the theory of self-organizing systems (e.g., White and Engelen, 1993a and 1993b; Batty and Xie, 1994; Cecchini, 1996; Batty et al., 1997; Clarke et al., 1997). These have been commonly applied to simulate urban growth and urban change. However, the transport component of these models has been weak, in the sense that a transport network is usually considered, but traffic flows are not simulated. Moreover, the cells did not comprise decision-makers. The need for a framework able to capture the behavior of relevant actors in the urban context, representing real individual decision making processes and accounting for heterogeneous preferences has therefore been advocated. In particular, the introduction of activity-based models and micro simulation methods was proposed since the end of the 1990s, characterizing the third generation of LUTI models (for a review, see Timmermans et al., 2002; Miller et al., 2004; Batty, 2005).

In these activity-based micro simulation models, land use and transportation patterns are seen as the result of individual's activities (e.g., people, households, developers, businesses, corporations, etc.) and related decisions (e.g., to buy a car, to open a firm, to travel to work, etc.) within the urban context. These are simulated at a very disaggregate level, taking into account the smallest subject possible (Moeckel, 2009). The key concern described by such micro simulation models is the behavior of the actors involved in such LUTI processes. Examples include: ILUTE (Miller and Salvini, 1998); RAMBLAS (Veldhuisen et al., 2000); ILUMASS (Moeckel et al., 2002); and PUMA (Ettema et al., 2007). It should be noted that most of these model systems have to date only been partially developed or represent limited land use.

In order to better represent these individual's activities and related decisions within the urban environment, advances in LUTI models also indicated a path toward multi-agent systems (see, for example, Bonabeau, 2002). Multi-agent models represent individual actors as autonomous agents that behave and have an identity. It means that such agents are able to make decisions within the system, based on the attributes that they have. Processes such as competition, negotiation and bidding among multiple actors can be also incorporated in multi-agent systems (see, for example, Arentze and Timmermans, 2003; Devisch et al., 2006 for applications in urban planning, and Huhns and Stephens, 1999; Sandholm, 1999 for explorations in other domains). Another important benefit of agent-based modeling regards the fact that it captures emergent phenomena. It allows the design of individual entities (i.e., households, firms) that will interact among each other following simple rules, generating complex behavior from the bottom up (for further reading, see, for example, Epstein and Axtell, 1996; Axelrod, 1997).

Studies attempting to incorporate such agent-based micro simulation modeling framework in LUTI systems include, for example, Hunt (2002); Miller et al. (2004); Salvini and Miller (2005); Saarloos et al. (2005 and 2008); Wagner and Wegener (2007); Ettema et al. (2007); Katoshevski-Cavari et al. (2010); and Arentze et al. (2010). These comprise frameworks for the simulation of entire systems, addressing the interactions of all actors in the urban environment (e.g., households, firms, etc.) and the related travel behavior. It should be noted, however, that some of these interactions are indirect, since agents respond to aggregate system characteristics. As stated by Miller et al. (2004), agent-based modeling "provides an extremely efficient, effective and natural way of both conceptualizing and implementing complex, dynamic, disaggregate models of human decision-making". Theoretically, it has a strong appeal for LUTI modeling, capable to represent the various actors involved, along with their related processes. Nevertheless, the development of empirical applications still finds limitations, especially regarding the need of detailed, updated and available data. But efforts have been made and these are still an ongoing process, as reported by this brief summary of the trajectory of LUTI models.

Overall, as can be noted from the above, the research agenda is currently focused on the multi-agent system approach, given its capability to incorporate and represent richer activity behavioral concepts in a micro simulation framework. However, as opposed to investigations in the household behavior (e.g., Devisch, 2008; Anggraini, 2009), the firm domain has received much less attention in multi-agent modeling applications. Hence, the development of approaches to investigate firm dynamics comprises the topic of interest of this research project. In line with this, the next section will present an overview of studies dealing with this specific topic, as a way to demonstrate the achievements in the field and to indicate possible contributions that could be made.

2.3 Investigations on firm development

Firms in general have an important influence on the development of the urban environment. As source of employment, they determine the locations of jobs, to which people travel to work. Firms provide services to the population, resulting in trips related to the consumption of such services. Firms require the delivery of both raw materials and manufactured products, also generating freight transport flows in the network. In sum, firm development has a strong interaction with mobility. However, differently from individual persons and households, firms present an inherent heterogeneity that makes modeling their behavior rather difficult, i.e., a larger variety of economic activity types compared to household types, and diversity in firm size as opposed to the average household size. As will be introduced hereafter, this interaction has been addressed by many theoretical frameworks and modeled in different ways in subsequent generations of LUTI models.

2.3.1 Theoretical approaches

Traditional industrial theories of location decisions of firms have had a strong economic perspective and gave rise to well articulated predictions of spatial distributions of firms, derived from the principle of profit maximization in isotropic space (Launhardt, 1882; Weber, 1909; Fetter, 1924; Christaller, 1933; Lösch, 1940). Later, these economic theories were gradually replaced with psychological theories, emphasizing the relevance of concepts such as awareness and imperfect knowledge (Pred, 1967 and 1969; Townroe, 1969, 1972, 1975 and 1991; Cooper, 1975; Taylor, 1975; Wood, 1978). While the former were concerned with the best or optimal location, the latter looked at locations that were expected to be satisfactory. This led to the development of operational models and quantitative methods, which were supposed to represent the location decision making process.

While these behavioral models have provided significant thoughts about some aspects of the business activity, they assumed that location preferences were time-invariant or, more specifically, not related to the characteristics and the lifecycle of the firms. The attention shifted to the notion that location decisions should not be treated in isolation, but rather as an integrated part of a business cycle. In other words, the behavioral theory did not take into account the relevance of the firm’s structural factors (i.e., lifecycle, business activity) on location decisions. In response to this shortcoming, the institutional or organizational approach was developed, claiming a more integrated study on firm dynamics, emphasizing the activity-related behavior of firms (Walker and Storper, 1981; Storper and Walker, 1983; Healey and Ilbery, 1990). In the context of production, investments and accumulation of capital inputs, the problem of location suitability had a minor importance and it should be analyzed within economic, social, historic and political processes. Hence, location choices should be understood in the context of the larger organization, its lifecycle and the timing of other strategic decisions of the organization.

Given the above, the organizational ecology (Hannan and Freeman, 1989; Carroll and Hannan, 2000) also made its contributions, leading to the development of the firm demography approach. It attempts to identify and quantify the processes related to the firm dynamic behavior, drawing similarities with the events in population demography. Following van Dijk and Pellenbarg (1999a), classic firm demographic studies refer back to Cameron and Clark (1966); Stanback and Knight (1970); Allaman and Birch (1975); and Birch (1979) reporting regional economic development in the United States based on basic demographic processes of individual firms. For a firm demographic framework, such dynamics basically refer to the processes of starting-up, growing, declining, closing, locating and relocating.

In line with the above, the so-called evolutionary economic geography approach was developed. It explains the spatial distribution of economic activities based on the historic development of individual firms. That is, an evolving basic firm-related process of birth, growth, exit, and (re)location that results in spatial structures of the economy, emerging from the behavior of actors at the micro-level. In this approach, a notion of path dependency is emphasized, i.e., the current firm behavior is largely determined by past experiences. But at the same time, future patterns will be influenced by experiences acquired in the present moment. In this evolutionary framework, time and space are then inter-related (Krugman, 1998; Boschma and Lambooy, 1999; Boschma and Frenken, 2006a, 2006b and 2011; Boschma and Martin, 2007; Frenken and Boschma, 2007; Martin and Sunley, 2007; Boschma, 2009).

From a modeling practice perspective, these investigations on firm development have found strong linkages with the current stage of LUTI models development, based on multi-agent systems and micro simulation models. These are able to accommodate the notion of lifecycle, evolving firm-related processes, and the development of modeling frameworks based on individual activity-behavior principles. Next section will detail the empirical efforts to this end.

2.3.2 Firm modeling practices in LUTI frameworks

A first complete micro simulation model of firm demographic events was developed in the Netherlands by van Wissen (2000). The processes related to firm start-up, growth/decline, failure and relocation are modeled by means of statistical models. In addition to this, the author introduces the concept of carrying capacity into the related models, accounting for differences between market supply and demand in relation to such firm demographic events.

Although focusing only on the mobility process of business establishments, Khan et al. (2002a and 2002b) present a micro simulation framework extending the original firm location component of the MEPLAN land use model (Abraham and Hunt, 1999). The system comprises a spatial input-output framework that captures the business transactions among the different markets of the city, which is divided into discrete zones. These transactions across zones (or markets) are modeled by means of a nested logit model. If a zone comprises several sectors, the probability of choosing a certain market m will depend on selecting a destination zone, resulting in a two level nested logit model.

The UrbanSim model (Waddell and Ulfarsson, 2003; Waddell et al., 2007) is another example of a micro simulation framework, addressed as one of the first operational micro simulation of location decisions of households, firms and real estate markets. The firm component is still not a complete firm demographic framework, as it deals with only the location choice, modeled by means of discrete-choice models. A major difference in their approach, however, regards the fact that the system does not simulate firms, but employment instead. They argue that it captures the emergence of employment clusters and centers, addressing the concept of agglomeration economies. These, in turn, describe positive externalities associated with spatial proximity to firms within the same or related economic sectors.

The works developed for the city of Hamilton, in Canada (Maoh and Kanaroglou, 2005, 2007a, 2007b and 2009; Maoh et al., 2002 and 2005) comprise a complete agent-based firm demographic micro simulation framework, based on a set of integrated sub modules. These comprise a set of econometric models, implemented to evolve a population of business establishments over time. The initial step regards the closure sub module that determines the surviving firms between the current period and a time step (t and $t+1$). The relocation sub module verifies then the firms that will move within the city as well as out of it. Intra-migrant firms along with newly born firms are added to a list, comprising objects searching for a location. This is handled by the location choice sub module. In the end, a growth/decline sub module updates the changes in the firm size during the time period.

Another Dutch example of a micro simulation model of firm dynamics was developed by de Bok and Sanders (2005a and 2005b); de Bok and Bliemer (2006a and 2006b); and de Bok (2007). Following some of the ideas developed by van Wissen (2000), the authors implemented a firm demographic micro simulation, calibrating the framework for the South Holland region. The specific focus of their studies was on examining the influence of accessibility and agglomeration economies on the firm demographic development.

Moeckel (2005 and 2009) also presented a firm simulation framework integrated into the ILUMASS model, developed in Germany. The author uses Markov models to simulate firm birth, growth/decline, and closure, along with logit models to simulate firm relocation. The specific interest was on investigating alternative strategies for limiting the urban sprawl of jobs, i.e., suburbanization of employment showed in several simulated scenarios.

Most recently, Bodenmann (2011a and 2011b) and Schirmer et al. (2011) also presented some contributions in Switzerland with the investigation of firm (re)location choice and spatial accessibility within a micro simulation model. The focus was on initiatives to attract firms that could be taken by the authorities, such as: improvements in transport infrastructure, designation of new building zones, and tax reductions. In addition to this, the authors have also been making an effort to implement the UrbanSim model in Zurich, as part of the SustainCity project. This project aims at extending the UrbanSim model to the European context, as a modeling framework for the interactions between land use and transport.

In sum, this overview presented the efforts on investigating firm dynamics, since the development of traditional location theories to the most current approach, based on firm demography principles. It also claims that this approach suits well the advances obtained in LUTI models, addressing the use of agent-based micro simulation models to investigate firm dynamics in land use and transportation planning. The investigations found in the literature are mostly applied to firms in general, but from a more pragmatic perspective, the development of office firms has been the contemporary stream of representing urban economic activities. In other words, a city's wealth is usually related to the service sector nowadays. The study of office firms and their relationships with the urban environment is, hence, important and requires specific contributions. As the specific interest of this research refers also to the examination of these types of economic activities, the next section will present some background on this topic.

2.4 Focus on office firms

The crescent urbanization process observed in the last decades deeply changed not only the organization of space, but also the type of economic activities concentrated in the cities. The icon of the urban economic development shifted from the factory, manufacturing labor to the service sector (Daniels, 1975). The types of professionals in such activities involve, for example, architects, engineers, lawyers, technicians, government officials, managers, directors, accountants, real estate and insurance agents, telephone operators, health and educational practitioners, among others.

Although these professions are typically associated with office buildings, there may be some discussions on which jobs are really regarded as an office job (Alexander, 1979). For example, scientists, retail salesmen, and teachers might not be located in offices, but at universities, shops, and schools, respectively. More recently, with the advent of home offices (Hill et al., 2003), professionals may also not be performing activities in offices, but at home instead. Regardless these discussions, however, they all play a significant role in society, given the importance of the service sector in total employment and the economy at large.

Given the above and in line with the object of study in this research, office firms strongly influence land use patterns and related travel behavior. For governments and developers alike, it is very important to evaluate this interaction properly. As mentioned, these have been examined in LUTI models and studies on firm dynamics. However, apart from the works of Khan and Maoh (and their respective colleagues) as well as some modeling applications on office location and relocation decisions (Elgar and Miller, 2006; Elgar et al., 2008 and 2009), and examinations on the supply side of offices in terms of push, pull and keep factors (Appel-Meulenbroek, 2008; Elgar and Miller, 2007 and 2010), the applications found in the literature mostly consider firms in general. The specific interest in office firms within a LUTI modeling framework is relatively rare. But one can identify avenues of further investigations in this research topic. The next section will give an overview of the findings acquired from this literature review, in light of possible advances that the present research might contribute.

2.5 Conclusions

This chapter has outlined the development of modeling approaches to examine the interactions between land use and transportation, namely LUTI models. These have evolved from initial principles of aggregate spatial interaction and entropy-maximization, to agent-based micro simulation modeling that account for the activity-travel behavior of actors within the urban environment.

Along with the above, the investigations in firm dynamics also progressed from traditional location theories based on profit maximizing in isotropic space to more activity-behavior principles of individuals, emerging the firm demographic approach. This approach examines the dynamics of firms as events compared to population demography, comprising processes such as: birth, growth, decline, and closure. Location and relocation, although not very demographic, are analyzed together.

Applications linking firm demography with multi-agent systems can be found in the literature, commonly comprising part of larger urban simulation systems that consider all types of economic activities. However, specific investigations on office firms are still rare, although these are of special interest nowadays. Increasingly, the service sector has been the main stream of the urban economic development. Consequently, approaches delineating the influence of office firm dynamics and related travel behavior in terms of land use and transportation planning are needed. The present research tries to cover some gaps in the literature, exploring the current trends of LUTI models applied to office firm dynamics.

Firstly, although existing models do consider a relevant set of location attributes to represent the characteristics of land use and travel attributes, these could be extended in more detail. One of the contributions aimed at with this research is to elaborate the influence of location attributes on office firm dynamics, as a way to improve the level of detail and realism of such models. For example, regarding the specific influences of transportation infrastructure, the inclusion of airports, train stations, public transportation, etc. are considered, as opposed to the simple accessibility indicator based on the proximity to highways. Along with this, the use of more continuous measures to calculate accessibility, as opposed to some qualitative or discrete proximity measures, commonly used, is aimed at.

Secondly, there is a challenge regarding the scale of analysis taken in this research project. While most studies comprised a city or a region, here a nationwide scale of investigation is considered. In the Netherlands, van Wissen (2000) developed his study within such scale. However, there was a need for further elaborations on the consideration of the location attributes from a LUTI modeling perspective. De Bok and Sanders (2005a and 2005b); de Bok and Bliemer (2006a and 2006b); and de Bok (2007) had a view in this direction, but their application was carried out only for the South Holland region. Therefore, this research project is also an exercise on trying to complement those investigations for the Netherlands.

Finally, another contribution aimed at this research project regards the application of alternative statistical/econometric methods, some data mining techniques and use of geographic information systems (GIS) to bring together empirical investigations of agent-based micro simulation modeling and firm demography to the examination of office firm behavior. These methods will be specified in the respective chapters, but, for example, the location choice model is developed using a data mining technique based on Bayesian classifier networks, as opposed to the discrete choice models that are usually applied. Duration models are adopted and tested to analyze the firm closure and the firm relocation decision processes, accounting for the elapsed time before the event occurs. This is different from the binary logit models that are often applied in such cases.

The next chapter will present a proposed framework along with further details about the firm demography concept, developed to examine office firm dynamics as aimed at this research project.

3

A framework for modeling office firm dynamics

3.1 Introduction

As described in the previous chapter, the specific relationship between transportation and land use is examined in LUTI models based on disaggregate micro simulation and agent-based models. These are built on behaviorally richer concepts for examining firm dynamics, such as firm demography. As stated by van Wissen (2000), "demography of the firm is rooted in a view of the economy as being driven by the behavior of individual actors, which may be more appropriate in understanding various economic processes than the traditional macro level".

The objective of this chapter is to introduce a conceptual framework to simulate the evolution of office firms in time and space. The chapter first outlines the firm demography concept, introducing the several components along with the important notion of lifecycle. A modeling framework is then presented, with the underlying ideas of agent-based micro simulation models. The specific relationships with the urban environment are also delineated. Finally, some concluding remarks are drawn.

3.2 Firm demography

Since the works of Cameron and Clark (1966); Stanback and Knight (1970); Allaman and Birch (1975); and Birch (1979), there is an overwhelming interest in the firm demography concept for investigating the evolutionary cycle of firm dynamics. As mentioned by Pellenbarg and van Steen (2003), the first experience in The Netherlands with this method of analysis refers to Wever (1984), with studies on firm formation, relocation and closure using data from the Dutch Chambers of Commerce.

Lifecycle is a key notion in firm demography. It has been examined by many authors in industrial economics, business, and marketing, as for example, Dean (1950); Levitt (1965); Vernon (1966 and 1979); Cox (1967); and Mueller (1972). Basically, as

Figure 3.1 shows, lifecycle comprises four stages, usually associated with the evolution of the sales of a product over time: introduction, growth, maturity, and decline (see Klepper, 1996 and 1997 for details).

The first stage, introduction, refers to the period when the firm starts-up followed by the need to establish its market position, along with innovation and large investments. The second stage, growth, is characterized by the period when the firm reaches its maximum (monthly) revenue. The third stage, maturity, is defined when the firm starts to die commercially, usually given when the revenue declines 10 % to 20 % of the maximum monthly revenue. The final stage, decline, is associated with the period between the commercial death and the firm exit from the market.

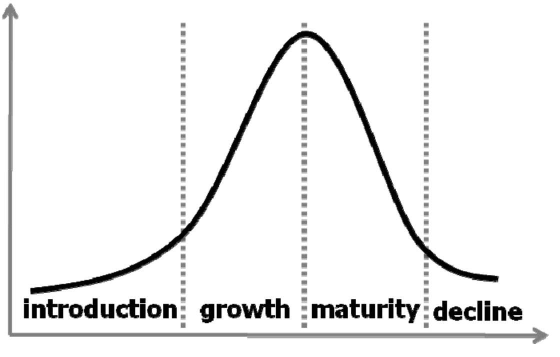


Figure 3.1. Stages of a lifecycle

In addition to the above, according to O'Rand and Krecker (1990) in social sciences, the notion of lifecycle was also addressed in organizational ecology (e.g., Hannan and Freeman, 1977 and 1987; Freeman, 1982; Carroll, 1984), where biological models were adapted to study the dynamics in organizations. Van Wissen (2002) extensively discusses this perception, arguing that there are significant comparable mechanisms between population dynamics and firm dynamics, although "the demographic metaphor does not arise because of applying biological laws to firms". Instead, these mechanisms rely on "methodological similarities in population dynamics and micro-macro linkages".

As outlined by van Dijk and Pellenbarg (1999a and 2000b), start-up and closure comprise the natural initial and end processes for the firm lifecycle. Like people get married and reproduce, firms may merge and generate spin-offs. Growth and decline stages are not very demographic, because of a lack of a direct relation in population demography, although one could think of a relationship with population ageing. Location and relocation are not directly related to the firm lifecycle, but consequences of its processes. For example, the establishment of a new firm requires a location, or a relocation decision might take place when firms grow. Nevertheless, the general processes of starting-up, finding a location to establish a business, growing or declining, relocating and going out of business are central in firm demography, as these are concerned with the changes in firm population in time and space. As reviewed by Maoh and Kanaroglou (2005), the firm demography concept is applied to examine the spatial dynamics in economy in various sectors, understood as the result of individual firm's behavior.

Firm demography is multidisciplinary, receiving attention from, for example, economics, sociology, and geography. Advances in firm demographic studies comprised influences from organizational ecology (Hannan and Freeman, 1989; Carroll and Hannan, 2000), industrial organization (van Dijk and Pellenbarg, 1999b) and evolutionary economic geography (see Frenken and Boschma, 2007; Boschma and Frenken, 2011, for a recent overview). In details, as reviewed by van Wissen (2000) and Maoh and Kanaroglou (2005), firm start-up has usually been of interest in regional economy (e.g., Reynolds et al., 1994; Geroski, 1995; van Wissen, 1997; Berglund and Brännäs, 2001), whereas firm closure has not only been examined by economists (Mata and Portugal, 1994; Audretsch and Mahmood, 1995; van Wissen, 1997; Baldwin et al., 2000; Berglund and Brännäs, 2001), but also by demographers (Ekamper, 1996) and sociologists (Brüderl and Schüssler, 1990; Baum and Singh, 1994; Hannan et al., 1998a and 1998b). Firm growth and decline has also been of interest in economy (Evans, 1987; Audrestsch, 1995; Hart and Oulton, 1996), while location and relocation decisions have been investigated in geography (Kemper and Pellenbarg, 1993, 1995 and 1997; van Dijk et al., 1999; van Dijk and Pellenbarg, 2000a; Brouwer et al., 2004).

In urban planning, firm demography has been treated in the LUTI frameworks, as introduced in Chapter 2. From this perspective, the expectation is that there is a strong relationship between firm demographic processes and the urban environment. The urban and transportation planning field tries to combine the various findings from the different areas, building a more integrated approach to investigate the evolution of firms in the urban environment. As also mentioned in Chapter 2, it regards the combination of the most recent developments on agent-based micro simulation models with the firm demography approach. The next section will introduce a modeling framework to this end, as a way to contribute to this specific field.

3.3 Description of the modeling framework

The modeling framework proposed in this research project builds upon the firm demography approach in a LUTI system, based on agent-based modeling and micro simulation. There are strong linkages between firm demography and multi-agent micro simulation systems that make the investigation of this relationship interesting in the current research agenda. For example, the notion of lifecycle can be accommodated by micro simulation, in which the different stages of lifecycle are simulated over time. From the multi-agent modeling practice, various processes can be defined to deal with the specific firm demographic events. These are then integrated in a larger system.

Along with the above, the underlying approach taken in this research project relies on the idea that the evolution of office firms is strongly influenced by the urban environment. More specifically, accessibility, land-use policies, infrastructure conditions, availability of buildings, economic and market prospects, employment locations, agglomeration economies and competition would influence office firm dynamics. However, it should be expected that some constraints apply when dealing with such dynamics in the urban environment. Therefore, following van Wissen (2000 and 2003), the notion of carrying capacity is also taken into account here, as a way to control the office firm demographic events.

Figure 3.2 presents the structure of the proposed framework to model office firm dynamics. It is composed by six components. First, the usual demographic cycle that firms usually go through is comprised by the events: start-up (or birth), growth and decline, and closure. In due course, it could also be expected that firms may merge or split-up, motivated by growth or decline processes. However, these are not particularly addressed in the current research project. Underlying this demographic level, a component related to the carrying capacity is introduced to control for the effects of firm dynamics within the urban environment. Van Wissen (2000 and 2003) considers the carrying capacity as an additional variable in the various sub models related to firm demographic processes. Here, the carrying capacity is estimated separately from the specific sub models, but comprises a controlling component within the overall system, as will be detailed later.

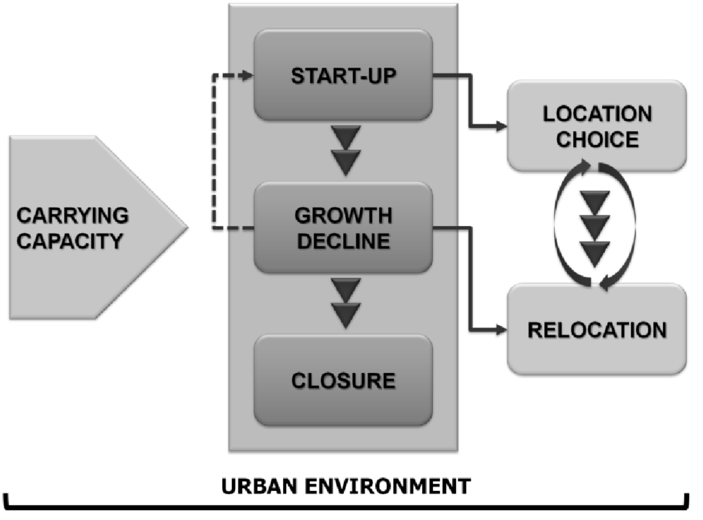


Figure 3.2. Scheme of firm dynamics and their interrelationships

Next, there are two other components related to the location choice and relocation decision. These are defined more as part of a firm's demographic event, usually associated with the spatial component in firm demography. Hence, they are designed as processes linked to the demographic events of starting-up and growth/decline, respectively. Moreover, the location choice and relocation are complementary events in the sense that a relocation decision leads back to a location choice problem. Finally, reinforcing the underlying idea of this research project, the entire process of firm dynamics is influenced by the urban environment.

Within a micro simulation framework, these components are part of a system that is dynamic in the sense that the collective behavior of office firms (as simulated) has an influence on the urban environment where firms operate. In turn, the urban environment has an influence on the firms' behavior to start-up, growth, relocate, etc. Hence, there is a dynamic loop. However, the purpose of this research project is not to develop such simulations. Instead, the focus is on the specification of the various models that can be used to simulate the behavior of office firms, which would be needed for eventually developing the overall framework. The specification of these models relies on the exploration of empirical approaches and modeling techniques for the individual components and their relationships with the urban environment. Essentially, such modeling techniques are built up on a set of statistical and econometric models as well as some data mining techniques. These will be specified later, at the related chapters depicting such individual components.

3.4 Conclusions

This chapter has outlined an approach for examining firm dynamics, based on the firm demography concept. It investigates the evolution of firms applying the notion of lifecycle and similar models found in population dynamics. The firm demography constitutes a valuable approach from an urban planning perspective, by using behaviorally rich concepts and methods.

In research practice, the approach has been successfully applied along with multi-agent micro simulation systems for examining the interactions between land use and transportation, namely in LUTI frameworks. Agent-based micro simulation models can easily accommodate the firm demography approach, by building specific components related to each firm demographic process that will eventually integrate a larger system. Moreover, the notion of lifecycle can be also captured and simulated over time, by means of micro simulation.

The modeling framework proposed for this research project specifies six main components. The first three regard to firm demographic events of start-up, growth/decline, and closure. These are dependent on a component defining the carrying capacity, which controls these processes based on the location characteristics. Two additional components are also defined, i.e., location choice and relocation decision, accounting for the spatial dimension in the proposed framework. Although the overall system reflects a simulation model, it is in fact not the focus of this research project. The introduction and specification of the various components along with the analysis of empirical applications is, instead, the object of study here, as a prerequisite for the development of such a simulation framework.

As mentioned, the focus of this research project is on the examination of the dynamics of office firms within a LUTI framework. Hence, the next chapter will present the data used to this end, along with the description of its source and related preparations. Also, the various datasets related to the urban environment used here will be presented.

4

Data

4.1 Introduction

This chapter presents the data used in the models and analyses developed in this study. The investigation comprises a nationwide scale for The Netherlands, for which data about office firms were obtained, along with several geographic databases regarding location attributes and socioeconomic characteristics. By means of Geographic Information Systems (GIS) and database management functions, the datasets were organized to be used according to the analyses involved, deriving several parameters of interest. The next sections detail the datasets and the procedures involved.

4.2 Office firm data: sources and attributes

The office firm data were obtained from two different sources: one provided by the company DTZ and the other related to the National Information System of Employment, the so-called LISA data. Their details are described in the following subsections.

4.2.1 DTZ data

DTZ Zadelhoff is a real estate company responsible for business properties, not only in The Netherlands, but also in other parts of the world. The data obtained for this study concerns the offices that DTZ manages in The Netherlands. It contains a sample of 3,303 objects at the address level (street name and number), for the year 2007, regarding offices that were occupied by a firm, i.e., non-vacant offices. Among the attributes included, the following are of interest: 6-digit postcode; address given by street name and house number; economic sector given by either a label or by a code according to the SBI (Standard Industrial Classification); office space area in m²; and rent price per area (Euro/m²).

Regarding the economic sector classifications, the records were not well organized in terms of the official references. Although the SBI code concerns one of the included attributes, most records did not have this code. Instead, they were mainly classified by labels defined by the company, which not always matched the official ones. To better prepare the data for analyses, a reclassification was carried out based on the provided labels, adjusting these as closely as possible to the official classification. It resulted in 17 office firm categories, as follows: Agriculture; Mineral extraction; Industry; Production and distribution of energy, gas and water; Building industry; Real estate and retail services; Hotel and catering industry; Traffic and communication infrastructures; Financial institutions; Public services and social security; Commercial and professional services; Repair of consuming products; Computer services and information technology; State organizations; Health and welfare; Education; and Environmental, culture and recreation services.

The DTZ data were geocoded allowing analyses in GIS. To this end, a database of postcodes at the 6-digit level was used, whose geographical coordinates were transferred to the DTZ data. Figure 4.1 presents the spatial distribution of the office firm database obtained from the company DTZ.

Figure 4.1. Database of office firms obtained from DTZ

4.2.2 LISA establishment data

The data used on firm establishments refer to the records of the National Information System of Employment, the so-called LISA data, and were analyzed at Utrecht University in cooperation with dr. Ettema and prof. van Oort. It concerns a longitudinal dataset, comprising all types of firms, i.e., not only office firms, recorded from 1996 to 2006. The attributes include the LISA unique identifier code; the size of the firm given by the number of jobs; the economic sector given by the 5-digit SBI code; and the postcode at the 6-digit level.

As mentioned above, the LISA data comprise all firms. Given that the interest in this study regards office firms, a selection of observations referred to only office firms was carried out, based on that 5-digit SBI code. Every classification that could be referred to an office firm was included. Table 4.1 presents an overview of the number of firms present in each year and the resulting number of office firms selected, along with the totals (sum of the number of firms existing between 1996 and 2006, according to their LISA codes).

Table 4.1. Number of firms and office firms in the LISA datasets

Year	Firms	Office firms
1996	659,920	137,090
1997	683,381	145,725
1998	715,174	159,928
1999	734,461	169,507
2000	752,707	178,095
2001	770,745	188,149
2002	787,247	195,489
2003	798,556	199,579
2004	808,060	203,288
2005	832,657	210,287
2006	859,683	218,697
Total	**1,618,846**	**425,241**

Differently from the DTZ data, the LISA datasets were better organized in terms of the economic sector classifications. They were obtained at the most detailed level, i.e., 5-digit SBI code, following the official references of the Dutch Bureau of Statistics. However, for the analyses conducted in this study, a more aggregate level was used, i.e., the 2-digit SBI code level. This involves 15 categories of office firm types, which are relatively similar to the ones obtained for the DTZ data (numbered firm types between parenthesis to ease later discussions): Agriculture (type 1); Industry (type 2); Basic infrastructures (energy, gas and water – type 3); Building industry (type 4); Retail and horeca (hotels, restaurants and cafes – type 5); Traffic and communication (type 6); Financial institutions (type 7); Social security (type 8); Real estate (type 9); Business service (type 10); Computer and information technology (type 11); Research and development (type 12); Public administration (type 13); Education and health (type 14); Environmental services, culture and recreation (type 15).

Except for the location choice model, the LISA datasets were used for estimating the various models developed in this study. Similar to the DTZ data, the LISA data was geocoded using the postcode database. Figure 4.2 presents its spatial distribution.

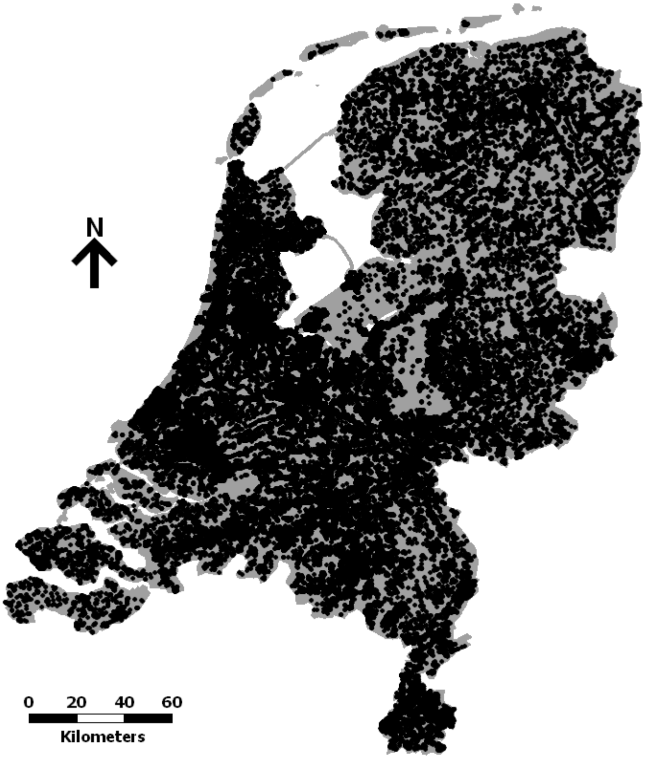


Figure 4.2. Database of office firms obtained from the LISA

4.3 Socioeconomic and spatial data: sources and attributes

The socioeconomic and spatial data were obtained from several sources and at different scales. In the subsequent sections, more details are discussed.

4.3.1 Transportation infrastructure

The data about the transportation infrastructure mainly refer to the location of airports, train stations, public transportation stops (bus, tram, and metro stops), and highways. As for the airports, the international airport (Schiphol), regional airports and the closest major airports in Germany and Belgium were considered. Their locations, i.e., the geographic coordinates were obtained from the internet, which resulted in a database of airports created manually. This included Schiphol, Eindhoven, Groningen-Eelde, Maastricht-Aachen, Rotterdam, Düsseldorf, and Brussels.

Regarding train stations, a differentiation between international train stations, intercity (IC) train stations, and local train stations was made. As for the international train stations, these include Amsterdam Central, Schiphol, Den Haag HS, Rotterdam Central, Utrecht Central, and Arnhem. They were selected based on the fact that high speed train services connect them, such as Thalys, ICE International, Eurostar, TGV, among others. However, these do not fully operate at high speed services within The Netherlands. That is the reason for labeling them as "international train stations" and, along with that, not including other international train stations such as Venlo, Heerlen, Enschede, Roosendaal, among others in this category. Specifically, these are included as IC train stations. Data about the location of such international train stations were obtained from the internet, which were used to manually create a database as well. The intercity and local stations were obtained from the Dutch National Railways, comprising a geographic database from 2003. Only the intercity stations were used in the analyses.

The location of bus, tram, and metro stops was also obtained from the internet, specifically from the OpenStreetMap service. The decision to use this data source was made at a later stage of the study, given that, initially, obtaining an official dataset had been tried. Analyses that were then carried out before this decision, i.e., the closure and location choice model, do not include the location of public transportation. The data used refers to January 2010.

Finally, as for the highways, a 2003 geographic database was obtained from the Dutch Ministry of Transportation. It was basically used to compute distances over the roadway network between several points of interest (e.g., distance from offices to train stations, airports, etc.), and to extract the location of highway junctions. The maps in Figure 4.3 present an overview of the transportation infrastructure datasets.

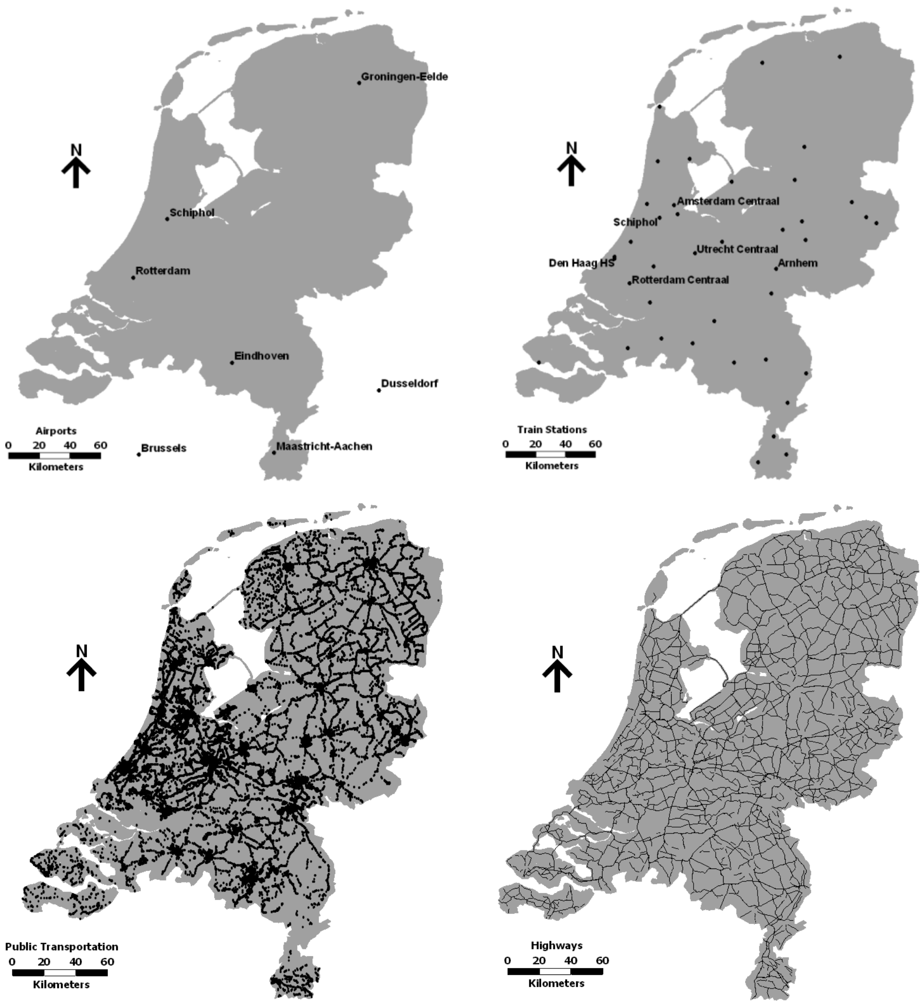


Figure 4.3. Maps of transportation infrastructure datasets

The fact that the (constructed) databases stem from different years and do not include any dynamics is, of course, not ideal. In principle, perhaps exact details for the various years could be reconstructed. However, in the present case, it was decided to use these data as the amount of time and effort involved would not outweigh the main aim of this PhD study, concerned with the development of the framework for an agent-based simulation of office firm dynamics. Possible small differences in the databases are not expected to dramatically change the major findings of this study.

4.3.2 Demographic attributes

The demographic attributes used in the models refer to urbanization levels, population, number of households, labor force, and average income. They were obtained from different sources and at different scales. A first source comprised the 2004 data underlying the NRM (New Regional Model), available at the 4-digit postcode level and obtained from the Ministry of Transportation.

Similar demographic data along with geographic databases were then obtained from the Dutch Bureau of Statistics, from 2005, at the levels of municipality, neighborhoods and districts. As expected, differences in the statistics were observed when comparing data of the two sources, which might be partly caused by the different time periods. The most recent values were hence used, although some information from the NRM data was still used in the development of the various models, as will be reported later. Additional information concerning a 1996-2006 database of population was also obtained from the Dutch Bureau of Statistics.

In general, the abovementioned demographic attributes were defined as follows: population, in number of inhabitants; labor force, in number of employed persons, and average income, given in Euros. The urbanization levels are defined as:

- Level 1: very highly urbanized area (more than 2,500 addresses/km²);
- Level 2: highly urbanized area (between 1,500 and 2,500 addresses/km²);
- Level 3: moderately urbanized area (between 1,000 and 1,500 addresses/km²);
- Level 4: lowly urbanized area (between 500 and 1,000 addresses/km²);
- Level 5: non-urbanized area (less than 500 addresses/km²)

4.3.3 Facilities

Some location attributes related to general facilities, such as: shopping centers, schools, and parking spaces were also considered in the models developed. The location of shopping centers was obtained from the Dutch Council of Shopping Centers (NRW). This database concerns the 2006-2007 time period and comprises 984 objects. Using the 6-digit postcode included, this database was geocoded. The number of educational places and parking spaces were obtained from the NRM data at the 4-digit postcode and geocoded accordingly.

4.3.4 Regions

In addition to the geographic databases from the Dutch Bureau of Statistics, some regional databases were also obtained. A first one concerns a 40-region dataset related to the so-called COROP (literally, Coordination Commission Regional Research Program) areas. These refer to a statistical classification that divides The Netherlands into 40 regions. They are delineated by a core area (usually a city) surrounded by a catchment area, depicting for example, considerable residential relationships. This classification matches with the so-called NUTS-III, often used within the European Union. COROP areas were meant to describe the functional regions in the Netherlands.

Next, at a more aggregated level, a database referring to the 12 Dutch provinces (NUTS-II) was obtained, also from the Dutch Bureau of Statistics. It entails the administrative layer between the national government and the local municipalities, having the responsibility for sub national matters or regional importance. Finally, a subdivision of The Netherlands into three major regions was used to ease some analyses. It comprises a Northern and Southern part, along with the Randstad area, labeled here as NSR regions. Figure 4.4 presents maps of these regions, including the municipality and 4-digit postcode databases mentioned before.

These various regions were used at different stages of this study. For example, the 4-digit postcode areas were used as the units of observations for several attributes related to demography and facilities. The COROP areas were also used as units of observation in some models, given their inherent statistical characteristic. In addition to the start-up and carrying capacity models, they were used to compute some agglomeration economies measures, as will be detailed in the next section.

The Dutch provinces were basically used to capture some regional effects in the developed models. The use of COROP areas should perhaps be preferred in the sense that these areas are more homogeneous and represent functional areas. However, these potential advantages come at the cost of dramatically increasing the number of parameters in the models, especially if interaction effects are taken into account. The use of provinces rather than COROP areas reduces the number of additional variables from 39 to 11. Nevertheless, some models still had a limitation in terms of the inclusion of possible independent variables, considering the number of available observations. Hence, for some analyses, the NSR regions were used, reducing the number of additional variables from 11 to only two. Basically, it differentiates effects obtained in the models across the Northern, Southern and Randstad areas.

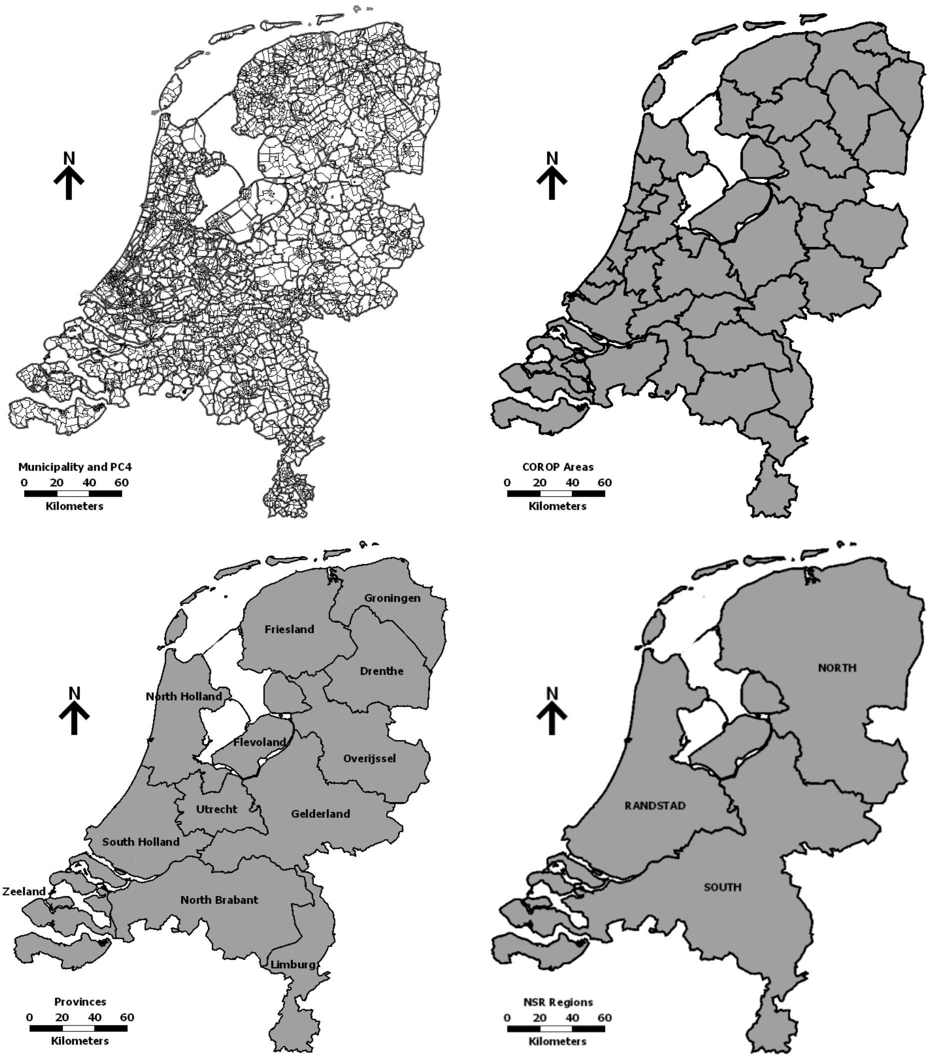


Figure 4.4. Maps of regions in The Netherlands at different scales

4.4 Derived attributes

Based on the numerous described datasets, several attributes were derived in order to be used in the developed models. They regard both office firm characteristics and socioeconomic and spatial attributes. It is important to mention that further details concerning these attributes can be found in the related model specifications. Also, not all derived attributes were used in all models. This was due to: either the definition of an attribute happened at a later stage, when the model had already been developed, or because the specific attribute was not considered relevant for the related model. An overview of such derived attributes is presented hereafter.

4.4.1 Office firm characteristics

These comprised the following attributes: economic sector (office firm type); frequency of office firm start-up; density of existing office firms; density of existing firms (not related to office firms); office firm size (number of jobs); office firm growth; duration (in years) in the current address before a relocation; duration (in years) of existence before going out of business; and distance (km) of relocation patterns.

4.4.2 Accessibility to transportation infrastructures

These comprised the following attributes: density of public transportation facilities, given by the number of bus, tram, and metro stops within a circle area of 500 m of radius around the office firm location; distance to the closest airport; distance to the international airport (Schiphol); distance to the closest intercity (IC) train station; distance to the closest international train station; and distance to the closest highway junction. It is important to observe that the distances were computed over the roadway network.

4.4.3 Demographic and economic aspects

These comprised the following attributes: urbanization levels at both municipality and 4-digit postcode levels; population (number of inhabitants) at both municipality and 4-digit postcode levels; number of households at the 4-digit postcode level; labor force at the 4-digit postcode level; and average income (Euro) at the 4-digit postcode level.

4.4.4 General facilities

These comprised the following attributes: distance (km) to the closest shopping center, computed over the roadway network; density of shopping centers, given by the number of shopping centers within a circle area of 1 km of radius around the office firm location; number of places at schools at the 4-digit postcode level; and number of parking places at the 4-digit postcode level.

4.4.5 Regional effects

These comprised the effects of the Dutch provinces (Drenthe, Flevoland, Friesland, Gelderland, Groningen, Limburg, North Holland, South Holland, Overijssel, Utrecht, Zeeland, North Brabant), as well as the effects of the NSR regions (North, South, and Randstad area).

4.4.6 Agglomeration economies

The effects of agglomeration economies were derived from the number of firms related to the same type, ideally computed within a COROP area. However, for some models, this scale did not turn out to be significant. Hence, the same measure was additionally calculated within a circle area of 1 km of radius around the office firm location.

4.4.7 Rent price estimations

As mentioned previously, one of the attributes included in the DTZ data was the rent price per m². However, about 35 % of the records comprised missing values. Attempts to solve this problem were carried out, using several modeling approaches. These included linear regression, spatial regression, neural network, and digital elevation models. It is not the focus of this research project to detail them. Instead, only the modeling approach that resulted more significant on estimating rent price values across the country is reported here, which referred to the digital elevation model (DEM).

Based on a triangulated irregular network (TIN) method, which is usually available in GIS software, a continuous surface representing rent price values across the country was generated from the existing figures in the DTZ data. The goodness-of-fit of this model, expressed in terms of a correlation coefficient between the observed and the estimated values, was equal to 0.89. Despite this fairly good result, one might question whether rent prices are in fact continuous across space. Although it may not be the case, more importantly, this approach could fulfill the needs of obtaining estimated figures to complete the missing values observed in the database.

On the top of the above, the estimated model could estimate not only the missing values for the DTZ data, but also allowed an extrapolation for the LISA data. This resulted, hence, in an additional attribute to be used in the various models.

4.5 Conclusions

The findings discussed in this chapter allow the following conclusions. The data obtained for this research project are generally very rich in both quality and quantity. Specifically, the office firm data comprise information at the individual level, which is very important for the nature of research in this project. Along with that, the location attributes were obtained from various datasets that also provided very detailed information about the urban environment. This is essential in LUTI frameworks, as the case investigated here.

The nationwide scale of study of this research project may bring several challenges, from analyzing the dynamics of office firms across different parts of the country, to dealing with a large amount of data. However, it is expected that the objectives of this research project are fulfilled, improving the understanding on modeling office firm dynamics in a LUTI system through empirical investigations. The next chapter will introduce the first model developed to this end, regarding the process of office firm start-up.

5

Start-up model

5.1 Introduction

This chapter presents the office firm start-up model. Van Wissen (2000), for example, reviews that firms start due to numerous reasons, but a first, usually crucial reason is related to the decision whether to start a new business or not. Geroski (1995) provides an overview on this topic, describing several stylized facts that contribute to firm start-up and related influences in the market. A first fact is that firm start-up is common, natural, and comprises relatively high rates of firms entering most markets in most years. However, market penetration is an issue seen as the successful establishment of a business, usually happening at a much lower rate. The author also reports that small-scale start-ups are relatively easier than large-scale start-ups.

In line with the above, another reason for firm start-up refers to the decision of individuals on being self-employed and start a business (Reynolds, 1997). According to the theory of entrepreneurial choice, the start-up propensity of a firm will be higher if the expected income related to the establishment of a new business is higher than the income from being employed (Berglund and Brännäs, 2001). Along with this, new firms also emerge from spin-off processes, when new entrepreneurs seem to benefit from the experiences acquired from the parent employment and decide to start a new business (Wenting et al., 2011).

Economic prospects and access to capital is fundamental to start a new business. Regions with good investment incentives respond positively to firm start-up, especially firms related to research and development. Employment structure in regions where there is specialized work force also contributes to explain firm start-up (Hart and Scott, 1994; Garofoli, 1994; van Dijk and Pellenbarg, 2000a; Berglund and Brännäs, 2001; Frenkel, 2001). Another important influencing factor reported in the literature refers to the effects of agglomeration economies. Some authors (Huisman and van Wissen, 2004; van Oort and Atzema, 2004) showed evidence of increased firm start-up in locations with spatially dense economic activity. Knoben et al. (2011) also report these influences along with investigations on the role of knowledge spillover. In turn, policies for regional innovation (Ponds et al., 2010) may be associated with firm start-up, especially in regions with universities and research institutes.

In addition to the developments in the fields of economic geography and regional science, the topic has also been of interest in urban planning, more specifically, in the role of transportation infrastructure on stimulating economic activities. Holl (2004) has investigated the role of investments in highways and its influences on the spatial patterns of firm start-up. The findings suggest some attractiveness of these improvements on the firm start-up behavior. Melo et al. (2008) have also examined the relationships between firm start-up and proximity to highway and railway networks. Although the results indicated that transportation networks do not seem to contribute to firm start-up, which is contrary to their initial hypothesis, the authors mention some constrains in the data used. This might explain the reported findings. However, they indicate that more detailed data could show results as initially expected, yielding to further investigations.

In line with the above, applications in firm demography (van Wissen, 2000; Khan et al., 2002a and 2002b; Maoh and Kanaroglou, 2005; de Bok, 2007; Moeckel, 2009) have investigated the relationships between firm dynamics and the urban environment within LUTI models, although there is a need for more detailed empirical approaches with regards to data inputs. Therefore, the objective of this chapter is to contribute to this field, by developing a model to estimate firm start-up within a LUTI framework for office firm dynamics. As part of the larger multi-agent micro simulation system, this firm start-up component derives parameters that translate the effects of location attributes on office firm start-ups, which can be eventually used in a simulation. The chapter is organized as follows: a modeling approach is described, followed by an empirical test. The results and analyses are then presented and some concluding remarks are drawn.

5.2 Modeling approach

The model presented in this chapter regards a component of a multi-agent system specified to examine the process of firm start-up in relation to a set of characteristics. As part of a LUTI framework, the design of the firm start-up model is based on the idea that firms would start-up given certain urban-related conditions. This does not comprise only transportation infrastructures, but also includes population aspects and agglomeration economies. The latter especially assumes that existing firms within a defined area would influence this process. These regard not only firms from the office sector, but also from other industries.

The approach taken here models the frequency of start-ups within a defined area, examining how it responds to the urban conditions. The reason for using here an aggregate approach as opposed to an agent-based approach is that a firm must exist before it could be simulated. Hence, the fact of coming into existence can be only modeled by taking an area as a "decision unit". So, this particular model is not intended to simulate the start-up behavior of firms, but rather is part of the simulation of an environment where firms come into existence every year, given certain urban conditions.

Given the above, the response variable refers to count data, i.e., the frequency of firm start-ups within an area. In such statistical-based modeling framework that this research project is developed, the choice of a specific statistical model depends on the nature of the estimated outcome. As dealing with count data, the application of standard least squares regression is not correct, because it yields inconsistent predicted values that are non-integers and possibly negative. Therefore, there are some statistical models that can be used to properly examine such type of censored data, including the Poisson regression, the negative binomial regression, and the Tobit regression. For the analyses developed here, the Poisson regression model was adopted. It is given by its general form, as presented by Equation 5.1.

$$\ln y_i = \beta_0 + \sum_{n=1}^{n} (\beta_n X_{ni}) \tag{5.1}$$

where y_i is the dependent variable, in this case the frequency of office firm start-ups in area i; X_{ni} is a vector of inputs, in this case the n independent variables calculated for area i; and β_n is a vector of coefficients to be estimated.

In the analysis depicted in this chapter, the interpretation of the findings is based on the effect that a given variable has on expressing the frequency of office firm start-up. For positive coefficients, the increase (decrease) in the value of such variable leads to an increase (decrease) in the number of start-ups, whereas for negative coefficients, the increase (decrease) in the value of such variable means a decrease (increase) in the number of start-ups. Having outlined the theoretical notions along with the modeling approach for examining the office firm start-up process, the next section will detail the empirical application.

5.3 Empirical modeling setup

The LISA register data for the 1997 to 2006 time period were selected, as observations had a known starting time. COROP areas were used as units of observation. Both firm-related and socioeconomic and spatial attributes were aggregated to the level of these COROP areas. The dependent variable is the number of starting office firms of each economic sector in any given year. This results in 15 separate models. As for the set of independent variables, calculated at the level of COROP areas, these included:

1. NOFE: number of existing office firms related to the same type;
2. NOFN: number of existing office firms of other modeling types;
3. NNOF: number of existing non-office firms;
4. POP: population;
5. NJE: number of jobs of existing office firms of the same type (i.e., economic sector size);
6. Regional effects related to the NSR regions;
7. Density of public transportation facilities;
8. Average distance (km) to Schiphol international airport;
9. Average distance (km) to IC train stations;
10. Average distance (km) to highway junctions;
11. Average distance (km) to shopping centers;
12. Average rent price / m².

These independent variables were defined to work as proxy measures related to effects of agglomeration economies, influences of populated areas on start-up, regional differences across the country, and influences of transportation infrastructures, facilities and pricing. It should be observed that the aggregated averages calculated at the level of COROP areas were based on the values attributed to each individual firm.

Given the available data, a set of 15 models could be estimated for each of the 10 years available. However, estimating a model for each year separately leads to unstable estimates because of a relatively low number of observations. Therefore an alternative pooling data procedure across the years was used. This means that all separate datasets for each year are merged into one dataset, resulting in 400 observations for each firm type (related to 10 years versus 40 COROP areas). The observations across firm types were not mixed. By performing such pooling procedure, tests also showed that a better model fit is attained when the relative value between two consecutive years is taken for (some of) the independent variables. Therefore, instead of using the absolute value of a given attribute, the difference of its actual value compared to the previous year was used. For example, consider that a certain number of office firms starts-up in 2000, and it can be explained by the number of existing non-office firms. Instead of taking the actual number of existing non-office firms in 2000, the difference between 2000 and 1999 is used. This reasoning can be explained by the fact that the dependent variable (number of start-ups) already reflects a change from one year to the next year. By using then the computed change of the values of independent variables across the same period, one can expect that this relative measure is better able to explain the number of start-ups. It should not go without saying that this was only performed for the attributes that had data available across the 10-year period of analysis. For the attributes whose information was present at only one point in time, that is, accessibility to transportation facilities, distance to shopping centers and rent price, the absolute value was taken instead.

Another consideration taken into account regards the lagged effects of some independent variables on firm start-up. The reasoning behind it refers to the fact that the influence of a given attribute on start-up may not result from the immediate variation of this attribute in the same period. Instead, a change in, e.g., the number of existing non-office firms in a past period could be responsible for observing start-ups in the actual period. The question of how many years backward should be considered regards a decision for the analyst in exploring such influences. However, the available period of observation in the data is also a constraint, as the more lagged periods are considered, the less observations will be available in the sample. In the case explored, a lagged effect over 2 years was computed, which was adopted to not decrease excessively the size of the sample, but also to give an insight of this influence in the problem addressed here. Also, these lagged effect variables were computed only for the variables where data over several periods were available, i.e., the number of existing office firms related to the same type; the number of existing office firms of other modeling types; the number of existing non-office firms; population; and the number of jobs of existing office firms of the same type. Again, accessibility to transportation facilities, distance to shopping centers and rent price were not included in the lagged effects.

In sum, three components regarding these lagged effects can be defined as follows: the current relative value (i.e., the difference across two years explained previously), the lagged relative value in 1 year, and the lagged relative value in 2 years. These are, hence, additional variables in the model. For example, consider that a random observation in the sample comprises the number of firm start-ups that is identified for the year 2002. As for a given independent variable, the current relative value refers to the period between 2002 and 2001. Next, the 1-year lag term is computed as the difference between 2001 and 2000, and the 2-years lag term is computed as the difference between 2000 and 1999. The scheme in Table 5.1 exemplifies this procedure.

Table 5.1. Scheme showing the related periods in which lagged effects were calculated based on the reference year

Number of start-ups in:	Current relative value	1-year lag	2-years lag
2006	2006-2005	2005-2004	2004-2003
2005	2005-2004	2004-2003	2003-2002
2004	2004-2003	2003-2002	2002-2001
...	...	...	...
2000	2000-1999	1999-1998	1998-1997

The resulting sample comprised 278 observations for each firm type, as two blocks of 40 observations were missed due to the lagged effects procedure (here the exclusion of the outliers are also already regarded). Although multiple observations per COROP area are involved and unobserved characteristics of COROP areas may play a role, a panel model structure was not implemented. Note however that time-dependency that is observed in the data is assumed to be captured by the calculations of the differences across two periods, as described. The related Poisson regression models were estimated in the software NLOGIT version 4.0 (Greene, 2007) and the findings are presented in the next section.

5.4 Analyses and results

The results of the estimated Poisson regression model are presented hereafter. These were carried out by firm type, comprising 15 separate models. The parameters are presented in Table 5.2 to Table 5.8, which refer to each independent variable used, where "z" represents the statistical z-test used for estimating the significance of the parameters. Especially regarding the lagged components, a note about their interpretation and logical consistency is in order. Logical consistency requires that the three components should have the same sign, indicating that the influences are consistent across the years. If the 2-years lag would differ from both current and 1-year lag, it should be checked whether the current component and the 1-year lag are consistent. On the other hand, it may also be that the current value and the 2-years lag are consistent, but not the 1-year lag. In this case, it is assumed that only the current component influences start-up and there are no lagged effects. Of course this requires further investigations, including more past lagged components, but those were the assumptions taken in the case studied here. Another possibility is when only the lagged components are consistent. In this case, it is considered that only these influence firm start-up and the current component has no impact.

In Table 5.2 the parameters for the number of existing office firms related to the same type (NOFE) are presented. Firm types 4, 11, 12, and 14 (respectively, building industry; computer and information technology, research and development; and education and health) have positive signs for the three components. This means that office firms of such types are motivated to start when the number of the related existing office firm type increases, or has increased in the past. Similarly, but holding negative coefficients, firm types 5 and 15 (respectively, retail and horeca, and environmental services, culture and recreation) are discouraged to start-up when the number of the related existing office firm type increases (or has increased in the past).

Office firms related to traffic and communication, financial institutions, real estate, and public administration (types 6, 7, 9, and 13, respectively) hold parameters with the same sign and magnitude only for the current and the 1-year lag components. Specifically, real estate companies hold positive coefficients, meaning that these are motivated to start when the number of existing real estate companies increases, or has increased in the past year. The other firm types have negative coefficients and, hence, they are discouraged to start-up when the number of the related existing office firm type increases (or has increased in the past year).

For firm types 3 and 8 (basic infrastructures and social security), only the current component is consistent, given that logical consistency across the lagged components cannot be observed. As both have negative signs, they are discouraged to start-up when the number of the related existing office firm type increases. Conversely, only the lagged components influence firm start-ups related to types 1, 2, and 10. Agriculture (type 1) has positive signs whereas industry and business services (types 2 and 10) have negative signs. That is, if the number of office firms related to agriculture has increased in the past, then this stimulates more start-ups of office firms related to this type. On the other hand, if the number of office firms related to industries and business services has increased in the past, then this discourages start-ups of office firms related to these types.

Table 5.2. Parameters for the variable NOFE

Firm type	Current		1-year lag		2-years lag	
	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]
1	-4.39E-02	0.869	4.00E-01	0.030	1.52E-01	0.437
2	7.29E-04	0.846	-1.61E-03	0.630	-4.73E-05	0.986
3	-9.03E-02	0.181	1.13E-01	0.109	-6.33E-02	0.328
4	3.90E-02	0.064	1.17E-01	0.000	1.57E-01	0.000
5	-2.61E-03	0.000	-7.45E-03	0.000	-2.52E-03	0.001
6	-2.55E-02	0.365	-1.57E-03	0.953	6.65E-02	0.043
7	-2.02E-03	0.000	-6.23E-03	0.000	9.86E-04	0.001
8	-8.89E-02	0.000	3.17E-03	0.880	-1.43E-02	0.442
9	4.86E-03	0.000	2.91E-03	0.000	-4.42E-03	0.000
10	4.08E-04	0.000	-3.66E-04	0.000	-3.46E-04	0.000
11	2.42E-03	0.000	1.67E-03	0.000	3.07E-03	0.000
12	1.12E-02	0.000	5.76E-03	0.054	6.40E-03	0.044
13	-1.53E-02	0.000	-2.56E-02	0.000	-2.79E-03	0.592
14	8.86E-02	0.000	5.06E-03	0.782	1.34E-02	0.483
15	-1.79E-03	0.099	-1.05E-03	0.253	-3.34E-04	0.719

Turning to the analysis of the findings for the number of existing office firms of other types (NOFN), presented in Table 5.3, it can be observed that firm types 4, 6, 8, 10, 11, 12, 13, 14, and 15 (respectively, building industry; traffic and communication; social security; business service; computer and information technology; research and development; public administration; education and health; and environmental services, culture and recreation) have consistent results for the three components. Moreover, except for traffic and communication (type 6), they all have negative coefficients. Therefore, if the number of other firm types increases or has increased in the past, then this discourages start-ups of office firms (related to the modeling type). On the other hand, for traffic and communication, if the number of other firm types increased or had increased in the past, this stimulates the start-up of this office firm type.

Holding consistent parameters for the current and for the 1-year lag components are office firm types 2, 5, and 9 (industry, retail and horeca, and real estate), with negative coefficients. Hence, if the number of other firm types (different from the modeling one) increases or has increased in the past year, then this discourages start-ups of office firms related to these modeling types. The same pattern can be observed for firm types 1 and 7 (agriculture and financial institutions), but only the current component is considered to influence it, as the lagged components were not consistent across the period. Conversely, companies related to basic infrastructures (type 3) are influenced only by changes observed in the past years, as the lagged effects seem to be consistent. Holding positive coefficients, these office firm types are motivated to start-up if the number of office firms related to all other types has increased in the past.

Table 5.3. Parameters for the variable NOFN

Firm type	Current		1-year lag		2-years lag	
	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]
1	-3.51E-04	0.538	3.43E-04	0.452	-2.02E-04	0.694
2	-5.77E-04	0.000	-2.76E-04	0.003	9.61E-05	0.263
3	-3.91E-04	0.077	1.63E-04	0.590	3.45E-04	0.195
4	-4.01E-04	0.047	-7.89E-04	0.001	-5.61E-04	0.020
5	-3.97E-04	0.000	-1.35E-04	0.001	9.82E-05	0.013
6	9.63E-04	0.000	5.60E-04	0.010	1.47E-04	0.440
7	-5.82E-04	0.000	2.14E-04	0.000	-2.18E-04	0.000
8	-5.88E-04	0.000	-3.85E-04	0.023	-1.54E-04	0.409
9	-7.22E-04	0.000	-6.42E-04	0.000	6.55E-04	0.000
10	-1.27E-03	0.000	-2.73E-04	0.000	-6.66E-05	0.058
11	-8.55E-04	0.000	-6.72E-04	0.000	-4.12E-04	0.000
12	-7.67E-04	0.000	-5.13E-04	0.000	-4.82E-05	0.502
13	-9.29E-04	0.000	-8.91E-04	0.000	-1.01E-04	0.464
14	-5.07E-04	0.004	-5.70E-04	0.001	-1.50E-04	0.321
15	-1.18E-03	0.000	-8.03E-04	0.000	-2.40E-05	0.704

Next, as shows Table 5.4 regarding the number of existing non-office firms (NNOF), it can be observed that most office firm types hold positive coefficients and are consistent across the lagged effects as well. Namely, types 2, 5, 7, 8, 9, 10, 11, 12, 13, 14, and 15. It means that increasing the number of non-office firms in the current and past years results in more office firms (of such types) starting up. Office firm types 3 (basic infrastructures) and 4 (building industry) follow the same pattern, but with one difference: type 3 is influenced by the current and 1-year lag components, whereas type 4 is influenced only by the current component. Office firms related to agriculture (type 1), and traffic and communication (type 6) are also influenced by the current and 1-year lag components only, but these hold negative coefficients. Therefore, for these office firm types, the effect of an increase in the number of non-office firms observed in the current and past year results in less office firms of such types starting up.

Table 5.4. Parameters for the variable NNOF

Firm type	Current		1-year lag		2-years lag	
	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]
1	-1.09E-04	0.854	-9.23E-04	0.101	1.05E-03	0.037
2	4.80E-04	0.000	1.22E-04	0.038	3.87E-04	0.000
3	2.04E-04	0.441	4.62E-04	0.103	-1.16E-04	0.635
4	8.42E-04	0.001	-2.43E-04	0.276	6.46E-04	0.003
5	5.21E-04	0.000	3.45E-04	0.000	3.61E-04	0.000
6	-1.31E-04	0.642	-1.41E-03	0.000	3.11E-04	0.206
7	5.31E-04	0.000	3.42E-04	0.000	4.42E-04	0.000
8	3.47E-04	0.030	5.00E-05	0.770	7.92E-04	0.000
9	4.16E-04	0.000	1.87E-04	0.000	2.69E-04	0.000
10	4.50E-04	0.000	2.30E-04	0.000	3.41E-04	0.000
11	5.78E-04	0.000	9.94E-05	0.000	1.94E-04	0.000
12	5.72E-04	0.000	3.81E-04	0.000	3.32E-04	0.000
13	2.47E-04	0.021	1.51E-04	0.160	7.51E-04	0.000
14	1.01E-04	0.588	4.73E-05	0.801	5.81E-04	0.004
15	6.49E-04	0.000	3.49E-04	0.000	6.94E-04	0.000

Table 5.5 presents the effects of population on firm start-up. Similarly to what was depicted for the number of non-office firms, most of office firm types hold positive parameters and are consistent across the three components. These specifically refer to types 2, 4, 5, 10, 11, 12, 13, 14, and 15. Hence, the effect of an increase in population, observed in the current and past years, results in more office firms of these types

willing to start-up. Following the same pattern, but holding a negative coefficient is office firm type 7 (financial institutions). Although not expected, it means that the effect of an increase in population, observed in the current and past year, results in less financial institutions willing to start-up. The other office firm types also hold positive parameters, following similar interpretations. The only difference refers to the influence of lagged effects. Office firm types 3, 6, and 9 are influenced by the 1-year lag component; and types 1 and 8 are only influenced by the current component.

Table 5.5. Parameters for the variable POP

Firm type	Current		1-year lag		2-years lag	
	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]
1	4.00E-05	0.267	-2.04E-05	0.701	2.84E-05	0.630
2	9.23E-06	0.067	6.58E-06	0.227	2.95E-06	0.563
3	3.75E-05	0.156	3.38E-06	0.896	-1.65E-05	0.569
4	4.07E-05	0.047	2.50E-05	0.310	4.09E-05	0.050
5	1.64E-05	0.000	1.38E-05	0.000	2.56E-05	0.000
6	3.58E-05	0.081	6.00E-05	0.026	-7.30E-06	0.766
7	-6.20E-07	0.780	-3.72E-07	0.879	-4.69E-06	0.037
8	2.31E-05	0.100	-3.89E-06	0.811	1.20E-05	0.385
9	1.05E-05	0.000	6.00E-06	0.024	-8.56E-06	0.000
10	1.18E-05	0.000	1.67E-06	0.039	5.58E-06	0.000
11	2.40E-06	0.154	8.45E-06	0.000	8.93E-06	0.000
12	1.77E-05	0.005	1.46E-05	0.025	1.18E-05	0.073
13	2.44E-05	0.002	1.47E-05	0.066	6.10E-06	0.404
14	2.00E-05	0.197	2.48E-05	0.174	4.16E-05	0.028
15	1.62E-05	0.002	3.42E-05	0.000	9.86E-06	0.036

Regarding the influence of the size of the economic sector related to the modeling type (NJE), whose figures are presented in Table 5.6, different patterns apply. It can be observed that office firm types 1, 3, 8, 11, and 13 (respectively, agriculture; basic infrastructures; social security; computer and information technology; public administration) hold positive coefficients for the three components. Office firms of these types are motivated to start-up when the size of the related sectors increases. On the other hand, following the same pattern but holding negative coefficient, are firm types 4 and 9 (building industry and real estate). These are discouraged to start-up if the size of their sector increases.

Still holding positive coefficients, but with different patterns regarding the influence of lagged effects are office firms of type 7 (financial institutions), type 15 (environmental services, culture and recreation), and types 6, 12, and 14 (traffic and communication; research and development; and education and health). These are motivated to start-up when the size of the relative sectors increases. In contrast, holding negative coefficients, are type 10 (business services), and types 2 and 5 (industry, and retail and horeca). Therefore, the influence of the lagged effects can be observed accordingly. These office firms are discouraged to start-up if the size of their sector increases.

Table 5.6. Parameters for the variable NJE

Firm type	Current		1-year lag		2-years lag	
	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]
1	7.05E-04	0.923	3.78E-03	0.244	4.64E-03	0.380
2	-2.83E-04	0.000	1.04E-04	0.219	-3.65E-06	0.963
3	3.70E-04	0.525	1.40E-04	0.771	1.45E-04	0.762
4	-2.36E-04	0.855	-1.82E-03	0.098	-2.08E-03	0.004
5	-1.08E-04	0.329	1.89E-04	0.042	-3.22E-05	0.725
6	-1.50E-05	0.983	1.70E-03	0.017	1.42E-03	0.037
7	5.91E-05	0.000	8.60E-05	0.000	-5.76E-06	0.578
8	1.26E-05	0.946	3.15E-04	0.180	1.58E-04	0.529
9	-9.91E-04	0.000	-3.09E-04	0.000	-2.89E-04	0.003
10	-7.74E-05	0.000	-3.04E-06	0.297	8.19E-05	0.000
11	6.95E-05	0.000	2.21E-05	0.001	5.74E-05	0.000
12	-3.91E-04	0.004	3.53E-04	0.055	3.25E-04	0.063
13	3.99E-05	0.218	2.23E-04	0.000	3.62E-05	0.405
14	-1.58E-03	0.017	4.04E-04	0.640	1.99E-03	0.033
15	4.87E-04	0.000	-6.36E-05	0.566	2.34E-04	0.024

Moving to the regional effects, as shown by Table 5.7, the outcomes are obtained considering the Randstad area as the base category, which is used to contrast the results. Office firms are considered to start-up in the Randstad area when both the North and the South regions result in negative coefficients. These refer to types 9, 12, 13, and 14 (respectively, real estate; research and development; public administration; and education and health). Conversely, when both North and South regions result in positive coefficients, office firms will start-up outside the Randstad area. These refer to types 2, 3, 4, 5, and 15 (respectively, industry; basic infrastructures; building industry; retail and horeca; and environmental services, culture and recreation). When the coefficient for the North region is positive and the one for the South is negative, office firms will start in the Northern region. This is the case of social security companies (type 8). On the other hand, office firms related to agriculture; traffic and communication; financial institutions; business services; and computer and information technology (respectively, types 1, 6, 7, 10, and 11) will start in the Southern region, as the coefficient for the North is negative and for the South is positive.

Table 5.7. Parameters for the variable related to regional effects

Firm type	North		South	
	Coefficient	P[|Z|>z]	Coefficient	P[|Z|>z]
1	-4.22E-01	0.698	7.04E-01	0.456
2	4.68E-01	0.000	2.71E-01	0.008
3	6.22E-01	0.283	1.04E+00	0.049
4	1.17E+00	0.040	1.51E+00	0.001
5	1.89E-01	0.009	7.44E-01	0.000
6	-3.73E-01	0.467	2.07E-01	0.642
7	-1.46E-02	0.771	6.14E-02	0.166
8	1.39E-01	0.685	-2.72E-01	0.380
9	-2.41E-01	0.000	-3.64E-02	0.463
10	-4.05E-02	0.023	2.60E-01	0.000
11	-1.60E-01	0.000	8.85E-02	0.015
12	-1.26E-01	0.388	-5.63E-02	0.674
13	-3.95E-01	0.041	-7.17E-01	0.000
14	-1.28E+00	0.001	-2.61E-01	0.419
15	6.54E-01	0.000	1.19E-01	0.265

Finally, the influence of location attributes can be examined in the figures presented by Table 5.8. Looking at the density of public transportation facilities, given by the number of bus, tram and metro stops, the estimated coefficients are positive for all firm types and most of them are significant. That is, office firms are encouraged to start-up where the level of such facilities is higher. If this variable were considered to be a proxy measure for urban density, it means that denser areas encourage start-ups.

Regarding the distance to highway junction, the estimated coefficients are positive (and significant) for most firm types as well. Hence, the proximity to highway junction does not stimulate office firm start-up. However, this is in line to what one might expect if highway junctions were considered to be located in the periphery of a city, for example. Therefore, central areas or neighborhoods farther from highways are preferable for the start-up of some office firm types. Nevertheless, for the ones with a negative coefficient, the proximity to highway junctions stimulates start-up. This is the case of, for example, companies related to traffic and communication, research and development, and education and health; although only for research and development the coefficient resulted significant.

Analyzing the influence of the Schiphol international airport, most firm types hold a negative coefficient, i.e., the proximity to the airport stimulates start-up, although only some of them result significant. That is the case of types 2, 4, 7, and 15 (respectively, industry; building industry; financial institutions; and environmental services, culture and recreation). On the other hand, positive and significant coefficients can be observed for firm types 9, 10, 11, and 12 (respectively, real estate; business service; computer and information technology; and research and development), which are not encouraged to start-up by the proximity to the airport.

The proximity to shopping centers has a significant influence on office firm start-up, resulting in a negative and mostly significant coefficient for all firm types. Therefore, the expectation is that office firms start-up near shopping centers. Regarding the proximity to IC train stations, the resulting significant coefficients are positive. This holds for firm types 2, 7, 9, 10, 11, and 15 (respectively, industry; financial institutions; real estate; business services; computer and information technology; and environmental services, culture and recreation), which are not stimulated to start-up by the proximity to train stations. Lastly, rent prices result in positive coefficients for most firm types. Although it is not expected that firms would start-up if prices increase, a possible explanation is that higher rent prices mean better quality of a location. Therefore, the improvement of a location could stimulate office firm start-ups.

Table 5.8. Parameters for the variables related to location attributes

Var.	Firm type	Coefficient	P[|Z|>z]	Var.	Firm type	Coefficient	P[|Z|>z]
Density of public transportation facilities	1	5.92E-05	0.836	Distance to highway junction	1	2.23E+00	0.442
	2	1.91E-04	0.000		2	1.27E+00	0.000
	3	3.14E-04	0.010		3	1.46E+00	0.259
	4	1.04E-04	0.337		4	5.60E-02	0.964
	5	1.78E-04	0.000		5	1.39E+00	0.000
	6	1.59E-04	0.270		6	-6.71E-01	0.614
	7	8.41E-05	0.000		7	2.72E+00	0.000
	8	1.49E-04	0.053		8	1.75E+00	0.016
	9	1.84E-04	0.000		9	2.88E+00	0.000
	10	2.09E-04	0.000		10	1.72E+00	0.000
	11	1.60E-04	0.000		11	1.30E+00	0.000
	12	1.96E-04	0.000		12	-1.08E+00	0.001
	13	2.25E-04	0.000		13	6.64E-01	0.087
	14	1.93E-04	0.036		14	-5.58E-02	0.950
	15	2.44E-04	0.000		15	2.04E+00	0.000
Distance to Schiphol international airport	1	-1.04E+00	0.615	Distance to shopping centers	1	-1.82E+00	0.311
	2	-4.85E-01	0.014		2	-8.64E-01	0.000
	3	-1.51E+00	0.141		3	3.71E-01	0.673
	4	-3.06E+00	0.001		4	-8.38E-01	0.315
	5	-2.34E-01	0.071		5	-1.38E+00	0.000
	6	-6.29E-01	0.497		6	-1.66E+00	0.062
	7	-1.29E+00	0.000		7	-3.41E-01	0.000
	8	-4.02E-03	0.995		8	-1.19E+00	0.038
	9	3.41E-01	0.000		9	-7.52E-01	0.000
	10	2.07E-01	0.000		10	-1.10E+00	0.000
	11	6.63E-01	0.000		11	-1.42E+00	0.000
	12	1.06E+00	0.000		12	-1.17E+00	0.000
	13	-1.68E-02	0.956		13	-1.08E+00	0.001
	14	2.19E-01	0.739		14	-8.54E-01	0.210
	15	-1.55E+00	0.000		15	-4.74E-01	0.008

Table 5.8. Parameters for the variables related to location attributes

Var.	Firm type	Coefficient	P[|Z|>z]	Var.	Firm type	Coefficient	P[|Z|>z]
Distance to IC train station	1	1.14E+00	0.400	Rent price / m²	1	-6.42E-03	0.817
	2	3.56E-01	0.007		2	2.59E-02	0.000
	3	1.09E+00	0.108		3	1.26E-02	0.361
	4	8.08E-01	0.182		4	1.87E-02	0.141
	5	-3.10E-03	0.969		5	1.35E-03	0.466
	6	1.09E+00	0.072		6	6.03E-03	0.636
	7	1.85E-01	0.005		7	8.75E-03	0.000
	8	5.38E-01	0.171		8	9.08E-03	0.342
	9	7.65E-01	0.000		9	2.78E-02	0.000
	10	4.57E-01	0.000		10	2.65E-02	0.000
	11	1.76E-01	0.000		11	1.76E-02	0.000
	12	-8.50E-02	0.627		12	2.22E-02	0.000
	13	1.20E-01	0.599		13	-1.10E-02	0.030
	14	-4.52E-02	0.927		14	-1.41E-02	0.166
	15	1.04E+00	0.000		15	2.42E-02	0.000

5.5 Conclusions and discussion

The findings discussed in this chapter allow the following conclusions to be drawn. The aim was to develop a component of a firm demographic approach to examine the start-up process of office firms within a LUTI framework. The methodology used here considers that in order to simulate firm behavior, a firm must exist. The fact of coming into existence can only be modeled by taking an area as a decision unit. Therefore, instead of simulating the firm start-up behavior of individual firms, this component is intended to be part of the simulation of an environment where firms come into existence every year. To model that, the Poisson regression was used. It is able to accommodate the nature of (count) data used in this study, investigating the frequency of firm start-ups as a function of various location attributes included in the analysis.

As for the specific findings, one of the aspects addressed here regarded the influence of agglomeration economies, which was tested by some variables related to the density of office firms and non-office firm types. The presence of office firms, both of same and different types along with their size, influences the start-up of new office firms. This has already been reported in the literature and the findings here are in line with this. Despite specific patterns across various office firm types, the overall outcome suggests that there is a significant effect of agglomeration economies on the start-up process of office firms. Some additional tests performed here also indicated that events occurred in the past might be important, as the effects of several lagged variables turned out to be significant. These were, however, not exhaustively explored here. Data from a longer period would be needed to refine the investigations in this matter.

Regarding the location influences that may arise in the start-up process as addressed in this chapter, it can be observed that increases in population levels also motivate new office firm start-ups. This can be explained by basically two reasons: the presence of more people means not only a higher demand for services, but also more potential entrepreneurs willing to start a business. Also, regional effects apply and the preference for regions to start-up is fairly well defined. Although differentiated by only three major regions, it can be observed that specific office firm types display specific preferences to start-up either inside or outside the Randstad area; if the latter is the case, the preference is for the Southern rather than the Northern area.

As for the influence of accessibility to transportation, the findings here partially support earlier findings suggesting that the start-up process is not significantly influenced by transportation infrastructures. An exception should be, however, addressed for the accessibility to public transportation, which proved to have an influencing effect. Nevertheless, the issue might be pointed out to the fact that the density of public transportation facilities is directly related to the urban density of a region. Therefore, denser areas could encourage office firm start-ups. This is also in line with the results obtained from price levels and shopping infrastructures. These are somehow related to locations with better quality levels, which could eventually motivate office firm start-ups.

In sum, by estimating the number of firms starting-up in certain areas based on observed location conditions, this model can be used in a firm demographic approach, defining the (initial) parameters of a simulation framework. As there may be a limit in terms of location conditions for evolving firms, this simulation framework should count on some sort of mechanisms that control for restrictions inherent to the location attributes. This is addressed in the next chapter, where the concept of carrying capacity is developed along with investigations on a modeling alternative for that end.

6

Carrying capacity model

6.1 Introduction

This chapter presents the carrying capacity model. In a firm demographic approach, a very important difference compared to the demography of people is the presence of interactions between demand and supply. These interactions are the underlying foundations of firm start-up, growth and decline, which translate the notion of market that drives firm dynamics. If demand is high, there will be new firms starting-up, existing firms tending to grow, and less firms going out of business. However, these interactions are bounded by the resources available in a given spatial unit, which can be used by the supply side. Overall, this defines the notion of carrying capacity, an ecological concept that measures a maximum population (usually of animals or people, but here referred to firms) that can be supported by the conditions of a defined environment (Hannan and Freeman, 1989; Carroll and Hannan, 2000).

Following the above, van Wissen (2000 and 2003) investigated the carrying capacity notion for a multi-sector set of organizations in The Netherlands. He assessed to what extent the existing number of Dutch firms in a region has an impact on further development of economic activity. In this respect, he assumed that a region has a maximum limit in the number of firms and jobs that can be accommodated depending on the number of consumers and workers in various sectors. If the number of firms exceeds the carrying capacity, market stress increases, with negative implications for firm growth, birth and closure probabilities. The methodology uses a spatial input-output framework that defines the market demand of each economic sector in each location, which accounts for the interactions among industry sectors and for the spatial distribution between demand and supply. Hence, the model claims to operationally capture the concept of agglomeration economies. Overall, the findings indicated a promising direction in exploring the carrying capacity notion in firm demography.

From an urban planning perspective, along with the market view of carrying capacity, it is also important to consider the resources available at the urban environment for determining economic development. Energy, green areas, transportation infrastructures, urban public facilities (shops, schools, etc.), among other types of assets should be taken into account when investigating carrying capacity, as these will be demanded in the related development process. Population aspects would also be important sources of demand, especially regarding the search for services provided by firms. At the same time, population is an important source of labor force, which imposes a maximum level of employment that a region can support. For example, economic development would only occur if there is demand for that, but it may be limited to the existing conditions of the (skilled) labor force.

Oh et al. (2005) have investigated the urban carrying capacity to evaluate the environmental conditions upon urban growth, as a way to define strategies to plan and manage urban dynamics, accounting for sustainability issues. More recently, Lane (2010) has also explored carrying capacity methodologies to assess sustainable land-use planning. Although these studies are not directly in line with the investigations of the topic aimed at this specific chapter, they illustrate some of the developments regarding the carrying capacity notion in the urban planning field. They view carrying capacity more in terms of sustainability, whereas the focus here is on assessing the available resources or demand in the urban environment for economic development. Nevertheless, it is interesting to register such investigations.

The question now moves toward how to implement the considerations regarding the carrying capacity notion into the overall modeling framework defined in this research project. On the one hand, firm dynamics rely on market prospects, and, on the other, the urban environment constraints such dynamics in terms of available resources. For the LUTI framework for office firm dynamics developed here, a modeling approach is proposed in this chapter, trying to elaborate the carrying capacity notion in such overall simulation model of this research project. This component is intended to work along with the demographic processes of firm start-up, growth and decline, by controlling such processes according to the existing conditions of the environment. The next section will describe the modeling approach investigated to this end, followed by an empirical application. The results and analyses are then presented and some concluding remarks are drawn.

6.2 Modeling approach

The model presented in this chapter regards a component of a multi-agent system specified to incorporate the carrying capacity concept into the study of firm dynamics.

From the principles of Economic Base Analysis (EBA), market dynamics can be distinguished as processes resulting from basic and non-basic industries. Basic industries regard those firms whose businesses depend on external factors, exporting their products to outside regions. Non-basic industries, in contrast, usually depend on local business conditions, whose products are consumed by the local demand. These principles also suggest that economic activities are often bounded by the resources available within a production environment, but at the same time, require a minimum level of supply in order to maintain the business activities.

In economic modeling, stochastic frontier analysis (SFA) has a long history in applied production economic systems, and could be useful to examine the above related principles. SFA was originally introduced by Aigner et al. (1977), and Meeusen and van den Broeck (1977), who proposed the production frontier model as given by Equation 6.1.

$$y_i = f(x_i;\beta)TE_i \tag{6.1}$$

where y_i is the observed output of producer i; x_i is a vector of inputs; $f(x_i;\beta)$ is the production frontier; β is a vector of coefficients to be estimated; and TE_i comprises the technical efficiency, which is defined as the ratio of the observed output to the maximum feasible output ($TE_i \leq 1$).

Apart from TE_i, which is assumed to be a stochastic variable, another stochastic (normally distributed) component is added, **exp** $\{v_i\}$, to account for statistical noise. If TE_i = **exp** $\{-u_i\}$ is written, where $u_i \geq 0$ (now defined as technical inefficiency) and u_i assumes some distribution forms, such as: normal-half normal, normal-exponential, normal-gamma, or normal-truncated normal, the model can be written as Equation 6.2:

$$y_i = f(x_i;\beta)e^{\{-u_i\}}e^{\{v_i\}} \tag{6.2}$$

Usually, the model takes the log-linear Cobb-Douglas form, as shown by Equation 6.3.

$$ln\, y_i = \beta_0 + \sum_{n=1}^{n}(\beta_n\, ln\, X_{ni}) + v_i - u_i \tag{6.3}$$

Equation 6.3 is defined as the stochastic *production* frontier. The dependent variable is the actual production. This is a function of the maximum amount (the frontier) that can be obtained from a given input, but subtracted the technical inefficiency (although the return in any production system is expected to be maximum, there is always some inefficiency). SFA also presents a *cost* frontier form, defined when $v_i + u_i$ is written instead. In this case, the dependent variable represents costs of production associated to a given set of inputs. The frontier then tends to represent a minimum or a baseline, but there is always some inefficiency as well. Thus, the technical inefficiency is summed to the frontier, regarding to an extra amount considered on the top of this "ideal" baseline. The notion of carrying capacity implies that a production form applies to the case of office firms. In the present analysis, production is measured as the size of employment in the sector of interest and the input variable as the size of the population in a region. The population imposes a maximum to the production either as a resource of labor, in case of basic industries, or as limiting the demand for the products, in case of non-basic industry. In other words, if the notion of carrying capacity holds for the firm type of interest, the expectation is that a production specification for the frontier model better fits the data than a cost function specification. By trying both functional forms, the analysis presented here will test the validity of this notion for each sector. The next section will detail the empirical application.

6.3 Empirical modeling setup

The office firm data used concerns the 2006 LISA register. COROP areas were used as units of observation, given their strategic delineation on capturing the functional relationships across regions. The dependent variable is the number of jobs (aggregated in these areas) in each of the 15 office firm economic sectors. These 15 types of office firms were used to estimate and analyze the models separately. The set of independent variables considered includes, in addition to population size, measures of transportation infrastructure facilities to also account for possible influence of accessibility on production. More specifically, the set of independent variables, calculated at the level of COROP areas as well, includes:

1. Population (number of inhabitants);
2. Average income of households (Euro);
3. Total length of highways (km);
4. Straight distance (km) of the COROP centroid to the closest airport;
5. Straight distance (km) of the COROP centroid to Schiphol international airport;

6. Straight distance (km) of the COROP centroid to the closest intercity (IC) train station;
7. Density of public transportation facilities.

It is important to mention that, ideally, other variables should be included, as a way to capture the existing location attributes, making the model more comprehensive. In line with this, the use of straight distances to calculate accessibility to transportation infrastructures should be changed, probably to an average calculated distance obtained from each individual firm, similar to what was done for the start-up model. Also, the investigation using the other periods would be interesting, to acquire the dynamic behavior over time. On the top of those issues, more importantly, this empirical modeling setup lacks variables describing agglomeration economies, as a way to account for market demands among multiple sectors. However, given the exploratory nature of this analysis, it was decided to stick to the results of this initial empirical application setup only. The expected gain from this empirical application is, therefore, on investigating the findings obtained through the stochastic frontier analysis in light of the carrying capacity notion.

Still regarding this empirical application, the initial set of inputs comprises 7 independent variables. Although there are 40 units of observation, which makes statistically possible the estimation of 39 parameters, there is a rule of thumb usually followed in empirical applications stating that between 10 and 20 observations are minimally required to reliably estimate one parameter, as a way to reduce the risk of overfitting. If the present dataset contains 40 observations, about 3 parameters can be estimated (excluding the constant term). Hence, as 7 variables are initially defined, these have to be tested in steps in order to find out their significance levels and eventually compose a set of about 3 parameters for the final model. To this end, several combinations of 3 out the 7 variables were performed, for each office firm type. The partial results will not be presented here. Instead, the resulting significant parameters comprise, overall, *population*, *distance to IC train stations*, and *distance to Schiphol international airport*. For some office firm types, however, only some of these are significant.

In sum, the related stochastic frontier model is estimated using the software NLOGIT version 4.0 (Greene, 2007). Both production and cost functions are tested across the datasets for each office firm type. As for the term u_i, only a normal-half normal distribution is adopted for this exploratory analysis, although other distributions could be tested as well. The findings are presented in the next section.

6.4 Analyses and results

A stochastic frontier model is estimated for each office firm type separately. However, the model for office firm type 1 (agriculture) did not give significant results, probably due to its low number of firms (consequently, number of jobs) observed across COROP areas. Table 6.1 presents the results, where "z" regards the statistical z-test used for estimating the parameters. It was mentioned in the previous section that population, distance to IC train stations, and distance to Schiphol international airport are, in general, the significant variables. In fact, population has the main influence in the sense that its coefficients are significant across all models. But distance to IC train stations and distance to Schiphol international airport are only significant for some models. In addition, these variables produce errors related to the estimation of singular variance matrices if included in the model specification of some office firm types. Hence, these variables were excluded, as marked with a "*" in Table 6.1.

Table 6.1. Estimated coefficients of the stochastic frontier model

Office firm type	Function type	Constant	P[|Z|>z]	Population	P[|Z|>z]	Distance to IC train	P[|Z|>z]	Distance to Schiphol	P[|Z|>z]
2	C	-7.655	0.001	1.263	0.000	-0.256	0.118	-0.471	0.002
3	P	-8.923	0.044	1.274	0.000	*	*	*	*
4	C	-8.834	0.041	1.131	0.000	-0.034	0.912	-0.521	0.066
5	P	-5.865	0.001	1.156	0.000	0.080	0.546	-0.495	0.000
6	C	-22.031	0.000	1.792	0.000	*	*	*	*
7	C	-7.201	0.000	1.158	0.000	*	*	*	*
8	P	-18.136	0.000	1.895	0.000	*	*	*	*
9	C	-5.856	0.000	1.075	0.000	-0.036	0.603	-0.165	0.005
10	P	-5.237	0.000	1.230	0.000	*	*	-0.221	0.000
11	C	-9.370	0.000	1.369	0.000	-0.079	0.345	-0.280	0.000
12	C	-15.700	0.000	1.657	0.000	-0.137	0.551	0.044	0.843
13	C	-5.833	0.000	1.143	0.000	-0.224	0.004	0.065	0.326
14	P	-21.124	0.000	2.148	0.000	0.616	0.056	-0.844	0.005
15	C	-7.464	0.000	1.096	0.000	-0.335	0.009	*	*

*(Excluded variables that produced errors related to singular variance matrices are marked with *)*

Given the above, the analysis carried out here is focused only on the effect of population in terms of examining the carrying capacity through the stochastic frontier model. The results in Table 6.1 show that a specific function type, production (P) or cost (C), is found for each office firm type. In the case explored here, it is not possible to choose between one or the other function a priori. The finding is that only one of them fits the data, which is determined by the skewness of the residuals obtained after the ordinary least squares (OLS) estimation. If this results in a wrong skew for the production function, usually the cost function will work, and vice-versa. Thus, for the exploratory analysis carried out here, the function type that fitted the distributional structure of the data is adopted.

The question now refers to the relationship of these findings with the notion of carrying capacity. It was discussed before that the production frontier could be associated with the carrying capacity of a region, as it would determine the maximum number of jobs that can be supported by the population. On the other hand, in cases where a cost function rather than a production function fits the data, the conclusion is that the notion of carrying capacity is not applicable. Such a finding indicates that the population defines a minimum rather than a maximum frontier. The size of the population, as a source of demand or supplier of labor, is not a limiting factor for the sector's size: the sector could grow further even if the population stays constant.

The results obtained from this exploratory study revealed that most office firm types have a cost frontier function specified, whereas only 5 office firm types are better described by a production function. It suggests that most office firm types are best represented by providing a baseline of employment rather than being subject to the carrying capacity. In other words, the size of the sector is not constrained by the size of the population. Those office firm types in which a cost function is found regard industry; building industry; traffic and communication; financial institutions; real estate; computer and information technology; research and development; public administration; and environmental services, culture and recreation. Nevertheless, this reasoning relies on the available knowledge, skilled labor force and industrial structure of a specific region.

Regarding the office firm types in which a production function is found, i.e., basic infrastructures; retail and horeca; social security; business service; and education and health, the size of the population limits the size of these sectors and they can be assumed to be subject to the carrying capacity. These sectors cannot produce more than the demand for its production and can neither produce more than the available labor force (or other resources) allows. This reasoning would then hold for consumer-oriented services that cannot be easily exported to other areas as well as for markets that can become saturated.

6.5 Conclusions and discussion

The findings discussed in this chapter allow the following conclusions to be drawn. The incorporation of strategies to account for market demand and control the availability of resources in firm demographic developments is very important. In line with this, the approach introduced by van Wissen (2000 and 2003) on using the carrying capacity notion constitutes an advance in this field and the exploration of alternative approaches could be fruitful to this end, especially in empirical applications. The exploratory model developed here tries to bring a contribution to this field using the stochastic frontier analysis.

In this specific approach, the carrying capacity of a region is represented by a frontier, which is obtained from a given set of inputs. The production frontier function is associated with the carrying capacity, as it gives the maximum output (in this case, measured as the number of jobs) based on an input, here given by the population in a local area. The stochastic frontier analysis can be also defined by a cost function. In this case, however, there is no association with the notion of carrying capacity. Instead, a minimum size might be a better description for some sectors that defines, therefore, a baseline of employment.

Overall, the findings suggested that most sectors are defined by a cost function, which indicates that they are not subject to the carrying capacity. The sector is not constrained by the size of the local population and it could grow further even if the population stays constant. However, it would depend on the available knowledge, skilled labor force and industrial structure of a specific region. On the other hand, for the sectors that a production function was found, the carrying capacity seems to apply in terms of the limits of the local population size. These sectors cannot produce more than the demand for its production and can neither produce more than the available labor force (or other resources) allows.

Despite the above, as for the empirical application, there is a need for further investigations in order to obtain more detailed results. It would involve the use of larger datasets and the inclusion of variables to account for market demand and agglomeration economies. Also, the level of detail could be increased for the accessibility measures, other types of distributions for the one-sided component error should be tested, and a dynamic analysis over the years could be carried out. However, from a theoretical perspective, the representation and modeling of the carrying capacity using stochastic frontier analysis seems plausible, allowing future investigations in this direction.

From a LUTI framework perspective, the notion of available resources at locations could be well captured, in the sense that the carrying capacity has a relationship with changes in population and transportation infrastructures. If population increases and accessibility levels are improved, the carrying capacity of a region increases. At a larger multi-agent simulation model, this carrying capacity component could serve as a base framework to mediate office firm demographic processes. This would work as a function to control the probability and intensity of start-up, growth and closure, by economic sector. Next chapter will move forward with the introduction of the location choice model, which comprises another component of the multi-agent framework developed in this research project.

7

Location choice model

7.1 Introduction

This chapter presents the location choice model. It constitutes a critical part of firm demographic models, especially to simulate the dynamics of the spatial distribution of firms, which in turn influences traffic flows and accessibility. As indicated before in the discussions about the investigations on firm development, traditional location decision theories evolved from classical concepts of profit maximizing in isotropic space, but were gradually replaced by behavioral ideas, derived from psychological theories (see section 2.3 for related references). As opposed to the idea of a best or optimal location defended by the economic theories, the behavioral theories were more concerned with a location expected to be satisfactory. This involved a location decision making process that was based on models and methods developed to this end. Later, with the importance of taking the firm's characteristics into account (i.e., structural factors and lifecycle), an organizational approach emerged, emphasizing the firm activity-related behavior. This approach claimed that location decisions should be considered along with the firm's business cycle.

In empirical applications, most models of location decisions of firms are based on the principle of utility-maximizing behavior from the Random Utility Maximization (RUM) theory (McFadden, 1974). It advocates that the probability of a firm selecting a particular location is seen as the result of maximizing the utility derived from the various characteristics of a large number of locations. In transportation research, specifically in LUTI systems, the use of these discrete choice models can be found in, for example, Waddell and Ulfarsson (2003); de Bok and Sanders (2005); de Bok and Bliemer (2006a and 2006b); de Bok (2007); Maoh et al. (2002 and 2005); Maoh and Kanaroglou (2007a); Sivakumar and Bhat (2007); and Elgar et al. (2009).

In contrast, as acknowledged by Witlox et al. (2004), the location choice process can also be modeled from other perspectives. They suggested that a qualitative approach, based on logical expressions (e.g., "if, then" statements) "have sufficient flexibility to represent a wider variety of decision rules". In addition, these rule-based models can deal with some problems usually found in choice models, such as: multi-collinearity among independent variables or complexity related to numerous included attributes. On the other hand, the authors agree that "their 'crisp' (or exact) nature implies the lack of an error theory, limiting in some cases the realism of such systems". Nevertheless, the development of multidimensional fuzzy systems was seen as an alternative to overcome these restrictions.

The approach taken in this chapter is, however, slightly different from what has been found in the literature. It assumes that location decisions are based on the firm's business model. Due to several possible options, inertia, limited information and costs involved, firms will neither necessarily take all characteristics into account nor maximize their profit. Instead, the location decision can be best viewed as the problem of matching a set of firm requirements and location characteristics. This underlying idea involves the decision of an adequate modeling approach, and the interpretation of the matching relationships upon the selected modeling approach. To this end, Bayesian networks (BN) represent an alternative modeling framework. They are comparable to a rule-based model, but account for probabilistic responses. This also make them similar to RUM models, but BN better take into account interdependency relationships among independent variables, by identifying and incorporating casual relationships from data. In addition, BN can handle high-dimensional data, given by potentially large number of variables and interdependency relationships.

Given the above, the objective of this chapter is to explore the use of BN for modeling location decisions of office firms. In fact, a special-form of BN is investigated, named Bayesian *Classifier* networks (BCN). As part of the larger multi-agent micro simulation system, this firm location choice component derives parameters that could be used to best represent the interdependency relationships between office firm requirements and location attributes. The chapter is organized as follows: a modeling approach is described, followed by an empirical application. The results and analyses are then presented and some concluding remarks are drawn.

7.2 Modeling approach

The model presented in this chapter regards a component of a multi-agent system specified to examine the location decision of office firms. The design of the model regards a data-mining technique, based on a classification task. It identifies class labels for instances based on a set of attributes. More specifically, the model analyzes the relationship of, on the one hand, office firm types (the class labels or dependent variable) and, on the other, the location attributes (the independent variables). Based on the fact that firms are already assigned to specific locations, i.e., for each location there is only one firm, if the attributes of each location were specified and linked to the firm characteristics housed at the specific location, a matching relationship could be determined. This would derive a classification of location attributes for each specific firm's requirements. It suggests that the model proposed here does not actually describe the choice of discrete locations, but the probability of having certain location characteristics related to each office firm type, which will eventually result in location choice patterns.

Several methodologies can provide a solution for this classification task. Inference methods can be used to compute the conditional probability of one node (the class variable) according to values assigned to the other nodes (the attribute variables). Studies have applied, for example, decision tree induction (Arentze et al., 2000; Thill and Wheeler, 2000; Wets et al., 2000) and ordinary Bayesian networks (Janssens et al., 2004 and 2006; Verhoeven et al., 2006; Arentze and Timmermans, 2009). However, the use of Bayesian Classifier networks (BCN) has shown to outperform more conventional approaches for classification problems, where the probability of one variable is to be updated given the evidence for all the other variables (Cheng and Greiner, 1999 and 2001).

As for BCN, these are a special form of BN. Specifically, a BN is a Directed Acyclic Graph where nodes represent (stochastic) variables and links represent causal relationships (Pearl, 1988). A Conditional Probability Table (CPT) is attached to each node. This CPT defines, for the node, the conditional probabilities related to possible states of the parent nodes, whose inputs are received from child nodes. Based on the conditional probabilities, node probabilities can be computed and updated at each time, by hard-evidencing the states of the nodes entered into the network. Applying this reasoning to the case studied here, a single parent node represents office firm types, whereas child nodes represent location attributes. Hence, for a given office firm type and a set of location attributes, the probability of each possible state of each child node can be determined, deriving the location characteristics that are more relevant for the given office firm type.

Following Cheng and Greiner (1999 and 2001), who have explored several forms of these BCN, two specific forms of BCN are considered for the analyses presented in this chapter. These regard the so-called Naïve Bayes and Bayesian augmented Naïve

Bayes (BAN). Figure 7.1 schematically represents these special forms. Naïve Bayes is a simple structure that has the class node (C) as parent node and all attributes (X) as child nodes. In this structure, no connections are allowed between the child nodes. As the structure is given, no structure learning is required. Computing the probability of the class node comes down to backward reasoning using as a principle concept the basic Bayesian method of belief updating, according to Equation 7.1.

$$P(c|x_1, x_2, \dots) = \frac{P(x_1, x_2, \dots |c)P(c)}{\sum_{x_1, x_2, \dots} P(x_1, x_2, \dots |c)P(c)} \tag{7.1}$$

where c is a particular state of the class variable C and x_1, x_2, … are particular states of attribute variables X_1, X_2, … .

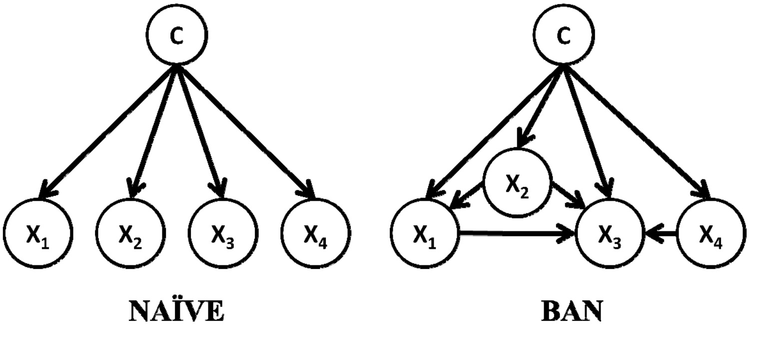


Figure 7.1. Schemes of Naïve Bayes and Bayesian augmented Naïve Bayes

This simple structure has been used for many years and appears to be surprisingly powerful. It works particularly well in cases where there are no strong dependencies among the attribute variables. Along with this, the BAN network is an extension that allows taking into account such interdependencies. To determine the interdependencies, this model involves network learning on the level of attribute variables, which is unrestricted.

A BN that defines the interdependency relationships among a set of variables can be learned based on data through a two-step process: 1) learn the structure of the network, referred to as structure learning, and 2) given the structure, learn the CPTs at the nodes, referred to as parameter learning. Despite this stepwise characteristic, these processes involve methods that have been developed independently from each other.

Parameter learning is rather straightforward. If there are no missing values in the data, this process is simply reduced to determining observed conditional frequencies for each child node and its parent nodes in the data. On the other hand, machine-learning and data-mining fields have predominantly focused on algorithms for structure learning (Andersen et al., 1989). Two groups of algorithms have emerged (Cheng et al., 2002): scoring-based learning methods, and constraints-based learning methods.

Scoring-based methods view a BN as a structure defining a joint probability distribution across the variables included in the network. These methods search for the structure that maximizes a goodness-of-fit on the observed joint probability distribution in the data. In contrast, constraints-based methods rely on tests of conditional independency among nodes to determine whether or not the nodes should be interconnected. Theoretically, constraints-based methods are better suited than scoring-based methods for developing classifier networks. Therefore, this is the method used in the analyses presented in this chapter. Further details are provided in Pearl (1988); Spiegelhalter et al. (1993); and Heckerman et al. (1995).

A basic concept is the mutual information between two given nodes, which is defined as:

$$I(A,B) = \sum_{a,b} P(a,b) log \frac{P(a,b)}{P(a)P(b)} \tag{7.2}$$

where $I(A,B)$ is the mutual information between nodes A and B; a and b represent possible states of A and B; $P(a, b)$ is the joint probability of $A = a$ and $B = b$; and $P(a)$ and $P(b)$ are the (marginal) probabilities of these states.

The existence of mutual information is not a sufficient condition for a link between two nodes, as the influence may also run through other nodes. Constraint-based algorithms use the *d*-separation concept: two nodes are *d*-separated when, loosely speaking, they are conditionally independent given possible paths through other nodes. The problem of finding the correct structure for a given set of variables is a NP-hard problem and, therefore, existing algorithms use heuristic search. The algorithm used in the current study, i.e., the Three-Phase Dependency Analysis (TPDA), uses a three-staged procedure:

1. drafting a network;
2. thickening the network, which adds links to the draft; and
3. thinning the network, which removes unnecessary links.

The resulting links are undirected. In a final step, an algorithm is applied to direct the links as far as possible by identifying the so-called collider structures. Links that remain undirected, if any, are presented to the user for making a decision based on knowledge about the domain. Furthermore, the TPDA algorithm uses a threshold parameter in conditional independency tests, meaning that conditional independence is falsified only if the (conditional) mutual information exceeds the threshold. This parameter has an influence on the complexity of a learned network: keeping everything else equal, the higher the threshold, the lower the expected number of links, and vice versa.

In sum, it is expected that the more flexible forms outperform the more restricted forms. Hence, the expectation is that BAN outperforms Naïve Bayes. However, overfitting the data is an (increasing) threat for the more flexible structures. The conditional probabilities in CPTs attached to nodes represent parameters that are fitted on training data. Too many of such parameters given the size of a training sample may result in rules that do not generalize well to unseen cases. Therefore, there are two complementary ways to reduce the risk of overfitting:

1. Reducing the complexity of the network consequently reduces the number of parameters. It can be controlled by the threshold parameter (at least for the BAN model, whose derivation involves a structure-learning phase), i.e., the larger the threshold, the smaller the risk of overfitting.
2. Feature selection is a well-known data mining technique used to reduce the number of attribute variables in the learning phase in a meaningful way (see Witten and Frank, 2005).

Regardless the approach taken, a holdout set needs to be used to test overfitting and to evaluate performance of a trained classifier on unseen cases (i.e., the holdout cases). Having outlined the theoretical framework along with the modeling approach for examining the location choice of office firms, the next section will detail the empirical application.

7.3 Empirical modeling setup

The DTZ data was used in the analyses presented in this chapter. The reason for using it instead of the LISA data refers to the fact that the DTZ was first obtained, at the time of the development of the location choice model. Also, regardless its limited number of observations compared to the LISA data, only the DTZ data contained the information about the office space that firms occupied, which could be used as an additional location variable.

Regarding the information about firm's characteristics available in the data, only office firm type was available. Therefore, this is used as an indicator variable of firm's requirements. Using the BCN terminology, the office firm types comprise the class variable (C) that will work as the dependent variable in the classifier structures. The 17 original categories of office firm types had to be merged into a smaller number of categories. For modeling reasons, the original amount of office firm types turned out to be very large in relation to the limited number of existing observations in the dataset. It resulted in a (sufficiently even) distribution of 9 classes of office firm types.

The independent variables (or 'X' attributes) comprised the socioeconomic and spatial aspects, detailed below in Table 7.1. As the BCN modeling requires discrete data, existing continuous variables were discretized, maintaining an even balance between the number of categories and the frequency in each category. Table 7.1 presents an overview of the results of these pre-processing steps, describing the variables, the categories obtained for each variable along with their meaning and the number of cases in each category.

From the modeling approach described previously, two forms of BCN are investigated here, i.e., Naïve Bayes and BAN. In addition, varying the network complexity by means of the threshold parameters and the use of feature selection (FS) comprise ways to evaluate the performance of the models. A question addressed in this analysis regards finding a BCN setup that best represents the problem under investigation. Therefore, various network setups were tested, as will be detailed in the sequel. However, in order to evaluate the performance of these BCN, a benchmark was defined based on a decision tree, which was induced for the same classification problem using the same data. Several decision tree induction methods can be used for this purpose, but a classifier that predicts class labels in a probabilistic manner is required. Hence, CHAID was chosen as it is more sensitive to full probability distributions of the class variable compared to tree growing-pruning methods such as C4.5 and CART. More specifically, Exhaustive CHAID was applied, because when compared to the basic CHAID method, it uses an improved splitting criterion.

Given the above, three major model structures outline the present analyses: Exhaustive CHAID (as a benchmark), Naïve Bayes and BAN. Turning back to the various BCN setups explored, some forms of FS were tested, especially due to the large set of attribute variables included in the office firm database. Regarding this FS, three approaches were compared:

Table 7.1. Description of the discrete variables

Variable	Cat.	Meaning	Frequency
Office firm type	A	Building industry; Industry; Agriculture; Mineral extraction	193
	B	Hotel and catering industry; Commercial and professional services; Production and distribution of energy, gas and water	1,209
	C	Health and welfare; Education	263
	D	State organizations; Public services and social security	177
	E	Repair of consuming products; Real estate and retail services	398
	F	Computer services and information technology	296
	G	Environmental, culture and recreation services	368
	H	Financial institutions	218
	I	Traffic and communication infrastructures	120
Dutch provinces	A	Groningen; Friesland; Drenthe	143
	B	Flevoland; Overijssel	229
	C	Gelderland	304
	D	Utrecht	450
	E	North Holland	805
	F	South Holland	775
	G	Zealand; North Brabant; Limburg	536
Urbanization (municipality level)	A	Very highly urbanized area (more than 2,500 addresses/km²)	1,245
	B	Highly urbanized area (between 1,500 and 2,500 addresses/km²)	1,231
	C	Moderately urbanized area (between 1,000 and 1,500 addresses/km²)	581
	D	Lowly or non-urbanized area (less than 1,000 addresses/km²)	185

Table 7.1. Description of the discrete variables

Variable	Cat.	Meaning	Frequency
Urbanization (post code level)	A	Very highly urbanized area (more than 2,500 addresses/km²)	1,061
	B	Highly urbanized area (between 1,500 and 2,500 address/km²)	463
	C	Moderately urbanized area (between 1,000 and 1,500 addresses/km²)	314
	D	Lowly urbanized area (between 500 and 1,000 addresses/km²)	387
	E	Non-urbanized area (less than 500 addresses/km²)	1,017
Distance to closest airport	A	Less than 10 km	511
	B	Between 10 and 25 km	1,186
	C	Between 25 and 50 km	699
	D	Between 50 and 75 km	421
	E	More than 75 km	425
Distance to Schiphol	A	Less than 15 km	406
	B	Between 15 and 45 km	921
	C	Between 45 and 65 km	649
	D	Between 65 and 100 km	445
	E	More than 100 km	821
Distance to closest international train station	A	Less than 3 km	568
	B	Between 3 and 12 km	1,023
	C	Between 12 and 25 km	507
	D	Between 25 and 50 km	395
	E	More than 50 km	749
Distance to closest intercity train station	A	Less than 2 km	613
	B	Between 2 and 5 km	1,057
	C	Between 5 and 12 km	1,060
	D	More than 12 km	512

Table 7.1. Description of the discrete variables

Variable	Cat.	Meaning	Frequency
Distance to closest highway junction	A	Less than 0.5 km	330
	B	Between 0.5 and 1 km	648
	C	Between 1 and 1.8 km	929
	D	Between 1.8 and 3 km	891
	E	More than 3 km	444
Distance to closest shopping center	A	Less than 0.5 km	504
	B	Between 0.5 and 1 km	644
	C	Between 1 and 1.8 km	886
	D	Between 1.8 and 3 km	708
	E	More than 3 km	500
Population	A	Less than 4,000 inhabitants/postcode area	865
	B	Between 4,000 and 7,000 inhabitants/postcode area	643
	C	Between 7,000 and 10,000 inhabitants/postcode area	768
	D	More than 10,000 inhabitants/postcode area	966
Number of households	A	Less than 1,500 households/postcode area	761
	B	Between 1,500 and 3,000 households/postcode area	503
	C	Between 3,000 and 4,500 households/postcode area	790
	D	Between 4,500 and 6,000 households/postcode area	731
	E	More than 6,000 households/postcode area	457
Labor force	A	Less than 2,000 employees/postcode area	912
	B	Between 2,000 and 3,500 employees/postcode area	700
	C	Between 3,500 and 5,000 employees/postcode area	765
	D	More than 5,000 employees/postcode area	865
Average income	A	Less than 30,000 Euro/postcode area	643
	B	Between 30,000 and 40,000 Euro/postcode area	786
	C	Between 40,000 and 50,000 Euro/postcode area	656
	D	Between 50,000 and 60,000 Euro/postcode area	542
	E	More than 60,000 Euro/postcode area	615

Table 7.1. Description of the discrete variables

Variable	Cat.	Meaning	Frequency
Number of places in schools	A	Less than 250 places/postcode area	672
	B	Between 250 and 1,000 places/postcode area	532
	C	Between 1,000 and 2,000 places/postcode area	738
	D	Between 2,000 and 4,000 places/postcode area	590
	E	More than 4,000 places/postcode area	710
Number of parking places	A	Less than 2,500 places/postcode area	1,019
	B	Between 2,500 and 5,000 places/postcode area	1,233
	C	Between 5,000 and 7,500 places/postcode area	660
	D	More than 7,500 places/postcode area	330
Percentage of green areas	A	No green area/postcode area	897
	B	Between 1 and 3 %/postcode area	434
	C	Between 4 and 9 %/postcode area	649
	D	Between 10 and 23 %/postcode area	627
	E	More than 24 %/postcode area	635
Number of shopping centers within 1 km	A	No shopping centers	1,512
	B	Just 1 shopping center	602
	C	Between 2 and 3 shopping centers	395
	D	4 or more shopping centers	733
Office floor space	A	Less than 180 m^2	636
	B	Between 180 and 300 m^2	676
	C	Between 300 and 500 m^2	644
	D	Between 500 and 1,000 m^2	681
	E	More than 1,000 m^2	605

1. Use of only the attributes identified in the CHAID decision tree induced as the benchmark. The subset of relevant attributes indentified is pre-selected for learning the BCN;
2. Use of all attributes in the data for network learning, meaning that the learned connections between attribute and class variable define automatically the selection of attribute variables. These are used for updating the probabilities of the class variable in each case during the prediction phase. Thus, the attribute selection is a component of BN learning in addition to the learning process of attribute-attribute connections;
3. Test the approach without feature selection, meaning that all attributes are (forced to be) included in the network.

Still regarding the network setups, in case of all models, different threshold (T) values were explored to find an optimal value in each case. The default value of 1.0 (no unit) is used to start and, then, the evaluation of the outcomes is carried out by either increasing or decreasing this threshold value. For threshold values higher than 1.0, only strong relations between attributes and the class variable will be found. In contrast, for threshold values lower than 1.0, also relatively weak relations will be found.

To summarize, the combination of the several modeling structures (CHAID, Naïve Bayes, and BAN) along with the explored network setups (regarding FS and variation in threshold values) derived the following specific models tested in this analysis:

1. *Exhaustive CHAID*: used as a benchmark.
2. *Naïve CH-FS*: predefined X-C links; no X-X links allowed; selection of features based on CHAID; no structural learning.
3. *BAN CH-FS*: predefined X-C links; structural learning of X-X links; selection of features based on CHAID.
4. *Naïve BN-FS*: structural learning of X-C links; no X-X allowed; selection of features during structural learning.
5. *BAN BN-FS*: structural learning of X-C links; structural learning of X-X links; selection of features during structural learning.
6. *Naïve No-FS*: predefined X-C links; no X-X links allowed; all the features included; no structural learning.
7. *BAN No-FS*: predefined X-C links; structural learning of X-X links; all the features included.

In addition to the above mentioned, two remarks should be made. First, the learning of X-C links is constrained by the fact that 'C' is defined as a root node, i.e., it cannot have incoming links. Second, to test the behavior of the BCN without any type of restriction, an additional model was included. This model was built without any feature selection, no predefinition of X-C or X-X links, neither constrains about root or leaf nodes. Thus, this comprises an ordinary BN model.

Finally, to evaluate the performance of the models, 75 % of the sample was used for training and the remaining 25 % was used for validation (test or holdout set). Two measures of prediction accuracy were used: the *ordinary* hit ratio and the *expected* hit ratio. The first represents the proportion of correctly predicted cases when predictions are generated in a deterministic fashion, whereas the latter represents the expected proportion of correctly predicted cases when predictions are generated in a probabilistic fashion. Deterministic prediction means that the class label having the highest predicted probability is predicted for each case. On the other hand, probabilistic prediction considers the case where Monte Carlo simulation would be used to predict the class label in each case. The expected number of correctly predicted cases is then given by the sum of predicted probabilities of the actual class labels across all cases. The expected hit ratio expresses this number as a ratio of the total number of cases. For the analyses carried out here, this is considered to be a more informative indicator of the goodness-of-fit than the ordinary hit ratio, given that a probabilistic prediction is more interesting for the sake of generating unbiased population distributions.

7.4 Analyses and results

Table 7.2 presents the results of the models. The class variable refers to the type of office firms, comprising 9 classes. A first observation is that the relative performance of the Naïve and BAN types of BCN compared to the benchmark (the CHAID-based decision tree) differs depending on which criterion is used for evaluating prediction accuracy. In terms of the *deterministic* R (the ordinary hit ratio), BCN do not perform better than the decision tree model, whereas in terms of the *probabilistic* R (the expected hit ratio), the relative performance depends on the chosen specification. As said previously, the probabilistic R is a more relevant criterion, and, therefore, it will be used to evaluate the performance of the models.

First, comparing the Naïve model and the BAN model under basic settings, that is, no FS and a threshold value of 1.0, the more complex BAN structure increases the prediction accuracy on the test set substantially (from 0.188 to 0.265). This means that taking interdependency relationships among attributes into account leads to an important improvement of prediction accuracy, at least in this case. At the same time, the drop in performance between the training set and test set is bigger for the BAN model (0.421 to 0.265 versus 0.218 to 0.188). This suggests that the BAN model is more sensitive to overfitting the training data, as expected, due to the increase in complexity (and number of parameters) of the network that it entails.

Table 7.2. Results for the explored models

Model			Deterministic R		Probabilistic R	
Structure	**FS**	**T**	**Training**	**Test**	**Training**	**Test**
Exhaustive CHAID	Not applicable	Not applicable	0.372	0.375	0.206	0.210
Naïve	CH-FS	1.0 (= any T)	0.372	0.352	0.215	0.201
BAN	CH-FS	1.0 (= 1.2; 1.4; 1.6)	0.441	0.332	0.285	0.226
Naïve	BN-FS	1.0	0.303	0.256	0.218	0.190
		1.2	0.302	0.269	0.217	0.191
		1.4	0.311	0.262	0.215	0.190
		1.6	0.315	0.272	0.207	0.186
BAN	BN-FS	1.0	0.400	0.354	0.251	0.211
		1.2	0.395	0.353	0.266	0.213
		1.4	0.397	0.347	0.245	0.208
		1.6	0.388	0.353	0.221	0.200
Naïve	No-FS	1.0 (= any T)	0.297	0.246	0.218	0.188
BAN	No-FS	1.0	0.512	0.305	0.421	0.265
		1.2	0.509	0.318	0.428	0.271
		1.4	0.514	0.303	0.425	0.266
		1.6	0.495	0.303	0.411	0.261
Unrestricted BN	Not applicable	1.0	0.400	0.354	0.251	0.211
		1.2	0.395	0.353	0.266	0.213
		1.4	0.397	0.347	0.245	0.208

Results also show that increasing the threshold value can reduce overfitting in case of the BAN model and, as a result, improve the performance of this model on the test set. Specifically, increasing the threshold from 1.0 to 1.2 increases the performance of the BAN model on the test set from 0.265 to 0.271. The performance of the Naïve model could not be improved by means of this instrument.

On the top of the above mentioned, benefits can be expected from FS. As it shows, the Naïve and BAN models respond differently in this regard. Performance of the Naïve model increases when some form of FS is used, whereby CHAID-based FS yields a bigger improvement (from 0.188 to 0.201) than BN-based FS (from 0.188 to 0.191). On the other hand, the BAN model does not benefit from FS: it even deteriorates the performance in case of the BAN model, both when CHAID-FS or BN-FS are used.

Coming back to the question on how the BCN perform relatively to the benchmark, it is noticed that it depends on the specific type of BCN. The Naïve model does not give a satisfactory result in this comparison. Under the best specification, i.e. CHAID-based FS, the expected hit ratio of the Naïve model equals 0.201 on the test set, which is lower than the hit ratio of 0.210 that the decision tree model attains. On the other hand, the BAN model does lead to an improvement of performance under all specifications in terms of FS and threshold tested in this analysis. The improvement is substantial. Under the best specification (no FS and a threshold of 1.2) the expected hit ratio on the test set is equal to 0.271, implying an improvement in prediction accuracy of 28.6 % compared to the benchmark.

Table 7.2 also shows the performance of an ordinary BN, i.e., a network that results when no learning restrictions are applied. As it appears, the performance of this model under the best threshold setting (T = 1.2) is significantly lower than that of BAN (0.213 versus 0.271). This indicates that BCN outperform ordinary BN when used for classification, at least in case of this dataset.

Given the above, the results suggest that BCN, when properly specified, outperform CHAID-based decision tree models in terms of prediction accuracy. The explanation for this refers to the existing significant dependency relationships among the attribute variables. That is, when taken these into account, it seems to improve prediction accuracy significantly. Furthermore, the BAN classifier also outperforms ordinary BN. This may seem counter-intuitive as ordinary BN imposes fewer restrictions on learning. It should be noted, however, that interdependency learning algorithms do not optimize a goodness-of-fit function, whereas treating, alternatively, the class variable as a root node allows the learner to take a full set of attributes into account in defining class node probabilities.

Finally, as for FS, the results do not show any benefit. A likely explanation is that

FS methods consider only relationships between attribute variables, on the one hand, and the class variable, on the other. The specific power of BAN, however, originates from its ability to take also interdependencies among attribute variables into account. Since FS-based methods do not consider this, their purpose for BAN is limited, at best.

As opposed to the above, Naïve BCN are more likely to benefit from FS. Because these do not use information contained in attribute-by-attribute interdependencies, attributes that do not have direct effects on the class variable can be safely removed. Nevertheless, there might be a good reason to consider FS in case of BAN. FS does reduce the risk of overfitting. A consequence of overfitting is that rules extracted from data are somewhat biased in the sense that they do not fully generalize to unseen cases. Hence, in applications where (behavioral) interpretation of revealed patterns (rules) rather than prediction accuracy is the primary purpose, FS may be considered. Obviously, the extent to which overfitting is a concern depends also on the sample size. For such larger databases, a tension between prediction accuracy and bias would decrease.

7.5 Structural relationships between office firm types and location characteristics

The interpretation of the matching relationships between type of office firm and location characteristics is best revealed by the BAN No-FS (T = 1.2) model. It is depicted in Figure 7.2, showing the way a learned network can be used by evaluating how the probability of each attribute's category changes when a particular class is hard-evidenced (e.g., a specific office firm type is selected). Thus, at the top of Figure 7.2 is the general configuration of the network. Since no evidence is entered, the probabilities represent a-priori beliefs. On the bottom, the updated beliefs are presented when a selected firm type is entered as hard evidence (represented in the network under the label "MyLab"). By instantiating the Mylab node successively with the firm types considered, the updated probabilities related to each attribute become apparent. Table 7.3 summarizes the findings of 9 hard evidence scenarios, related to the 9 types of office firms considered here.

Examining the results for office firm type "A" (industries, agriculture and mineral extraction) suggests that these firms are mostly located in medium-sized or big office buildings, with an average floor space of about 250 m² or more than 1,000 m². They seek a location in very low urbanized neighborhoods in high and moderate urbanized municipalities, especially in the South Holland region. These firms prefer locations not too close or too far from shopping areas (distance of about 1.5 km), and seem to be relatively satisfied with any level of service with regard to some facilities such as parking places or schools. Concerning the population aspects, they are likely to locate in zones with less than 4,000 inhabitants/postcode area and less than 1,500 households/postcode area. In terms of accessibility, office firms type "A" locate not very far from airports and train stations, and about 1 km from a highway junction.

Results of the analyses indicate that office firms in commercial and professional services (type B) are mostly located (i) in small and medium-sized office buildings, up to 500 m²; (ii) in highly and very highly urbanized municipalities, with preference for very highly urbanized neighborhoods, although very lowly urbanized neighborhoods also seem of interest for these firms; (iii) cities in the South Holland region, or cities such as Amsterdam, Hilversum or Haarlem situated in the North Holland region; (iv) within a moderate distance from shopping areas, with a moderate number of schools in the area; (v) in zones with more than 10,000 inhabitants/postcode area and more than 5,000 employees/postcode area; and (vi) within a moderate distance from airports and train stations, but not very close to highway junctions (between 1 and 3 km).

Analyzing the findings for office firms related to health and education services (type C), it can be observed that these firms are mainly located (i) in medium to large office buildings of more than 500 m²; (ii) in highly urbanized municipalities and very highly urbanized neighborhoods; (iii) predominantly in regions of North Holland and South Holland; (iv) not very close to shopping centers; (v) in zones with a high number of schools (more than 4,000 places/postcode area) and between 3,000 and 4,500 households per postcode area, where the average income is between 30,000 and 40,000 Euro; and (vi) close to intercity train stations and between 1 and 3 km from the highway junctions.

State organizations, public services and social security companies (type D) are predominantly located in very large office buildings (i.e., more than 1,000 m²) in highly urbanized municipalities and neighborhoods, mainly in the South Holland region, but also in North Holland and Utrecht. This type of office firms will be found far from shopping centers (between 1 and 3 km), and seem to be quite satisfied with any level of service in terms of parking places provided. They are located mainly in zones with a relatively high number of places in schools, and in highly populated zones (more than 10,000 inhabitants). This type of office firms seems to have strict requirements of being located close to transport infrastructure, except for the highway junctions.

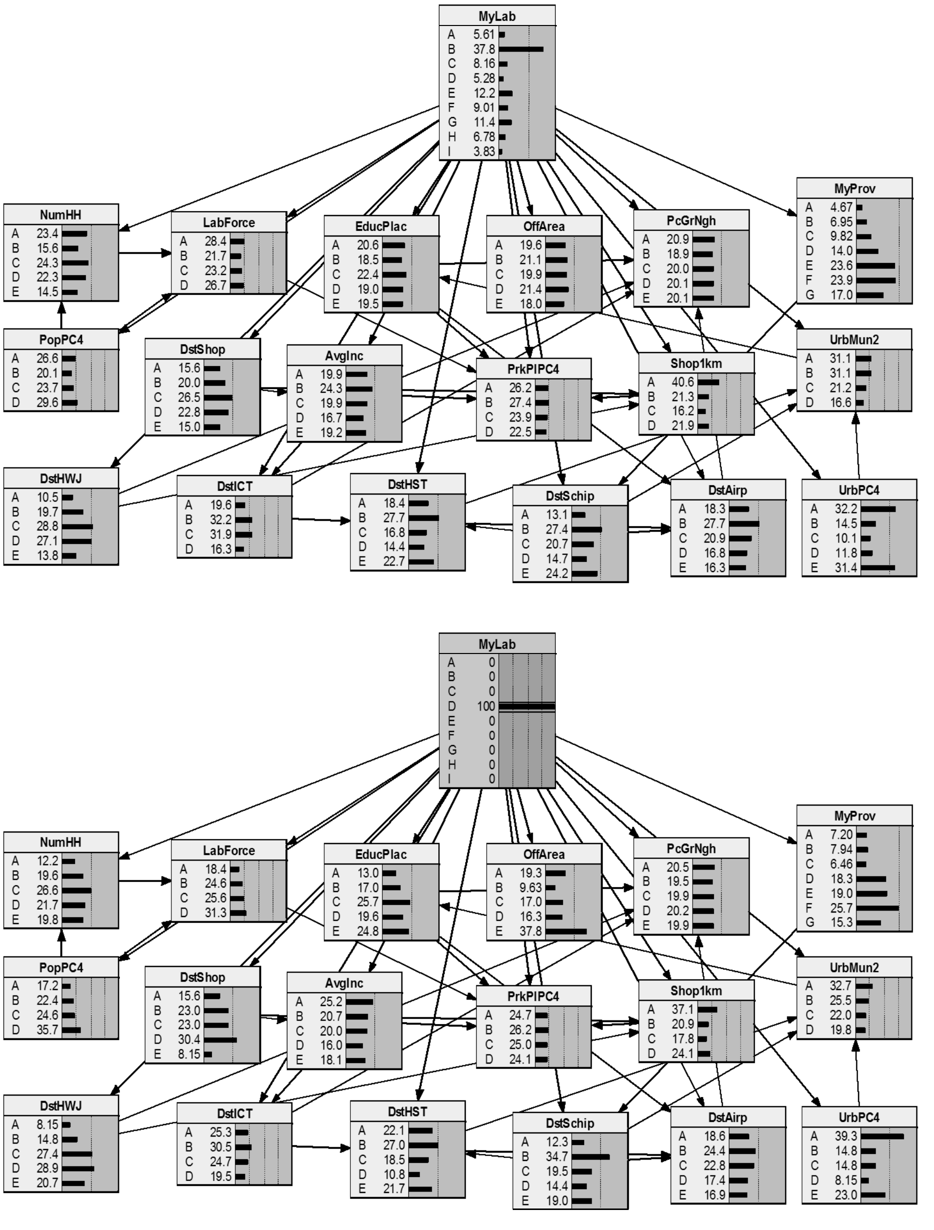


Figure 7.2. Representation of the BAN No-FS (T=1.2) model

Table 7.3. Summary of the results of the BCN for each office firm type

Attribute	Cat.	Office firm type class (values in %)								
		A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Dutch provinces	A	4.7	3.8	5.7	7.2	4.5	5.2	4.5	6.8	3.8
	B	5.4	6.4	10.5	7.9	6.1	6.0	9.0	6.8	4.8
	C	9.6	9.8	12.0	6.5	11.0	10.4	9.4	8.6	8.8
	D	13.8	13.5	14.9	18.3	10.4	16.6	16.0	14.4	9.8
	E	21.5	21.8	21.7	19.0	28.7	23.7	24.5	27.7	29.0
	F	27.1	26.2	23.2	25.7	22.1	17.9	19.5	20.2	35.1
	G	18.0	18.5	12.0	15.3	17.2	20.1	17.1	15.5	8.8
Urbanization (municipality level)	A	26.4	34.3	29.8	32.7	29.3	26.7	30.0	30.3	27.8
	B	28.8	34.6	32.1	25.5	27.5	30.6	31.9	25.7	25.3
	C	25.0	18.9	20.0	22.0	24.2	22.0	20.8	23.2	25.7
	D	19.8	12.2	18.0	19.8	19.0	20.7	17.4	20.9	21.2
Urbanization (post code level)	A	23.8	34.2	43.7	39.3	27.5	24.2	35.4	30.2	17.2
	B	15.4	15.2	13.6	14.8	13.4	15.0	14.4	12.2	13.1
	C	12.6	8.6	12.6	14.8	9.2	11.5	10.9	9.9	8.1
	D	13.3	11.3	8.7	8.2	12.7	10.6	13.7	16.3	12.1
	E	35.0	30.6	21.4	23.0	37.3	38.8	25.6	31.4	49.5
Distance to closest airport	A	19.4	18.0	16.8	18.6	18.0	20.7	17.2	17.3	22.5
	B	25.4	31.0	25.7	24.4	27.7	23.9	26.0	26.8	23.6
	C	19.8	20.8	21.4	22.8	19.6	20.7	23.0	20.2	19.0
	D	18.4	16.0	17.5	17.4	18.1	16.9	15.6	18.4	17.0
	E	17.0	14.3	18.6	16.9	16.5	17.7	18.2	17.2	18.0

Table 7.3. Summary of the results of the BCN for each office firm type

Attribute	Cat.	Office firm type class (values in %)								
		A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Distance to Schiphol	A	14.3	11.6	10.4	12.3	16.7	13.7	11.1	16.2	21.5
	B	24.3	26.9	29.6	34.7	25.5	23.4	30.2	29.8	23.8
	C	24.5	21.8	20.9	19.5	20.6	19.3	16.4	18.1	25.8
	D	12.5	14.0	12.5	14.4	14.4	19.9	16.7	13.4	14.1
	E	24.4	25.7	26.5	19.0	22.8	23.8	25.6	22.6	14.7
Distance to closest international train station	A	14.8	19.3	20.3	22.1	15.0	15.4	19.4	18.3	20.2
	B	30.2	28.1	22.6	27.0	31.4	26.5	24.6	29.2	30.2
	C	19.2	16.0	18.1	18.5	16.5	15.9	17.4	15.7	18.7
	D	14.3	13.2	18.0	10.8	16.0	16.7	13.8	14.0	15.0
	E	21.5	23.4	21.0	21.7	21.0	25.5	24.8	22.8	15.8
Distance to closest intercity train station	A	12.4	22.9	16.9	25.3	14.8	18.3	17.8	19.4	18.6
	B	30.7	33.5	36.9	30.5	27.7	28.4	34.9	32.9	28.8
	C	37.1	30.9	27.5	24.7	39.1	35.3	29.9	28.8	33.9
	D	19.8	12.7	18.7	19.5	18.5	17.9	17.4	18.9	18.7
Distance to closest highway junction	A	11.2	8.9	7.8	8.2	11.4	15.4	12.3	10.5	16.2
	B	21.7	19.8	18.4	14.8	25.8	20.3	14.0	19.8	21.2
	C	39.2	25.9	28.6	27.4	30.1	27.3	31.2	33.7	27.3
	D	18.9	29.0	31.6	28.9	19.9	28.2	29.1	25.6	26.3
	E	9.1	16.4	13.6	20.7	12.7	8.8	13.3	10.5	9.1

Table 7.3. Summary of the results of the BCN for each office firm type

Attribute	Cat.	Office firm type class (values in %)								
		A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Distance to closest shopping center	A	11.9	17.1	18.0	15.6	14.1	15.0	13.0	15.1	17.2
	B	14.0	21.5	25.2	23.0	13.7	17.6	23.2	25.0	7.1
	C	28.7	26.1	24.3	23.0	29.7	23.8	31.9	24.4	21.2
	D	25.2	20.4	24.3	30.4	23.9	26.4	22.1	15.7	31.3
	E	20.3	14.9	8.3	8.2	18.6	17.2	9.8	19.8	23.2
Population	A	30.2	26.4	21.5	17.2	30.5	29.6	19.4	25.7	51.8
	B	17.7	21.1	22.5	22.4	18.4	16.0	21.5	18.2	19.4
	C	25.3	22.4	25.8	24.6	25.9	21.3	27.1	27.5	10.4
	D	26.7	30.1	30.2	35.7	25.2	33.1	32.0	28.6	18.4
Number of households	A	27.6	23.9	15.2	12.2	28.6	26.7	17.1	18.7	47.6
	B	15.5	13.9	19.3	19.6	14.5	12.5	17.2	20.1	16.2
	C	21.4	24.2	29.4	26.6	23.0	22.9	25.7	27.2	13.0
	D	22.8	21.9	20.0	21.7	24.0	24.9	24.2	20.5	15.6
	E	12.7	16.0	16.0	19.8	9.9	13.0	15.8	13.5	7.6
Labor force	A	32.1	28.1	23.2	18.4	32.3	30.7	21.4	29.3	51.3
	B	18.6	21.1	25.2	24.6	20.6	19.2	25.5	21.0	18.8
	C	23.4	21.7	27.4	25.6	24.5	22.7	26.7	23.5	12.3
	D	25.8	29.1	24.2	31.3	22.6	27.4	26.3	26.1	17.6
Average income	A	19.3	19.7	24.1	25.2	17.1	18.2	20.0	17.2	22.6
	B	18.4	24.5	31.9	20.7	22.2	23.8	27.0	25.0	19.1
	C	20.9	19.5	17.2	20.0	20.0	20.6	19.7	20.8	24.2
	D	19.1	15.9	10.9	16.0	19.9	19.5	15.9	18.3	17.2
	E	22.2	20.3	15.9	18.1	20.7	17.9	17.3	18.7	17.0

Table 7.3. Summary of the results of the BCN for each office firm type

Attribute	Cat.	Office firm type class (values in %)								
		A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Number of places in schools	A	22.8	20.8	16.6	13.0	24.1	20.3	16.0	19.2	39.1
	B	18.2	18.1	17.8	17.0	22.2	18.8	18.7	16.8	16.2
	C	20.7	25.3	22.1	25.7	19.8	17.9	21.0	24.0	13.5
	D	20.6	18.8	18.6	19.6	15.6	24.2	18.4	19.7	17.8
	E	17.6	17.0	24.9	24.8	18.3	18.7	26.0	20.3	13.4
Number of parking places	A	26.0	26.5	25.3	24.7	27.0	26.0	25.0	25.6	29.2
	B	25.4	28.9	26.6	26.2	27.2	26.7	27.7	26.0	23.7
	C	24.7	23.5	24.7	25.0	23.4	24.0	24.0	24.7	23.9
	D	23.9	21.1	23.4	24.1	22.5	23.3	23.3	23.7	23.2
Percentage of green areas	A	19.6	22.0	20.5	20.5	19.8	20.5	20.9	20.0	20.2
	B	19.4	18.4	18.7	19.5	19.3	19.4	18.9	19.4	19.8
	C	20.2	19.9	20.0	19.9	20.2	19.7	19.9	20.2	20.0
	D	20.5	19.9	20.9	20.2	19.5	20.5	20.5	19.7	19.6
	E	20.3	19.8	19.9	19.9	21.2	19.8	19.8	20.7	20.3
Number of shopping centers within 1 km	A	40.8	42.5	33.7	37.1	45.4	42.1	37.3	33.9	42.4
	B	18.5	19.9	24.6	20.9	20.4	23.2	24.4	22.9	19.3
	C	21.5	13.8	19.2	17.8	17.7	15.8	13.6	19.7	20.3
	D	19.2	23.7	22.5	24.1	16.4	18.9	24.7	23.5	18.1
Office floor space	A	11.2	23.1	9.2	19.3	21.9	18.9	20.0	18.0	15.2
	B	23.8	22.8	21.4	9.6	23.5	18.5	19.6	16.3	27.3
	C	23.1	21.0	15.0	17.0	16.7	25.6	24.6	12.2	16.2
	D	18.9	20.8	26.2	16.3	19.6	23.3	21.8	27.9	16.2
	E	23.1	12.3	28.2	37.8	18.3	13.7	14.0	25.6	25.3

The results of office firm type “E” (repair of consuming products, real estate and retail services) suggest that these firms are located in small and medium office buildings, occupying about 200 m², in highly urbanized municipalities, in either very low or very highly urbanized neighborhoods in the North Holland, South Holland, Zealand, North Brabant and Limburg regions. These firms prefer zones farther than 1 km from shopping centers and zones with fewer places in schools. They are primarily located in lowly populated zones and with less than 2,000 employees/postcode area. As for accessibility aspects, locations not too far from airports and train stations or from highway junctions are preferred.

Office firm type “F” concerns firms that deal with computer services and information technology. Medium-sized (about 400 m²) office buildings in lowly urbanized neighborhoods in highly urbanized municipalities, mainly in the North Holland region, but also in Utrecht, Zealand, North Brabant and Limburg are preferred. These firms can be found in zones within a moderate distance from shopping areas, in highly populated zones with a relative number of places in schools (about 3,000 places/postcode area). Regarding accessibility aspects, locations slightly close to airports and train stations, but farther from highway junctions are preferred.

Environmental, culture and recreation services, represented by office firm type “G”, seem to require medium or large office buildings (between 300 and 1,000 m²) in very high urbanized municipalities and neighborhoods, primarily in North Holland, but also in South Holland, Utrecht, Zealand, North Brabant and Limburg. These firms are predominantly located at rather close distance from shopping areas in highly populated zones with a high number of places in schools. Results also suggest that this type of office firm seek locations not too far from airports and train stations, or between 1 and 3 km from the highway junction.

The results obtained for office type “H” (financial institutions) suggest that they require (i) large to very large office buildings (more than 500 m²) in highly urbanized areas, mainly in the North Holland region, which is the core region of the Netherlands; (ii) locations very close to shopping centers and in zones with more than 7,000 inhabitants/postcode area; (iii) locations not too far from airports and train stations, and within 1 and 3 km from a highway junction.

Finally, office firm type “I” representing companies related to traffic and communication infrastructures seem to require either small to medium (between 180 and 300 m²) or very large (more than 1,000 m²) buildings. They seem to be satisfied in municipalities with any level of urbanization, but prefer very lowly urbanized neighborhoods, probably to avoid any conflict with residential use. Consequently, locations often have few schools, few labor force, and few inhabitants and households. Their most critical location requirement seems to be a location very close to highway junctions.

7.6 Conclusions and discussion

The findings discussed in this chapter allow the following conclusions to be drawn. The current developments found in LUTI frameworks dealing with the location choice modeling typically comprise discrete choice models based on utility-maximizing principles. However, as opposed to these models, the approach taken here was different in the sense that firms will not consider all locations and related attributes for location decision making. Instead, the location choice problem was viewed as a matching process of firm's location requirements and location attributes. To this end, a methodology was developed, based on a data-mining technique. More specifically, a classification task was addressed, where office firm types, on the one hand, and location attributes, on the other, were used to establish the matching relationships. As a way to model these, BCN were explored and compared to decision tree models, usually applied for classification tasks.

As for more specific findings obtained from the empirical application, BCN turned out to be a valuable approach, outperforming decision trees in terms of prediction accuracy. The specific power of the BCN originates from the fact that they take into account possible interdependency relationships among attribute variables in the prediction of a class variable. Overfitting training data is a general concern in BCN learning, and therefore the use of feature selection was considered to minimize such effects. However, the value of current feature selection methods is limited for the forms of BCN that take indirect effects of attribute variables on the class variable into account. Feature selection may reduce overfitting, but, at the same time, decreases the ability of a learned classifier to account for network effects. In applications, different models could be tried to identify the best approach for the specific dataset and the specific modeling purpose at hand. Furthermore, techniques such as k-fold cross-validation can be used to increase the size of the training set and yet obtain insights in the prediction accuracy (on unseen cases) of a model. For the firm database explored here, an increase of almost 30 % in prediction accuracy for unseen cases was found.

A general concern regards, however, how the approach can be used for location choice predictions. In fact, in the way that the model is built to perform the classification task, firms are selected as opposed to locations, comprising an inverse mechanism as it would be expected in a typical location choice model. That is, as the case studied here, the model gives as output office firm types based on the location characteristics treated as inputs. In a location choice model, firms (along with their characteristics) are used as inputs, resulting in a set of possible locations to accommodate them. The results depicted in Table 7.3 allow such reasoning, where the probabilities of the various location characteristics are estimated according to each office firm type. Nevertheless, this model still gives only the probabilities of choosing specific location characteristics. What it lacks is a spatial component, where concrete locations can be defined. If those selected location characteristics could be matched, for example, with a geographical database of available locations, such location decision mechanism could be derived and an allocation process between firm and location could be established.

Some general findings in substantive terms regard the fact that office firms are predominantly located in the Center-Northwest region of The Netherlands, in the so-called Randstad area. Next, the Northwest-Center-Southeast comprises a major corridor. Office firms in the industry and infrastructure sector require a less urbanized neighborhood to locate their business, probably to avoid any conflict with residential use. These firms are more likely to be located close to highway junctions, and usually at the outskirts of the cities. Conversely, firms that provide services to the population, such as commercial and professional companies, and education, health, culture, recreation and financial institutions are, in general, located mainly in central areas, with good accessibility and at a moderate distance from facilities such as schools, parking spaces and shopping areas.

Next chapter will continue with the introduction of the growth model, which comprises another firm demographic component of the multi-agent framework developed in this research project.

8

Growth model

8.1 Introduction

This chapter presents the growth model. The process of firm growth/decline is often associated with the changes in firm size, which in turn influences economic development. If growth is positive, new jobs are created, reflecting good perspectives in economic development. On the other hand, when growth is negative (or decline), there may be job losses and firms going out of business, resulting in a decline in the economy.

Firm growth has been examined in the literature under several formulated hypotheses. In economy, the traditional Gibrat's law (Gibrat, 1931) of proportionate growth states that firm size and growth rate are independent. However, this assumption resulted incorrect in empirical applications. According to a review carried out by Maoh and Kanaroglou (2005), the findings obtained by Evans (1987), for example, indicated that firm growth decreases when firm size increases. This relationship is not linear, though, as variations across firm size distributions could also be observed. Similarly, Hart and Oulton (1996) found out that, overall, growth has a negative relation with initial firm size. But when stratified values of firm size were taken, they observed that smaller firms had larger growth. In contrast, there was no fundamental relationship between firm size and growth for larger firms.

Firm size is not the only factor influencing firm growth, though. As reviewed by de Bok (2007) and Raspe and van Oort (2011), firm structural factors in the industry sector are also determinants of firm growth. These include specifically age, which suggests that growth decreases with aging firms (Jovanovic, 1982; Evans, 1987), and type of economic activity, indicating that the emergence of dominant technologies can lead to the growth of specific economic sectors (Teece, 1986; Breschi et al., 2000). Van Wissen (2000) also discusses that growth is highly influenced by market conditions and may be triggered by lifecycle aspects, such as: relocation and merges. Relocation can be associated with an improvement in the location of the firm, for attracting new markets. Firm merges are related to, for example, shared market coverage or improved efficiency on using shared production technologies and labor force.

While the effects of firm internal factors have received a large attention on influencing firm growth, external factors have been, to some extent, rather overlooked. This is reported by Hoogstra and van Dijk (2004), who argued that "while in theory it is widely acknowledged that 'location' should be considered as a relevant growth determinant, empirical research has so far mainly focused on firm internal factors". The authors reveal that the specification of location or environment is often poor, sometimes regarding to a simple introduction of proxy measures to capture region-specific influences, for example. Hence, the authors investigated a set of explanatory variables related to location characteristics that were assumed to have influence on firm growth. This resulted in a general conclusion that location should not be disregarded on explaining firm growth.

From an urban planning perspective, the findings reported above are expected. Investigations in the field, specifically in LUTI models, have addressed firm growth as a function of transport infrastructure. De Bok (2007) reviews that empirical investigations (Hilbers et al., 1994; Rietveld, 1994; Chandra and Thompson, 2000; de Bok and Bliemer, 2006a and 2006b) revealed that the proximity to highways has a positive influence on firm growth for some economic sectors. The specific findings obtained by de Bok (2007) are also in line with this. In addition to the proximity to highways, the author considered the proximity to train stations. He reported that these had a significant effect on firm growth for various industry sectors.

Despite the above, there are some shortcomings in the empirical applications discussed. Although some efforts could be observed on incorporating location attributes for investigating firm growth, these are mostly limited to the influence of highways, for example. Hence, the examination of other transportation infrastructures is required. But location attributes should not be limited to accessibility to transportation only. Population aspects, urbanization levels, effects of agglomeration economies, urban facilities (parking, schools, etc.) may also be part of the process. Therefore, these are to be reviewed. Also, as for modeling approaches, methodologies should be developed in order to account for the factors influencing firm growth in an appropriate way.

Given the above, the objective of this chapter is to contribute to this field, investigating firm growth and its relation with location characteristics, as part of a LUTI system of office firm demography. It tries to extend the existing knowledge by using more detailed data about location attributes and by selecting a modeling approach capable to accommodate these attributes accordingly. An approximation based on the logistic growth function is tested to this end. The chapter is organized as follows: the modeling approach is described, followed by an empirical application. The results and analyses are then presented and some concluding remarks are drawn.

8.2 Modeling approach

The model presented in this chapter regards a component of a multi-agent system specified to examine the process of firm growth in relation to a set of characteristics. Following Tsoularis and Wallace (2002), a simple exponential curve can fit the growth patterns of most populations at initial stages. However, a population would not grow indefinitely, as constrains would apply, such as: predation, competition, saturation, and limited environmental resources available. Therefore, it is not realistic to consider growth as an unrestricted process.

Numerous curves have been applied to model growth, but the classical Verhulst logistic growth function (Verhulst, 1838) has been the base of most empirical applications found in the literature. What the Verhulst model does is to consider a saturation level, regarded as an upper bound on the growth size, generally associated with the concept of carrying capacity. As a way to account for this limit in the exponential curve, Verhulst introduced the logistic growth function. It has an 'S-shape' curve, in which growth is approximately exponential for initial stages, followed by a saturation stage in which growth slows down, and ending at a maturity stage in which growth stops.

A general form that is assumed for the logistic growth function can be formulated as in Equation 8.1.

$$S_i^t = \frac{S_i^{max}}{1 + e^{\,[\beta_i(\alpha_i - t)]}} \tag{8.1}$$

Equation 8.1 gives the size (S_i^t) of an individual (a firm in this case) at time t, as a function of the maximum firm size (S_i^{max}) and some parameters (α_i, β_i) to be estimated. However, if the interest is in a function that returned the firm size at time $t + T$ for any given value of firm size at time t, i.e.:

$$S_i^{t+T} = f(S_i^t) \tag{8.2}$$

By solving (8.1) for t (Arentze and Timmermans, 2006):

$$t = -\ln \frac{S_i^{max} - S_i^t}{S_i^t \; e^{(\beta_i \, \alpha_i)}} \tag{8.3}$$

From (8.1) it is further known that the size at T time steps after t is:

$$S_i^{t+T} = \frac{S_i^{max}}{1 + e^{\,[\beta_i(\alpha_i - t - T)]}} \tag{8.4}$$

Substituting (8.3) in (8.4) and rewriting the equation, the logistic growth function is given in the format of (8.2):

$$S_i^{t+T} = \frac{S_i^{max}}{1 + \left(\frac{S_i^{max}}{S_i^t} - 1\right) e^{\,(-\beta_i T)}} \tag{8.5}$$

where S_i^{t+T} is the predicted size of firm i at a time $t + T$; S_i^t is the current size of firm i at time t; S_i^{max} is the asymptotic maximum size of firm i; β_i is a parameter to be estimated; and T is a time step.

The question underlying Equation 8.5 is how it can be used to model growth of firms, as proposed here. The logistic growth function has a stage in which the predicted size increases at an accelerated rate, followed by a decelerating rate stage, and ending at a point in which the predicted size equals to its maximum. However, firms can also have a decline stage, in which decreasing values of predicted size would be observed, but these cannot be well represented by the logistic curve. Alternatively, if "growth" over time could be considered instead of "size" at a given time moment, it would result in a more convenient form in terms of parameters estimation. In fact, this is what is aimed at this approach, as well as having the estimation of such a form with the use of simple linear methods such as linear regression analysis.

Given the above, an approximation is proposed. The derivative of the logistic function describes the change in size (growth) and has a 'bell-shaped' form. Growth increases with size in an initial stage and starts to decrease when size is at the inflection point. However, this is still a not very suitable form that could be estimated by linear methods. In turn, this bell-shaped curve might be roughly represented by a quadratic function, i.e.:

$$S_i^{t+T} - S_i^t = a_i(S_i^t)^2 + b_i(S_i^t) + c \tag{8.6}$$

where a_i, b_i, and c are parameters to be estimated; and $\left(S_i^{t+T} - S_i^t\right)$ is the predicted growth for a given time step as a function of the current size S_i^t.

The quadratic $(S_i^t)^2$ and linear (S_i^t) terms are included in the right-hand side of the equation to make sure that growth might be an increasing function for initial values of current size, followed by a decreasing function for larger values of current size. That is, the expected shape of the predicted curve is a parabola opening downward, i.e., $a < 0$. From this approximation, one can also observe that Equation 8.6 allows negative values for growth, which can accommodate the firm decline process. As opposed to the logistic growth function, this approximation is expected to compose a more flexible approach, capable of accommodating different sets of firm size-development patterns that could be observed in a given sample.

Another issue raised with the development of this firm growth model regards the investigation of the influence of location attributes on this process. The underlying question is how these can be included in the approximation given by Equation 8.6. The location attributes can be taken into account as interaction terms of (firm) size with a given attribute, i.e., size x proximity to airport, size x population, and so on. Hence, by writing $\left(S_i^{t+T} - S_i^t\right)$ as G_i^{t+T} and defining X_k as the *k-th* location attribute included as interaction term with $(S_i^t)^2$ and (S_i^t), Equation 8.7 presents the derived model for estimating firm growth/decline. The parameters a_i and b_i represent a set of parameters $a_0, a_1, a_2, \ldots, a_n$ and $b_0, b_1, b_2, \ldots, b_n$ to be estimated.

$$G_i^{t+T} = \left(a_0 + \sum_{k=1}^{n} a_k X_k\right)(S_i^t)^2 + \left(b_0 + \sum_{k=1}^{n} b_k X_k\right)(S_i^t) + c \tag{8.7}$$

8.3 Empirical modeling setup

The data used concerned the LISA register. From Equation 8.7, one can observe that G_i^{t+T} can be calculated over any yearly sequence of time-intervals. Considering the 11-year database available, G_i^{t+T} can be obtained over 1, 2, …, 10 years. Using the size of firms, the difference in firm size between two time moments was hence computed for the time steps of 1, 2, 3, …, 10 years, resulting in a set of computed G_i^{t+T} observations for each of these period lengths. Thus, 10 separate sub datasets were obtained. That is, sub dataset 1 refers to the computed growth over 1 year; sub dataset 2 refers to the computed growth over 2 years; and so on. The reason for doing this is to analyze different sets of period lengths in which growth can be computed. In turn, this allows the evaluation of the modeling approach by searching for a distributional data structure that produces the best model fit.

It is important to observe that although data was acquired between 1996 and 2006 from the original database, this 'calendar' information was not taken into account. Instead, the firms were rearranged in terms of the number of years that they existed. For example, the first year that a firm is observed in the database is coded as year 1, independently on which calendar year this refers. By doing this, all observations were brought to a same reference, based on the years of existence (i.e., year 1, year 2, year 3, …, year 11), instead of the actual period that they existed (year 1996, year 1997, …, year 2006).

Depending on the number of years that a firm existed and on the observed period that growth was computed, it was possible to extend the number of observations. For example, considering that a firm existed for 8 years and growth is computed over 3 years, this firm will be represented by 5 observations, i.e., year 4-year 1; year 5-year 2; year 6-year 3; year 7-year 4; year 8-year 5. That is, one for each pair of period in which the firm existed and in which growth was computed. Likewise, if growth is computed for this same firm, but now over 6 years, this firm will be represented by only 2 observations, i.e., year 7-year 1; year 8-year 2.

After these preparations, there were 10 sub datasets relating to 10 different time step sizes over which growth was computed. Consequently, each sub dataset contains a different number of observations. This is due to the possible combinations among the number of firm observations, their period of existence, and the related time step in which growth was computed. Table 8.1 presents the resulting number of firm observations in these created sub datasets, under the "extended dataset" column label. However, one can observe that using this set of data would require a panel data model, as there is information of each firm over multiple periods. But a simpler structure is preferred in the sense of not taking interdependencies between observations into account in the present analysis. Hence, a sample of observations was extracted from each of the 10 sub datasets, comprising a random selection of one observation per firm. As a result of this, the data has no longer a panel structure. The observations included in the analysis are now independent of each other. Table 8.1 presents the resulting number of firms in these samples, under the "one-firm dataset" column label.

Table 8.1. Number of firm observations in the created sub datasets for analyzing firm growth

T	Number of observations	
	Extended dataset	One-firm dataset
1	1,092,598	256,348
2	836,250	188,850
3	647,400	149,950
4	497,450	122,285
5	375,165	99,668
6	275,497	81,999
7	193,498	68,309
8	125,189	53,407
9	71,782	39,072
10	32,710	32,710

Another issue considered when organizing such datasets regards the indication of whether a firm relocated during its existence. Growth and relocation can be interdependent events, but an examination of the data indicated that a large percentage (about 82 %) of office firms did not relocate. Given this quite large majority of non-relocating firms, only these were selected to carry out this analysis.

Finally, the attributes of this database included: office firm type (economic sector), already aggregated into the 15 types of office firms, the 6-digit postcode, which was used to link the socioeconomic and spatial characteristics to each firm, and the firm size, which was combined with the set of location attributes as interaction terms (for both quadratic and linear terms). The list of socioeconomic and spatial attributes used is presented hereafter.

1. Density of public transportation facilities;
2. Distance (km) to the closest airport, measured through the roadway network;
3. Distance (km) to Schiphol (international level) airport, measured through the roadway network;
4. Distance (km) to the closest international train station, measured through the roadway network;
5. Distance (km) to the closest intercity train (IC) station, measured through

the roadway network;

6. Distance (km) to the closest shopping center, measured through the roadway network;
7. Distance (km) to the closest highway junction, measured through the roadway network;
8. Urbanization level at the municipality level;
9. Urbanization level at the 4-digit post code area;
10. Regional effects, represented by the Dutch provinces;
11. Population measured at both municipality and 4-digit post code area levels;
12. Number of households measured at the 4-digit post code area;
13. Labor force measured at the 4-digit post code area;
14. Average income (Euro) measured at the 4-digit post code area;
15. Number of places in schools measured at the 4-digit post code area;
16. Number of parking places measured at the 4-digit post code area;
17. Effects of agglomeration economies, at the COROP area;
18. Density of shopping centers within 1 km;
19. Office rent price in Euro/m².

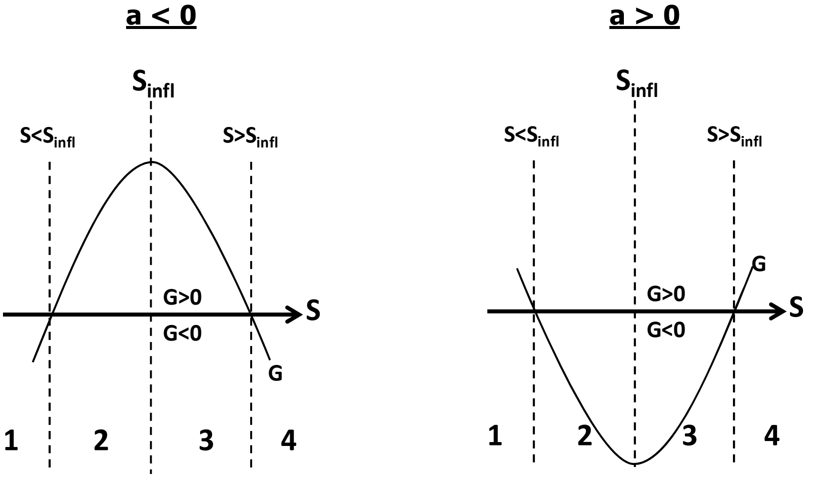
NLOGIT version 4.0 (Greene, 2007) was used to estimate an ordinary least squares linear regression model for the proposed approach. The next section presents the findings.

8.4 Analyses and results

8.4.1 Evaluation of the shape of the growth curve

A first question considered in the approach regards to what extent growth processes of firms can be indeed represented by a logistic form. In terms of the regression model approximation used, this comes down to the question whether estimations of the a and b parameters indeed result in a function that has the form of a parabola opening downward. To test these issues, the 10 sub datasets prepared before were used to estimate 10 separate linear regression models. This resulted in 10 sets of estimated parameters for the related independent variables described before. Using these parameters and keeping everything else constant, it can be noted that 10 separate forms of Equation 8.7 can be written, describing growth over some time step (T) as a function of firm size.

Using mathematical properties of a quadratic function, the terms *a*, *b*, and *c* can be determined along with the coordinates of the vertex of the curve. The term *a* indicates whether the curve opens upward ($a > 0$) or downward ($a < 0$). The x-coordinate of the vertex indicates the inflection point (e.g., S_{infl}), which refers to the size where growth is at a maximum (for $a < 0$) or minimum (for $a > 0$). By plotting these values in a graph along with the predicted values of growth (G_i^{t+T}) obtained from any given value of size (S_i^t), some possible classifications can be drawn as represented by the zones 1, 2, 3, and 4 in the graphs shown in Figure 8.1. As the main interest is for the findings where $a < 0$, these will be discussed first.



1) $G<0$; $S<S_{infl}$ 3) $G>0$; $S>S_{infl}$

2) $G>0$; $S<S_{infl}$ 4) $G<0$; $S>S_{infl}$

1) $G>0$; $S<S_{infl}$ 3) $G<0$; $S>S_{infl}$

2) $G<0$; $S<S_{infl}$ 4) $G>0$; $S>S_{infl}$

Figure 8.1. Classification scheme according to the properties of a quadratic function

Given the above, zone 1 represents firms whose predicted growth is negative and the current size is lower than S_{infl}. Zone 2 represents firms whose predicted growth is positive and the current size is lower than S_{infl}. For this part of the curve, growth is assumed to be at an accelerating phase, where growth increases for increasing values of size. Zone 1 is not consistent with the theory of logistic growth, as it represents that firms decline (negative growth) for an accelerating growth phase. Then, zone 3 represents firms whose predicted growth is positive and the current size is larger than S_{infl}. Likewise, zone 4 represents firms whose current size is larger than S_{infl}, but the predicted growth is negative. It represents declining firms. For this part of the curve, growth is assumed to be at a decelerating phase, where growth decreases for increasing values of size. In this case, in fact, zones 3 and 4 are in line with the firm lifecycle assumed in the theory of logistic growth, as firms would start to slow down their growth process, ending up in a declining phase.

Applying this classification scheme systematically to every firm observation in each of the 10 sub datasets, it is possible to obtain the percentage of firms in each zone. The results are reported in Table 8.2. It shows that the highest percentage (70.1 %) of overall firms following the expected curve (i.e., $a < 0$) refers to $T = 4$. Hence, this is the model taken for the analyses depicted in the sequel. However, it is important to emphasize that the choice of $T = 4$ is rather arbitrary. It should not mean that the temporal resolution of 4 years would work best in any case, as one may be interested in predicting growth for smaller or larger values of T, depending on the available data.

Table 8.2. Percentage of office firms in each category of the classification scheme

T	a < 0					a > 0				
	1	2	3	4	Total	1	2	3	4	Total
1	0.0	30.4	13.6	3.7	47.7	26.0	12.1	0.0	14.1	52.3
2	0.0	36.5	13.0	1.6	51.1	32.2	8.8	0.1	7.8	48.9
3	0.0	38.8	13.1	1.4	53.3	34.1	9.6	0.1	3.0	46.7
4	0.0	59.5	8.3	2.2	70.1	14.3	6.7	0.1	8.8	29.9
5	0.0	41.3	18.9	2.8	63.0	23.5	9.2	0.1	4.2	37.0
6	0.0	42.4	11.4	2.1	55.9	25.5	9.2	0.1	9.3	44.1
7	0.0	37.9	13.3	1.9	53.1	32.5	8.8	0.1	5.5	46.9
8	0.0	40.5	11.8	2.4	54.6	25.7	7.9	0.1	11.6	45.4
9	0.0	43.5	11.3	2.2	57.0	22.2	8.4	0.1	12.3	43.0
10	0.0	53.3	8.1	2.2	63.6	14.8	9.3	0.2	12.1	36.4

According to the results of this specific model, about 59.5 % of firms are in zone 2, which represents firms with predicted positive growth and are at an accelerating growth phase. In turn, in zone 3, there are about 8.3 % of firms, regarding firms with predicted positive growth, but at a decelerating growth phase, as the current size is larger than S_{infl}. Firms with predicted negative growth and at a decelerating growth phase, i.e., firms that are declining, correspond to about 2.2 % (referred to zone 4). As for the findings related to zone 1, these are in line with the expectations, i.e., there are no predicted firms in this zone.

Although the obtained model was able to accommodate a large percentage of firms according to the proposed assumptions, it can be noticed that a rather significant percentage of firms (i.e., about 29.9 %) is not in line with the expectations, as they follow an upward curve. However, it is still possible to reason in the same fashion as discussed above. Literally, in case $a > 0$, zone 1 represents firms whose predicted growth is positive and the current size is lower than S_{infl}. Zone 2 represents firms whose predicted growth is negative (decline) and, likewise, the current size is lower than S_{infl}. For this part of the curve, it is assumed that growth is at a decelerating phase. If only this descending part of the curve were taken into account and disregarding the fact that the curve will start an ascending part after S_{infl}, one can observe that this descending part of the curve represents that about 21 % of firms are at a decelerating growth phase. It is composed by 14.3 % of firms with predicted positive growth, i.e., firms whose growth is slowing down, and 6.7 % of firm with predicted negative growth, i.e., declining firms.

Looking now at the second part of the curve, where it comprises an ascending part (i.e., current size is larger than S_{infl}), this part regards to an accelerating growth phase. As zone 3 represents firms whose predicted growth is negative, this zone does not seem to make sense, as it would suggest that firms are declining (negative predicted growth), but in an accelerated growth phase. As for the share of firms found in this zone, it regards to an insignificant figure (0.1 %), which is in line with the expectations. On the other hand, zone 4 could still be considered consistent. It represents firms whose predicted growth is positive and in an accelerating growth phase. The share of firms in this zone is about 8.8 %.

8.4.2 Model parameters

The results of the growth model related to $T = 4$ based on a linear regression estimation are presented hereafter. Table 8.3 presents the coefficients for the size variables, for both quadratic and linear terms, along with the constant, i.e., the terms a_0, b_0, and c, where "z" regards the statistical z-test used for estimating the parameters. Table 8.4 presents the coefficients of the interactions terms between the size variables and the location attributes.

Table 8.3. Main effects of the linear regression model for modeling growth

Size variable	Coefficient	P[|Z|>z]
a_0	1.74E-04	0.000
b_0	-2.41E-01	0.000
Constant *c*	3.04E-01	0.000

The interpretation of the coefficients is not as straightforward as one might expect. Traditionally, a positive coefficient for a given variable means that growth increases as the value of such a variable increases. Conversely, if the coefficient is negative, growth decreases as the value of such a variable increases.

Given that the model is based on a quadratic and on a linear term, both built with additional interaction terms between size and location attributes, such a reasoning becomes more complicated. Therefore, to examine the individual effect of a given variable on growth, both quadratic and linear terms should be combined. It gives four possible classifications described hereafter, which are represented in a graph shown in Figure 8.2.

1. $a_k > 0$ and $b_k > 0$: growth is an *increasing* function of size and an *increase* (decrease) in the values for the variable leads to *accelerating* (decelerating) growth;
2. $a_k < 0$ and $b_k < 0$: growth is a *decreasing* function of size and an *increase* (decrease) in the values for the variable leads to *decelerating* (accelerating) growth;
3. $a_k > 0$ and $b_k < 0$: on the one hand, growth is a *decreasing* function of size for firms whose size is *smaller* than S_{infl} and *increasing* (decreasing) values for the variable leads to *decelerating* (accelerating) growth; on the other hand, growth is an *increasing* function of size for firms whose size is *larger* than S_{infl} and *increasing* (decreasing) values for the variable leads to *accelerating* (decelerating) growth;
4. $a_k < 0$ and $b_k > 0$: on the one hand, growth is an *increasing* function of size for firms whose size is *smaller* than S_{infl} and *increasing* (decreasing) values for the variable leads to *accelerating* (decelerating) growth; on the other hand, growth is a *decreasing* function of size for firms whose size is *larger* than S_{infl} and *increasing* (decreasing) values for the variable leads to *decelerating* (accelerating) growth.

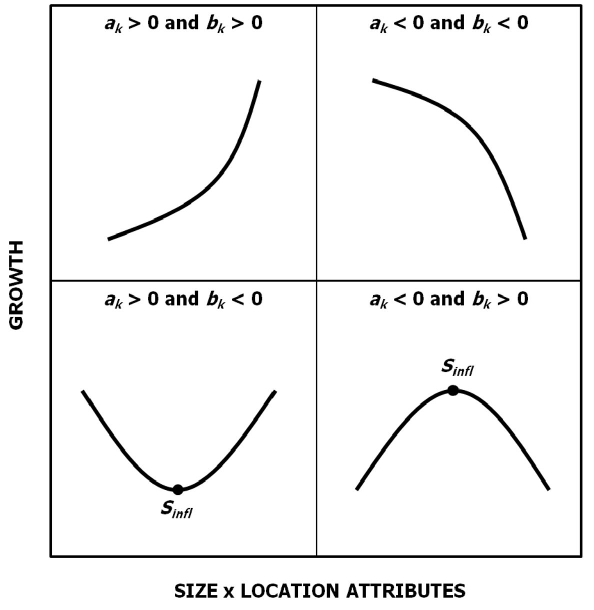


Figure 8.2. Representation of the classification scheme for the estimated parameters

To ease the discussion of the findings of Table 8.4 according to the classification scheme introduced, the terms "small (office) firms" and "large (office) firms" will be used to refer to as, respectively, firms whose size is smaller than the inflection point of the parabola (i.e., S_{infl}), and firms whose size is larger than the inflection point of the parabola. As the results for office firm types show, small firms related to agriculture; industry; basic infrastructures; retail and horeca; traffic and communication; financial institutions; social security; business service; research and development; and education and health have a slower growth compared to companies related to environmental services, culture and recreation. Conversely, large firms of those types experience a faster growth compared to the base type. Real estate companies follow an opposite pattern. Small firms grow faster than companies related to environmental services, culture and recreation; while large firms grow slower than environmental services, culture and recreation companies. Firms related to computer and information technology, and public administration have an accelerated growth for increasing values of size, compared to companies related to environmental services, culture and recreation. On the other hand, the building industry experience a decelerating growth for increasing values of size compared to environmental services, culture and recreation firms.

Table 8.4. Interaction effects of the growth model

Interaction variables with size	a_k	P[|Z|>z]	b_k	P[|Z|>z]
Firm type 1 – Agriculture	4.35E-03	0.123	-5.23E-01	0.087
Firm type 2 – Industry	3.36E-04	0.000	-9.14E-02	0.000
Firm type 3 – Basic infrastructures (gas, electricity and water)	6.61E-04	0.000	-2.88E-01	0.000
Firm type 4 – Building industry	-9.39E-05	0.036	-2.54E-02	0.575
Firm type 5 – Retail and horeca	5.52E-04	0.000	-1.58E-01	0.000
Firm type 6 – Traffic and communication	5.60E-04	0.000	-4.79E-01	0.000
Firm type 7 – Financial institutions	4.89E-04	0.000	-1.81E-01	0.000
Firm type 8 – Social security	3.47E-04	0.000	-3.52E-02	0.029
Firm type 9 – Real estate	-8.86E-04	0.000	2.28E-02	0.354
Firm type 10 – Business service	3.62E-04	0.000	-8.37E-02	0.000
Firm type 11 – Computer and information technology	2.21E-04	0.000	1.62E-01	0.000
Firm type 12 – Research and development	4.95E-04	0.000	-1.87E-01	0.000
Firm type 13 – Public administration	2.71E-04	0.000	1.70E-02	0.218
Firm type 14 – Education and health	4.96E-03	0.000	-7.90E-01	0.000
Firm type 15 – Environmental services, culture and recreation	Base type (dummy code)			
Province #1 – Drenthe	-5.34E-04	0.000	2.07E-01	0.000
Province #2 – Flevoland	2.01E-04	0.000	-2.82E-02	0.209
Province #3 – Friesland	-7.39E-04	0.000	3.91E-01	0.000
Province #4 – Gelderland	3.75E-04	0.000	-1.00E-01	0.000
Province #5 – Groningen	-5.02E-04	0.000	6.11E-02	0.045
Province #6 – Limburg	-2.02E-04	0.000	5.36E-02	0.036
Province #7 – North Brabant	2.05E-04	0.000	-3.39E-02	0.009
Province #8 – North Holland	-3.32E-04	0.000	1.99E-01	0.000
Province #9 – Overijssel	-4.08E-04	0.000	2.87E-01	0.000
Province #10 – Utrecht	1.81E-05	0.088	1.08E-01	0.000
Province #11 – Zeeland	-8.44E-04	0.000	3.00E-01	0.000

Table 8.4. Interaction effects of the growth model

Interaction variables with size	a_k	P[|Z|>z]	b_k	P[|Z|>z]
Province #12 – South Holland		Base type (dummy code)		
Urbanization level 1 (municipality)	5.54E-05	0.164	-8.99E-02	0.000
Urbanization level 2 (municipality)	-1.53E-04	0.000	-6.97E-02	0.000
Urbanization level 3 (municipality)	-1.30E-04	0.001	-4.63E-02	0.009
Urbanization level 4 (municipality)	-2.87E-04	0.000	6.74E-02	0.000
Urbanization level 5 (municipality)		Base type (dummy code)		
Urbanization level 1 (postcode area)	-4.26E-04	0.000	2.28E-01	0.000
Urbanization level 2 (postcode area)	-1.48E-04	0.000	8.48E-02	0.000
Urbanization level 3 (postcode area)	-2.45E-04	0.000	7.46E-02	0.000
Urbanization level 4 (postcode area)	-1.54E-04	0.000	1.13E-02	0.138
Urbanization level 5 (postcode area)		Base type (dummy code)		
Number of bus, tram and metro stops	-3.03E-06	0.000	-5.53E-04	0.004
Distance to Schiphol international airport	-7.37E-06	0.000	2.39E-03	0.000
Distance to the closest airport	1.74E-06	0.000	-1.71E-04	0.331
Distance to IC train station	-3.20E-06	0.000	-1.97E-03	0.000
Distance to international train station	9.28E-06	0.000	-2.48E-03	0.000
Distance to highway junction	-4.48E-05	0.000	1.80E-02	0.000
Distance to shopping centers	-2.52E-06	0.050	-7.96E-03	0.000
Rent price / m^2	1.06E-06	0.000	2.69E-04	0.001
Effects of agglomeration economies	3.48E-08	0.000	-4.37E-05	0.000
Population (4-digit postcode)	5.31E-08	0.000	-1.33E-05	0.000
Number of households (4-digit postcode)	-1.99E-09	0.702	2.94E-05	0.000
Labor force (4-digit postcode)	-1.08E-07	0.000	-1.19E-05	0.052
Average income (4-digit postcode)	-5.80E-09	0.000	4.74E-06	0.000
Places at schools (4-digit postcode)	-2.56E-09	0.000	-3.79E-07	0.327
Parking places (4-digit postcode)	2.78E-08	0.000	5.55E-06	0.000
Density of shopping centers within 1 km	9.97E-06	0.000	-1.12E-02	0.000

Regarding the effects of the Dutch provinces, it can be observed that, compared to South Holland, the provinces of Drenthe, Friesland, Groningen, Limburg, North Holland, Overijssel, and Zeeland have small firms growing faster while larger firms grow slower. Firms in Flevoland, Gelderland, and North Brabant grow slower if they are rather small, while they grow faster if they are large. Lastly, firms in the Utrecht province have an accelerated growth for increasing values of size.

The effects of urbanization levels indicate that very highly urbanized municipalities have small firms growing slower if compared to very lowly urbanized municipalities. However, large firms grow much faster if compared to very lowly urbanized municipalities. Next, lower levels of urbanization (i.e., highly and moderately) experience a decelerated growth for increasing values of size, compared to very lowly urbanized municipalities. Finally, small firms grow faster in lowly urbanized municipalities while large firms grow slower, both compared to very lowly urbanized municipalities. A similar reasoning can be described for the effects of urbanization at the postcode area level, but the general conclusion is that small firms grow faster while large firms grow slower, when compared to very lowly urbanized postcode area.

The influence of transportation facilities on firm growth can be also observed in the related coefficients. Firstly, the accessibility to public transportation seems to negatively influence growth. The increase in the number of bus, tram and metro stops slows down firm growth. It may be related to urban planning policies, where locations with good public transportation infrastructures is probably reaching a growth capacity compared to, for example, more remote locations where there is still a growth potential. On the other hand, the proximity to Schiphol international airport influences firm growth in such a way that, when closer to the airport, small firms grow slower while large firms grow faster. The proximity to highway junctions indicates the same pattern as for Schiphol airport: small firms grow slower while large firms grow faster when close to highway junctions. Examining the proximity to the closest airport, however, the pattern differs. Small firms have a faster growth and large firms have a slower growth, when close to these airports. This pattern is also observed for the proximity to international train stations. Nevertheless, as expected, when analyzing the proximity to intercity train stations, firms tend to grow faster if they are close to such transportation facility. Along with that, the same inference can be formulated with regard to the proximity to shopping centers.

Considering rent price, the interpretation is rather straightforward. For increasing price values, firms grow faster. In other words, growing firms would probably search for locations that can offer better quality, which reflect higher rent prices. The same outcome can be observed regarding parking places. Firms growing faster need more parking places. Analyzing the effects of agglomeration economies, it can be noticed that small firms grow slower and large firms grow faster, i.e., large firms seems to benefit more from these effects compared to small firms. This pattern is also observed for population and for the density of shopping centers. Increasing population reflects in slower growth for small firms, whereas large firms have a faster growth. Likewise, for an increasing number of shopping centers, small firms have slower growth and large firms have faster growth. The number of households and the average income in the postcode area present similar findings. When the values of these variables increase, small firms grow faster while large firms grow slower. Lastly, labor force and places at schools reflect a decreasing function for growth, which becomes slower when the values of these variables increase.

8.5 Conclusions and discussion

The findings discussed in this chapter allow the following conclusions to be drawn. The literature reports empirically that firm growth is driven by firm aspects, such as: size, age, lifecycle, and type of economic activity, for example. However, from an urban planning perspective, and more specifically, from the perspective of LUTI models development, the expectation is that location attributes do have an influence on firm growth. Hence, the study described in this chapter tried to investigate this process, bringing an alternative way to model the firm growth process within these LUTI frameworks.

The model proposed here comprised an approximation based on the well known logistic growth function. This logistic growth function has an appeal for modeling population growth due to its main characteristic of considering a limiting upper bound for evolving larger population sizes, which is often related to the concept of carrying capacity. However, the logistic growth function considers that growth stops when the population size reaches its maximum. But the firm development is different. Firms also have a declining phase, which cannot be accommodated by the logistic growth model.

Given the above, the approximation derived comprised a quadratic function, where growth is assumed to increase for initial population sizes, reaching a maximum size where growth would start to decrease for larger population sizes. At final stages, for increasing population sizes, growth would become negative, representing the decline process. As for the inclusion of location attributes, aimed at being investigated in this study, they were considered as interaction terms with firm size. This methodology resulted in a valuable approach, specifically for being a more flexible form capable to: accommodate both firm growth and decline, compare various firm patterns found in the data, and include the location attributes.

The findings suggest that firm growth is highly influenced by location attributes, which is expected from the perspective of LUTI models. Previous empirical applications had found influences of the proximity to highways and train stations and some socioeconomic aspects. The results obtained here are comparatively in line with those, although the set of attributes included in this study is relatively larger. Hence, it brings contributions to this field, advancing the analysis of firm growth from an urban and transportation planning perspective. In sum, a firm growth component is defined through these estimated parameters, which can be used to build a firm demographic approach. Next chapter will present the closure model, which is the process described when firms go out of business, following the same reasoning underlying this research project on examining firm demographic processes in relation to location aspects.

9

Closure model

9.1 Introduction

This chapter presents the closure model. In a lifecycle, firms will go out of business at some point in time, basically because of two reasons: bankruptcy or voluntary exit (Harhoff et al., 1998). The first is obviously associated with the firm's profit becoming lower than expected. The latter can be determined by some conditions, such as: market conditions and competition. If the economy presents good prospects, closure rates tend to be low, but there is also a chance that new firms will enter in the market. It may cause market competition resulting in the closure of some businesses. Effects of agglomeration economies are also another determinant that influences firm closure. Findings suggest that firm closure rates are low where these effects are observed (Berglund and Brännäs, 2001).

In addition to the above, the determinants of firms' closing patterns have also been explored in terms of firm internal characteristics, such as: age and size. As van Wissen (2000) reviews, older firms tend to have lower closure rates compared to younger firms. This regards to a hypothesis formalized as the liability of newness. As for firm size, larger firms tend to have lower closure rates compared to small firms, which is related to the liability of size hypothesis. Empirical investigations that have found these relationships between closure rates and age and size include: Mata and Portugal (1994); Audretsch and Mahmood (1995); Geroski (1995); and Baldwin et al. (2000). The dynamic characteristic of a firm associated with technological changes over time may also lead to a closure event. If a firm is not able to deal with these changes or lacks some resource (e.g., skilled workforce), then it will be subject to close (Maoh and Kanaroglou, 2005). This also varies across economic sectors and is influenced by the production environment, which includes the necessary conditions for the firm's economic production, such as: available infrastructure, access to other firms, and the setup in the local environment (van Wissen, 2000).

The above gives an indication that location characteristics may also influence the firm closure process. Maoh and Kanaroglou (2007b) give an overview on this topic, describing the findings obtained by some authors. They mention, for example, Baldwin et al. (2000), who showed that closure patterns of Canadian firms are different across provinces. These differences are even stressed when interactions with firm age are taken into account. In Greece, Fotopoulos and Louri (2000) showed that young firms located in Athens have lower closure rates when compared to their counterparts located elsewhere. In another review, de Bok (2007) argues that influences of accessibility on firm closure cannot be empirically found in the literature. However, the author included accessibility attributes in his closure model and found out a "very modest influence of infrastructure proximity". His results indicated that firm closures tend to be higher in areas close to highways, although the estimated coefficients were not fairly significant.

In short, the findings reported are not sufficient to conclude to what extent location characteristics influence the process of firm closure. In general, the literature suggests that this process is mainly influenced by firm aspects, such as: age, size, type of economic activity, etc. However, from an urban planning perspective, especially considering the development of LUTI models, further investigations are still required to evaluate the relationships between firm closures and location characteristics. Hence, the contribution aimed at the analyses presented in this chapter is to examine these relationships. To this end, a duration model is applied, whose findings can be eventually used as part of a LUTI system of office firm demography. The chapter is organized as follows: the modeling approach is described, followed by an empirical application. The results and analyses are then presented and some concluding remarks are drawn.

9.2 Modeling approach

The model presented in this chapter is intended to serve as a component of a multi-agent system specified to examine the process of firm closure in relation to a set of characteristics. Because the outcome usually associated with this type of process comprises whether a firm closed or not, a simple binary logistic regression model could be used. However, given the longitudinal data available, a model capable to not only take into account whether a firm is closed, but also consider the duration that a firm existed before going out of business was adopted here. This type of model is usually referred to as duration models, failure time analysis, hazard analysis, transition analysis, survival analysis, or, in general, event history analyses (Allison, 2004).

The Cox proportional hazards model is the most used in survival analyses. In these proportional hazard models, the parameters are interpreted as a function of hazards and the distribution of the outcome remains unknown. This structure has produced reliable results and has been used in many applications (see for example, Kleinbaum and Klein, 2005; Norušis, 2008).

As opposed to the Cox model, there is another type of duration models known as parametric models that are based on a specified distribution of the outcome. Many parametric models are considered to be accelerated failure time (AFT) models, in which survival time is modeled as a function of covariates. The parameters in AFT models are related to survival times, which are assumed to follow a known distribution. The most common distributions used for survival times are: Exponential, Weibull, and Loglogistic.

Table 9.1 presents the survival function, $S(t)$, and hazard function, $h(t)$, for the three distributions mentioned before, i.e., Exponential, Weibull, and Loglogistic. The details underlying such formulations will not be introduced here. For further reading, readers may consult Kiefer (1988); Collett (1994); Hosmer and Lemeshow (1999); Smith (2002); Vittinghoff et al. (2005); and Kleinbaum and Klein (2005). Regarding the specifications of such functions, the parameter λ is usually re-parameterized in terms of covariates and regression parameters, and the parameter p (called the shape parameter) is held fixed.

Table 9.1. Survival and hazard functions for the Exponential, Weibull and Loglogistic distributions

Distribution	$S(t)$	$h(t)$
Exponential	$\exp(-\lambda \,.\, t)$	λ
Weibull	$\exp(-\lambda \,.\, t^{p})$	$\lambda \,.\, p \,.\, t^{p-1}$
Loglogistic	$1 / (1 + \lambda \,.\, t^{p})$	$(\lambda \,.\, p \,.\, t^{p-1}) / (1 + \lambda \,.\, t^{p})$

9.3 Empirical modeling setup

The data used for the present model stems from the LISA register. It contains records at the 6-digit postcode and comprises a time series database, from 1997 to 2006. Although the original dataset included cases observed in 1996, these were removed from this analysis, as the year when these office firms were founded is unknown. Therefore, only office firms with a known start-up year were selected, comprising 286,874 observations.

The dependent variables used in the duration models comprise the duration (in years) that each firm existed between 1997 and 2006, and a binary status variable, which defines whether a firm closed or not. If the firm still existed in 2006, the variable is equal to zero, i.e., right-censored case. On the other hand, the set of independent variables includes firm attributes and location characteristics. In fact, regarding firm attributes, only the office firm economic sector was used. The literature suggests that firm age and firm size are important determinants of firm closure as well. However, given their endogenous nature considering the approach developed here, they are not included. The location characteristics are specified by the following set of variables:

1. Distance (m) to the closest airport, measured through the roadway network;
2. Distance (m) to Schiphol (international level) airport, measured through the roadway network;
3. Distance (m) to the closest international train station, measured through the roadway network;
4. Distance (m) to the closest intercity train (IC) station, measured through the roadway network;
5. Distance (m) to the closest shopping center, measured through the roadway network;
6. Distance (m) to the closest highway junction, measured through the roadway network;
7. Urbanization level at the municipality level;
8. Urbanization level at the 4-digit post code area;
9. Regional effects, represented by the Dutch provinces;
10. Population measured at both municipality and 4-digit post code area levels;
11. Number of households measured at the 4-digit post code area;
12. Labor force measured at the 4-digit post code area;
13. Average income measured at the 4-digit post code area;
14. Number of places in schools measured at the 4-digit post code area;
15. Number of parking places measured at the 4-digit post code area;
16. Effects of agglomeration economies, at the COROP area;
17. Office rent price in Euro/m².

NLOGIT version 4.0 (Greene, 2007) was used to estimate the duration model for the proposed approach. It is important to mention that all continuous independent variables are used with logarithm transformation, which results in better model fit when compared to the outcomes without such transformations (these comparisons are not shown here). The same transformation is done for the variable 'duration' as it is required by the software used. The next section presents the findings.

9.4 Analyses and results

The goodness-of-fit values of the parametric duration models are presented in Table 9.2. It shows that, although the values for the three estimated distributions (Exponential, Weibull, and Loglogistic) are similar, the Loglogistic distribution presents the best fit. It is indicated by the lower value of the Akaike Information Criterion (AIC) statistic, which is equal to 1.919. Hence, this is the adopted distribution to model office firm closure.

Table 9.2. Goodness-of-fit of Exponential, Weibull and Loglogistic distributions

	Exponential	Weibull	Loglogistic
Number of iterations completed	123	128	126
Log likelihood	-283,639.80	-281,094.70	-275,083.00
Number of parameters	118	119	119
Akaike Information Criterion (AIC)	1.978	1.961	1.919

The estimates of the main effects of the Loglogistic distribution are presented in Table 9.3, where "z" represents the statistical z-test used for estimating the parameters. It should be noted that only the variables that are statistically significant at the 5 % level are shown. Even though the parameters of most office firm types are not significant (at this level), they are kept in the estimated model to give an indication of their survival patterns separately. Also, interaction effects between office firm types and some location variables are included. These are shown in Table 9.4, whose statistically significant values at the 5 % level are marked with an "*". Not all combinations of location variables and office firm types could be included in the model. The software used allows the inclusion of a limited number of variables (i.e., about 200), so a representative set of interactions was selected, as Table 9.4 shows.

Table 9.3. Coefficients for the main effects of the Loglogistic duration model

Variable	Coefficient	P[|Z|>z]
Constant	-0.422	0.391
Distance to closest airport	0.158	0.010
Distance to Schiphol international airport	0.111	0.000
Distance to closest international train station	-0.174	0.000
Distance to closest intercity train station	0.263	0.000
Distance to closest highway junction	0.229	0.000
Population in the municipality	-0.035	0.000
Average income in the postcode area	-0.824	0.000
Rent price / m²	0.827	0.000
Firm type 1 – Agriculture	0.637	0.950
Firm type 2 – Industry	-1.467	0.052
Firm type 3 – Basic infrastructures (gas, electricity and water)	-0.593	0.814
Firm type 4 – Building industry	-0.553	0.839
Firm type 5 – Retail and horeca	-0.199	0.747
Firm type 6 – Traffic and communication	-3.692	0.185
Firm type 7 – Financial institutions	0.223	0.688
Firm type 8 – Social security	2.016	0.177
Firm type 9 – Real estate	-0.669	0.243
Firm type 10 – Business service	-0.652	0.190
Firm type 11 – Computer and information technology	-0.783	0.152
Firm type 12 – Research and development	-0.995	0.310
Firm type 13 – Public administration	1.970	0.041
Firm type 14 – Education and health	0.251	0.909
Firm type 15 – Environmental services, culture and recreation	Base	-
Province #1 – Drenthe	0.558	0.000
Province #2 – Flevoland	-0.354	0.000
Province #3 – Friesland	-0.434	0.000
Province #4 – Gelderland	0.082	0.000
Province #5 – Groningen	-0.166	0.000
Province #6 – Limburg	0.676	0.000
Province #7 – North Brabant	0.109	0.000

Table 9.3. Coefficients for the main effects of the Loglogistic duration model

Variable	Coefficient	P[|Z|>z]
Province #8 – North Holland	-0.344	0.000
Province #9 – Overijssel	-0.037	0.052
Province #10 – Utrecht	-0.121	0.000
Province #11 – Zeeland	0.353	0.000
Province #12 – South Holland	Base	-

The Loglogistic distribution is an accelerated failure time (AFT) model. The interpretation of the coefficients follows the logic of examining the influence of a variable on survival times (i.e., the timing that firms will go out of business). Thus, positive coefficients are directly related to a stretching out in survival times, indicating higher survival rates. Conversely, negative coefficients are directly related to a contraction of survival times, indicating lower survival rates.

Table 9.3 shows that the coefficients of the transportation infrastructure variables (i.e., airports, train stations and highways) are all positive, except for the distance to international train stations. The positive signs indicate that the farther a firm is from, for example, an airport or a highway junction, the higher its survival rate. Conversely, the negative coefficient for the distance to international train stations suggests that the closer a firm is from this facility, the higher its survival rate. In addition to these coefficient descriptions, it is important to interpret the mechanisms underlying firm survival patterns and transport accessibility. One may expect that the proximity to these infrastructures would generally support firm survival. However, this is not what most of the results obtained here indicate. The explanation could be that the presence of (some) transportation infrastructures hinders survival in the sense that regions with good levels of transportation accessibility are also very urbanized and, therefore, more dynamic, where stronger competition may be observed. Hence, there may be a higher tendency of firms going out of business in these areas compared to, for example, more remote areas.

Next, the coefficients for demographic and economic aspects can be observed. The influence of population resulted in a negative value, indicating that the survival rates in highly populated cities are lower. Results for the average income in the neighborhood also play a significant role in firm's survival, represented by its negative coefficient. Firms tend to go out of business sooner as the average household income in such areas increases. Regarding rent price, the positive coefficient indicates that the survival rates tend to increase when the rent price rises. Although one might expect the opposite (higher survival rates for lower rent prices), high rent price might be an indicator of high quality of the location. The positive coefficient might indicate that the quality aspect is more important than the cost aspect for firm survival. Nevertheless, the expectation is that it would hold mainly for large companies rather than small businesses.

Table 9.3 also evidences contrasts among office firm types presented. For example, office firms with positive values, such as agriculture, financial institutions, social security, public administration, and education and health companies have higher survival chances when compared to the firm type 15 – environmental services, culture and recreation. Conversely, all other types with a negative coefficient have a higher probability of going out of business compared to office firms related to environmental services, culture and recreation.

Finally, to interpret the regional effects represented by the Dutch provinces, it should be realized that estimated effects are relative to the effect estimated for the South Holland region (base category). Negative coefficients indicate regions with lower survival rates than South Holland, as for example, Flevoland, Friesland, Groningen, North Holland, Overijssel, and Utrecht. In contrast, positive coefficients indicate higher survival rates compared to the South Holland province. This is the case of, for example, Drenthe, Gelderland, Limburg, North Brabant, and Zeeland.

As said, the estimated interaction effects between office firm types and location characteristics are shown in Table 9.4. Note that these effects capture differences with the corresponding main effects of location characteristics, as presented in Table 9.3. Moreover, it should be realized that these estimated interaction effects capture differences from the effect of the base (environmental services, culture and recreation – type 15). For this office firm type, all interaction terms are equal to zero, whereas the interaction effects for all other types are non-zero.

Table 9.4 then shows that the estimated interaction effects related to transportation facilities differ. The main effect of distance to the closest airport is positive (+0.158). Significant interaction effects can be observed for office firm types 2 (industry), 9 (real estate), 12 (research and development), and 13 (public administration). Except office firm type 13, all are positive. Therefore, for these firm types, the effect of distance becomes more positive, suggesting that their survival rate increases with increasing distance to the airport. For firm type 13, however, the effect of distance to the airport becomes even negative, indicating that its survival rate increases with decreasing distance to the airport.

Table 9.4. Coefficients for interaction effects with office firm types of the Loglogistic duration model

Office firm type	Dist. to airport	Dist. to IC train station	Dist. to highway junction	Rent price	Average income
1	-0.484	-0.354	-1.092	-1.632	5.998*
2	0.251*	-0.064	-0.139*	-0.802*	1.126*
3	0.021	0.143	-0.090	-0.367	0.362
4	0.152	-0.111	-0.518*	-0.059	1.447*
5	-0.044	-0.060	-0.258*	-0.557*	1.028*
6	0.198	0.049	-0.162	0.641	1.334*
7	0.081	-0.159*	-0.205*	-0.830*	1.048*
8	0.193	-0.242*	-0.333*	-1.307*	0.851*
9	0.119*	-0.114*	-0.186*	-0.560*	1.111*
10	0.002	-0.129*	-0.222*	-0.645*	1.479*
11	0.070	-0.158*	-0.253*	-0.667*	1.439*
12	0.200*	-0.019	-0.206*	-0.280	0.615*
13	-0.364*	0.177*	-0.159*	0.524*	-0.543*
14	0.377	-0.490*	0.153	-1.388*	1.723*

*(Significant coefficients at the 5 % probability level are marked with an *)*

Regarding the proximity to IC train stations, the main effect is also positive (+0.263). Significant interaction effects can be observed for office firm types 7, 8, 9, 10, 11, 13, and 14. Respectively, these are related to financial institutions; social security; real estate; business service; computer and information technology; public administration; and education and health. Except office firm type 13, all are negative, though. It suggests that the effect of distance becomes less positive for these firm types. In other words, their survival rate increases to a lesser extent with increasing distance to train stations. In fact, for firm type 14, the effect becomes even negative, meaning that its survival rate increases with decreasing distance to IC train station. On the other hand, office firm type 13 has a positive sign, indicating that the effect of distance becomes more positive. Hence, its survival rate increases with increasing distance to the airport.

Following a similar reasoning for the proximity to highway junctions, the main effect is positive (+0.229). However, all significant interaction effects are negative. On the one hand, for some of these office firm types the effect of the distance to highway junctions becomes less positive. It indicates that their survival rate increases to a lesser extent with increasing distance to highway junction. This is the case of office firm types 2 (industry), 6 (traffic and communication), 7 (financial institutions), 9 (real estate), 10 (business service), 12 (research and development), and 13 (public administration). On the other hand, the effect of distance to highway junction becomes even negative for office firm types 4 (building industry), 5 (retail and horeca), 8 (social security), and 11 (computer and information technology). It suggests that the survival rate for these firm types increases with decreasing distance to highway junction.

The estimated interaction effects between office firm types and rent price also demonstrate differences. As the positive (+0.827) main effect indicates, survival rates increase for increasing rent prices. Hence, office firm type 13 (public administration) has this effect intensified, given by its (significant) positive interaction coefficient. On the other hand, most of office firm types have negative (and significant) interaction coefficients. For firm types 2 (industry), 5 (retail and horeca), 9 (real estate), 10 (business service), and 11 (computer and information technology), the effect of rent price becomes less positive, suggesting that their survival rate increases to a lesser extent with increasing rent price values. However, firm types 7 (financial institutions), 8 (social security), and 14 (education and health) have the effect of rent price becoming in fact negative. It suggests that the survival rate for these firm types increases with decreasing values of rent price.

Finally, as for the average income, the main effect is negative (-0.824). Therefore, firms go out of business sooner while income levels increase. According to the estimated interaction effects, office firm type 13 (public administration) has this effect even intensified, as indicated by its (significant) negative interaction coefficient. On the other hand, most significant interaction effects are positive. In fact, except office firm type 12 (research and development), the effect becomes even positive. It suggests that the survival rate for office firm types 1 (agriculture), 2 (industry), 4 (building industry), 5 (retail and horeca), 6 (traffic and communication), 7 (financial institutions), 8 (social security), 9 (real estate), 10 (business service), 11 (computer and information technology), and 14 (education and health) increases with increasing income levels. For firm type 12, the effect only becomes less negative, meaning that its survival rate decreases to a lesser extent with increasing income levels.

9.5 Conclusions and discussion

The findings discussed in this chapter allow the following conclusions to be drawn. The literature reports that the process of firm closure is basically influenced by firm aspects, such as age and size, for example. However, from an urban planning perspective, and more specifically, from the perspective of LUTI models development, it is expected that location characteristics also influence firms' closure patterns. Therefore, the analyses presented in this chapter tried to investigate this relationship.

The modeling approach used here was based on duration models, which not only fits the binary nature of the outcome observed for this type of analysis, but also accounts for the elapsed duration until the event is observed. In this case, the event relates to firm closure. In addition, by taking the effects of covariates into account, the model derives parameters that represent the influences of a set of variables in this process.

The findings suggest that firm closure is highly influenced by location attributes, as showed by most of the variables included. This is also in line with the aim of the analysis proposed in this chapter, suggesting an evident relationship between firm closure and location attributes. In addition, earlier efforts found in the literature that have investigated this relationship are not as detailed as developed here. These earlier studies usually adopt only some regional effects or limited accessibility characteristics to explore firm closure. Therefore, as proposed initially, the analysis carried out here also brought a better understanding about the firm closure process by using an extended set of location attributes.

Despite the above, some of those earlier findings can be compared to what is obtained here. For example, de Bok (2007) showed that firm closures tend to be higher in locations close to highways. This is also suggested by the results obtained here. In another study, Maoh and Kanaroglou (2007b) indicated that effects of agglomeration economies have a positive impact on firm survival, which could also be verified from the analysis carried out here.

In sum, a firm closure component is defined through the estimated parameters as described in this chapter. This component can be used to build a firm demographic approach to simulate the evolution of firm dynamics along with the previous components already discussed. Next chapter will present the relocation model, which is the process described when firms decide to change their location. It follows the same reasoning underlying this research project on examining firm demographic processes in relation to location aspects.

10

Relocation decision model

10.1 Introduction

This chapter presents the relocation decision model. Upon establishing a business, firms may consider relocating during its lifecycle. But they expect to minimize this event as much as possible, since relocating often involves high costs and risks of losing well established market shares, for example. As argued by Bade (1983), firms have a strong preference to remain in the current location and will not move unless this location has particular deficiencies that may be causing dissatisfactions to the firm. There are numerous reasons that can motivate a relocation decision, which are often associated with the quality of the location in order to maintain the firm's best level of service. The investigation of the firm relocation process is traditionally based on a twofold scheme that, on the one hand, analyzes the firm's reasons to leave the present location, and, on the other, specifies the characteristics that attract the firm to a new location. These are respectively denominated as push and pull factors (van Dijk and Pellenbarg, 2000a; van Wissen, 2000; Pellenbarg et al., 2002; Brouwer et al., 2004).

Push factors are usually related to changes in firm characteristics (e.g., firm size, market orientation, stage of lifecycle, technology) and deficiencies in the present location (e.g., poor accessibility, costs mismatch, insufficient office space, lack of skilled labor force, land use policy implications) that motivate a relocation decision. Pull factors, on the other hand, comprise mostly the location aspects that firms are searching to accommodate their new requirements upon the decision to relocate is made. Van Wissen (2000) observes that pull factors comprise mostly the same aspects of push factors, but with an opposite content. Hence, these aspects usually include: locations with better accessibility, an even match between location cost and location quality, adequate office space, satisfactory market demand and labor force supply, adequate technology resources, better land use incentives, among others.

In line with the above, although the relocation process is intimately related to the (spatial) characteristics of the new location to which firms will move, Pellenbarg et al. (2002) argue that relocation is relatively different from the initial location choice decision (i.e., by the time of a firm's first establishment), because there is not only the substitution of a location for another. In other words, there is not only the problem of finding a new location. The process involves the firm's decision to relocate, which is conditional on the firm's history and its characteristics. Therefore, the investigation of the overall relocation process can be best viewed as a sequential procedure based on two steps: first, the analysis of the firm's propensity to relocate, and second, the analysis of the choice of a new location.

As for the firm's propensity to relocate, behavioral foundations have been addressed as reviewed in Maoh and Kanaroglou (2007a). Taking an approach where the individual firm is the decision maker, the relocation behavior is highly influenced by firm's internal characteristics. Brouwer et al. (2004) tested some hypotheses about this relationship between firm characteristics and the relocation decision event. Their findings suggest that relocation chances decrease with increasing firm size and firm age; growing or declining firms are also more willing to relocate; and firms that serve large markets are more likely to relocate.

In addition to firm's internal characteristics influencing the relocation decision behavior, Maoh and Kanaroglou (2007a) also reviewed that the firm's propensity to relocate should be examined by an approach that takes into account the aspects of the location as well. Especially from an urban planning perspective, the expectation is that these aspects influence the firm's relocation decision. Along with this, in terms of empirical methods developed, discrete choice models have been usually used. Examples include: binary logit models (e.g., van Wissen, 2000; Brouwer et al., 2004; de Bok and Bliemer, 2006a and 2006b; de Bok, 2007), an ordered logit model applied by van Dijk and Pellenbarg (2000a) using stated preference data, and a nested logit structure as adopted by Maoh and Kanaroglou (2007a). More recently, duration models were also tested as presented by Elgar et al. (2008). The special interest in using this method relies on the fact that the propensity to relocate is not modeled considering only whether or not a firm decides to relocate, but also takes into account the elapsed time before this decision.

The general finding obtained from these empirical investigations points mainly to firms' internal characteristics, such as age, size, economic sector, and lifecycle. As for location aspects, the literature shows fairly few studies on the relationships of these aspects with the firm's decision to relocate. For example, Maoh and Kanaroglou (2007a) reported that core areas of the city are less attractive, which motivate firms to relocate out of these areas. Local competition also seems to motivate firm relocation behavior. On the other hand, effects of agglomeration economies have an influence on hindering firm relocation, which could also be verified in the study carried out by de Bok (2007). In addition, this author included some accessibility measures, based on two types of location profiles: locations close to highway on-ramps only and locations close to highway on-ramps and train stations. His findings suggest that firms close to highway on-ramps only are discouraged to relocate, whereas the proximity to highway on-ramps and train stations seems to motivate a relocation behavior. De Bok (2007) attributes this (unexpected) outcome to the fact that these latter locations may present higher urban density, parking problems, or higher levels of rent price. Nevertheless, he argues that these specific findings are of less importance due to their small influence in the model compared to the outcomes related to firm and agglomeration attributes. In another study, Elgar et al. (2008) found out that effects of agglomeration, competition and accessibility are not significant and they were excluded from their final model. However, the authors reported some influence of labor force around the firm on relocation trends, suggesting that it may still be important to consider location aspects on the firm's relocation decision.

Given the above, there seems to be, in general, fairly inconclusive responses about the influence of location aspects on firms' decisions to relocate. In addition, there is a need for further examinations using more detailed information about location aspects, especially from the perspective of LUTI models development. Therefore, the aim of this chapter is to contribute to the investigation of factors underlying the firm relocation event, especially with regard to the firm's propensity to relocate in terms of external characteristics. To this end, the set of location attributes defined for this research project is applied. Considering the possible alternatives in terms of modeling approaches that could be used, duration models are adopted, specifically because of their capability to account for the elapsed time before the occurrence of the relocation event (Elgar et al., 2008), and since the data available also allows to do so.

Regarding the second step of the relocation process, the analysis of the choice of the new location is often considered as a location choice problem. The investigation of this specific step can be thus referred back to location decision theories and modeling approaches as discussed in chapters 2 and 7. In the analysis carried out here, the location choice upon a relocation decision is examined following the same approach developed for the location choice model introduced in chapter 7, based on the Bayesian classifier networks model. However, the findings will not be discussed in much detail as presented in chapter 7. Instead, the analysis and discussions will be summarized in terms of the differences compared to the earlier location model. This chapter is structured as follows: as the theoretical background involving the techniques adopted here was already introduced in previous chapters, not much detail about them is required. Therefore, the subsequent sections will deal with the application of the selected models to the case investigated here, underlining the empirical modeling setup, the analyses of the results, and the concluding remarks at the end.

10.2 Empirical modeling setup

The data used for this relocation component concerns the LISA register, comprising the period between 1996 and 2006. A relocation was recorded if a change was observed at the 6-digit postcode level. Of the 425,241 observations, 77,664 (about 18 %) office firms relocated in this time frame and some office firms relocated more than once. Table 10.1 shows the frequency distribution of the number of relocations.

Table 10.1. Number of relocating office firms according to the frequency of relocation

Frequency of relocation	Number of office firms
1	64,210
2	11,251
3	1,841
4	300
5	56
6	6

As Table 10.1 presents, a vast majority of relocating firms in this period relocated only once. But there are cases in which more than one relocation is recorded, although it decreases substantially. In this dataset, a maximum of 6 relocations is observed. Tracking the historic relocation behavior of these firms is a possible type of analysis, but this is not the scope here and, hence, a panel structure model is not considered. The main interest is, in fact, in the relocation behavior from location “A” to location “B” and the factors influencing it. Hence, for the analyses carried out, firms that relocated more than once are treated as independent observations, included in the database as additional cases. It results in an extended dataset containing 93,751 observations.

Although firm observations are present within several time frames, with different starting years and different durations over the period between 1996 and 2006, the 'calendar' information is disregarded. Instead, firms are identified by the number of years that they existed, aligned to the same time reference related to the year that they started (i.e., named as year 1). From this reference, the duration (in years) that a firm is present in a given postcode until the period when this firm relocated is computed.

For modeling the firm's propensity to relocate, observations consisting of both relocating and non-relocating firms are used. Considering that office firms registered in 1996 did not have a known starting year, they are excluded from this analysis. Hence, the resulting dataset comprises 297,102 observations. The dependent variables required to run the duration model regard the duration (in years) that each firm was present in the present location until the relocation event (or the end of the observation period, whichever comes first), and a binary status variable defining whether a firm relocated or not. If the firm relocated, the value of this variable is equal to 1.

Regarding the independent variables, both firm related attributes and location characteristics are considered. The firm attributes include: whether or not the office firm was involved in a relocation before, office firm size at the time of relocation, and the office firm type in terms of economic sector. The location characteristics are described by:

1. Density of public transportation facilities;
2. Distance (km) to the closest airport, measured through the roadway network;
3. Distance (km) to Schiphol (international level) airport, measured through the roadway network;
4. Distance (km) to the closest international train station, measured through the roadway network;
5. Distance (km) to the closest intercity train (IC) station, measured through the roadway network;
6. Distance (km) to the closest shopping center, measured through the roadway network;
7. Distance (km) to the closest highway junction, measured through the roadway network;
8. Urbanization level at the municipality level;

9. Urbanization level at the 4-digit post code area;
10. Regional effects, represented by the Dutch provinces;
11. Population measured at both municipality and 4-digit post code area levels;
12. Number of households measured at the 4-digit post code area;
13. Labor force measured at the 4-digit post code area;
14. Average income measured at the 4-digit post code area;
15. Number of places in schools measured at the 4-digit post code area;
16. Number of parking places measured at the 4-digit post code area;
17. Number of shopping centers within 1 km
18. Effects of agglomeration economies, at both COROP area and 1 km radius area scales;
19. Office rent price in Euro/m².

NLOGIT version 4.0 (Greene, 2007) is used to estimate the duration model for the proposed approach. It is important to mention that all continuous independent variables are used with logarithm transformation, which resulted in better model fit when compared to the outcomes without such transformations (these comparisons are not shown here). The same transformation is done for the variable 'duration' as it is required by the software used. The next section presents the findings.

10.3 Analyses and results

The goodness-of-fit statistics for the three distributions (Exponential, Weibull, and Loglogistic) tested are shown in Table 10.2. Although the results are similar, the Weibull distribution presents the best fit. This is indicated by the lower value of the Akaike Information Criterion (AIC), which is equal to 0.979. Hence, this is the distribution used to model the firm's propensity to relocate.

Table 10.2. Goodness-of-fit of the Exponential, Weibull and Loglogistic distributions

	Exponential	**Weibull**	**Loglogistic**
Number of iterations completed	99	101	100
Log likelihood	-150,183.80	-145,383.70	-146,576.00
Number of parameters	90	91	91
Akaike Information Criterion (AIC)	1.012	0.979	0.987

The estimated main effects of the duration model based on the Weibull distribution are presented in Table 10.3, where “z” represents the statistical z-test used for estimating the parameters. Table 10.4 reports the estimated interaction effects between office firm types and some location variables. For the analysis carried out here, only the variables that are statistically significant at the 5 % level are kept in the final model. However, as the results in Table 10.3 show, the parameters of most office firm types are not significant (at this level), but they are still included in the model to give an indication of the specific relocation behavior across economic sectors. The same holds for provinces and urbanization levels that have one category each in which an insignificant parameter was found. As for the interaction effects, not all combinations of location variables and office firm types are included, due to the same limitations in the software used, as described in chapter 9. Hence, a representative set of interactions is selected, as presented in Table 10.4.

Since the Weibull distribution leads to an accelerated failure time (AFT) model, the interpretation of the coefficients can be carried out in a similar fashion as for the closure model discussed in chapter 9. The difference is that instead of analyzing “closure”, the effects of the various variables are related to “relocation”, or more specifically, the propensity to relocate. Therefore, positive coefficients indicate a lower propensity to relocate, and negative coefficients indicate a higher propensity to relocate.

Table 10.3 firstly presents the coefficients related to firm characteristics, i.e., whether a firm was involved in a relocation before and the office firm size. Both coefficients are positive, suggesting that firms that relocated before tend to avoid relocating again. As for firm size, increasing values of office firm size indicate lower relocation chances as well. Next, the propensity to relocate across specific office firm types can be observed. Firm types with negative coefficients tend to relocate more often compared to the base firm type (environmental services, culture and recreation). It comprises firms related to agriculture, basic infrastructures, retail and horeca, traffic and communication, real estate, research and development, and public administration (respectively, types 1, 3, 5, 6, 9, 12, and 13), although not all effects are significant. On the other hand, firm types with positive coefficients are less inclined to relocate compared to the base firm type. This is the case of firms related to industry, building industry, financial institutions, social security, business services, computer and information technology, and education and health (respectively, types 2, 4, 7, 8, 10, 11, and 14), but not all effects are significant as well.

Table 10.3. Estimated main effects of a Weibull duration model

Variable	Coefficient	P[|Z|>z]
Constant	1.798	0.000
Previous relocation	2.751	0.000
Office firm size	0.100	0.000
Firm type 1 – Agriculture	-0.573	0.471
Firm type 2 – Industry	0.135	0.326
Firm type 3 – Basic infrastructures (gas, electricity and water)	-1.268	0.051
Firm type 4 – Building industry	0.775	0.002
Firm type 5 – Retail and horeca	-0.044	0.713
Firm type 6 – Traffic and communication	-0.190	0.462
Firm type 7 – Financial institutions	0.089	0.424
Firm type 8 – Social security	0.385	0.237
Firm type 9 – Real estate	-0.044	0.704
Firm type 10 – Business service	0.098	0.320
Firm type 11 – Computer and information technology	0.424	0.000
Firm type 12 – Research and development	-0.020	0.922
Firm type 13 – Public administration	-0.650	0.011
Firm type 14 – Education and health	0.266	0.235
Firm type 15 – Environmental services, culture and recreation	Base	
Density of public transportation facilities	-0.159	0.016
Distance to Schiphol international airport	0.390	0.000
Distance to international train station	-0.113	0.048
Distance to highway junction	-0.153	0.000
Distance to shopping centers	-0.108	0.000
Province #1 – Drenthe	-0.722	0.000
Province #2 – Flevoland	0.155	0.000
Province #3 – Friesland	0.085	0.018
Province #4 – Gelderland	-0.138	0.000
Province #5 – Groningen	-0.237	0.000
Province #6 – Limburg	-0.455	0.000
Province #7 – North Brabant	-0.099	0.000
Province #8 – North Holland	-0.014	0.504

Table 10.3. Estimated main effects of a Weibull duration model

Variable	Coefficient	P[|Z|>z]
Province #9 – Overijssel	-0.090	0.002
Province #10 – Utrecht	0.037	0.045
Province #11 – Zeeland	-0.308	0.000
Province #12 – South Holland	Base	
Urbanization level 1 (municipality)	0.191	0.000
Urbanization level 2 (municipality)	0.176	0.000
Urbanization level 3 (municipality)	0.122	0.000
Urbanization level 4 (municipality)	0.050	0.031
Urbanization level 5 (municipality)	Base	
Urbanization level 1 (postcode)	-0.113	0.000
Urbanization level 2 (postcode)	-0.073	0.000
Urbanization level 3 (postcode)	-0.067	0.001
Urbanization level 4 (postcode)	-0.013	0.445
Urbanization level 5 (postcode)	Base	
Agglomeration economies (1 km)	0.104	0.000
Population in the postcode areas	-0.240	0.010
Number of households in the postcode area	0.285	0.003
Average income in the 4-digit post code area	-0.631	0.000
Number of places in schools	-0.035	0.000

Turning to the influences of transportation infrastructures, the negative coefficient for the density of public transportation indicates that increasing the number of bus, tram, and metro stops motivates firms to relocate from their current location. As for the proximity to Schiphol airport, the positive coefficient indicates that firms closer to Schiphol tend to have higher relocation rates. In general, these findings can be associated with a more active behavior in relation to firm dynamics, which can be likely observed in areas with such good levels of transportation infrastructures. In this sense, areas with good levels of public transportation are certainly more urbanized, and consequently, more dynamic. So, it is expected that more frequent relocations are observed. The case of Schiphol airport follows the same reasoning. The area is well known for its high concentration of businesses, which may also reflect an intense dynamic of firm location behavior. On the other hand, office firms that are close to international train stations and highway junctions tend to have a lower relocation rate, as indicated by the negative coefficient. This is in line with the findings obtained by de Bok (2007), who found that firms that are close to highway on-ramps are discouraged to relocate.

The effects of the Dutch provinces result in patterns of lower relocation rates in Flevoland, Friesland, and Utrecht, given by the positive coefficient, when compared to the South Holland province. On the other hand, higher relocation rates can be expected in Drenthe, Gelderland, Groningen, Limburg, North Brabant, Overijssel, and Zeeland, compared to South Holland, as indicated by the negative coefficient.

Examining the effects of urbanization levels at the municipality scale, the relocation rates decrease for increasing urbanization levels (compared to non urbanized municipalities). However, when looking at a smaller scale of analysis (e.g., the postcode area), more urbanized postcode areas seem to indicate higher relocation rates, as indicate the negative coefficients. These findings suggest that, on the one hand, the effects of higher urbanization levels on influencing the decision to relocate are stronger at the local level, rather than at the municipality level. On the other hand, firms initially established in lower urbanized municipalities would be encouraged to relocate, probably seeking for more urbanized municipalities.

Regarding the effects of agglomeration economies, although both COROP area and 1 km radius area levels are examined in the analysis, only the latter is significant. The positive coefficient indicates that when the number of office firms around each office firm increases, firms are less inclined to relocate. This confirms the idea that firms tend to benefit from market shares and effects of agglomeration economies, which would hinder the firm's decision to relocate. Although Maoh and Kanaroglou (2007a) used a slightly larger area of analysis, i.e., a circular area of 1.5 km of radius, the results obtained here are also in line with what these authors found.

In terms of socioeconomic aspects and other facilities (shopping centers and schools) considered in the analysis here, if the number of inhabitants in neighborhoods increases, relocation rates become higher, as indicated by the negative coefficient. Alike, increasing the average income and educational places also increase the relocation rates. In general, these findings are also in line with the reasoning described for the urbanization levels at the postcode area. If there are more people, higher income, etc., the area becomes more urbanized. Consequently, it suggests higher relocation rates. On the other hand, as the positive coefficient indicates, firms in neighborhoods with higher number of households tend to relocate less. This could indicate that if a firm is established in a more residential zone, it would hinder its relocation decision. However, this deserves further investigations, because it may be only the case for some smaller office firms of specific types of economic activities. Lastly, firms that are close to shopping centers tend to relocate less, as indicated by the negative coefficient. For some office firms, it may suggest the benefit of market shares and effects of agglomeration, as well as the possibility of making use of collective urban infrastructures available, for example.

Table 10.4. Estimated interaction effects of a Weibull duration model

Firm type	Previous relocation	Density of public transportation	Distance to international train
1	-0.311	0.192	0.484
2	-0.252*	-0.038	0.165*
3	-0.109	0.892*	0.327
4	-0.664*	-0.026	-0.041
5	-0.256*	0.201*	0.169*
6	-0.649*	0.551*	0.368*
7	-0.177*	0.062	0.053
8	-0.110	-0.485*	0.081
9	-0.159*	0.157*	0.094
10	-0.240*	0.135*	0.045
11	-0.446*	0.117	-0.023
12	-0.008	0.299*	-0.116
13	0.373	0.145	-0.188
14	-0.302	0.115	-0.034

*(Significant coefficients at the 5 % probability level are marked with an *)*

Examining the interactions effects between office firm types and location attributes presented in Table 10.4, it turns out that only three variables are significant: whether a firm was involved in relocation before, density of public transportation, and distance to international train. Yet, observing the figures across firm types, it can be noted that there is a relatively low percentage of actual significant values at the 5 % level even in terms of these variables. The main effect related to the variable that describes whether a firm was involved in relocation before is positive (+2.751). Significant interaction effects can be observed for office firm types 2, 4, 5, 6, 7, 9, 10, 11 (respectively, industry; building industry; retail and horeca; traffic and communication; financial institutions; real estate; business service; and computer and information technology), and they are all negative. It suggests that for these office firm types the effect of having relocated before becomes less positive. In other words, their relocation rates tend to be higher compared to office firms related to environmental services, culture and recreation (type 15, the base). Nevertheless, given the difference in the magnitude of the main effect compared to the related interaction effects, the initial interpretation still holds for the above mentioned office firm types. That is, they tend to have lower relocation rates if they were involved in a relocation before.

Looking at the results for the influence of public transportation, the main effect indicates that higher density of bus, tram and metro stops motivates firms to relocate from their current location, as given by the negative coefficient (-0.159). For the related interaction effects of office firm types that are significant, this effect is intensified for social security firms (type 8), which also have a negative coefficient. On the other hand, office firm types 9 and 10 (real estate and business service) hold significant positive coefficients. This indicates that the effect of higher density of public transportation infrastructure becomes less negative for these office firm types, suggesting that their relocation rates increase to a lesser extent with increasing density of bus, tram and metro stops. However, for office firm types related to basic infrastructures, retail and horeca, traffic and communication, and research and development (respectively, types 3, 5, 6, and 12), the effect of higher density of public transportation becomes even positive. It indicates that their relocation rates decrease with increasing density of bus, tram and metro stops.

Finally, regarding the influence of international train stations, the main effect indicates that firms close to these infrastructures have lower relocation rates, as given by the negative coefficient (-0.113). The significant interaction effects indicate, however, that office firm types related to industry, retail and horeca, and traffic and communication (respectively, types 2, 5, and 6) have this effect of the distance to international train stations becoming even positive. It suggests that these office firm types have their relocation rates increased with closer distance to international train stations.

In sum, the findings above have presented the analyses of the relocation decision model, which is the first step in the overall firm relocation process. The second step, which is the choice of the new location, will be discussed in the next section presented hereafter.

10.4 Location choice patterns upon a relocation decision

As said, the relocation event also comprises a step in which firms will look for new locations once the decision to relocate is made. In the case studied here, the analysis of the location choice patterns upon a relocation decision is carried out as a location choice problem. Therefore, similar theories and modeling approaches as introduced in chapter 7 can be applied. More specifically, Bayesian classifier networks are used again to this end. However, since the modeling setup remains similar in relation to the earlier location choice model, the analyses presented in this section will be summarized in terms of the differences and important factors underlying the choice of the new location.

10.4.1 Data

The LISA data is used, but regarding only the observations that are involved in relocation. In fact, it comprises the extended dataset prepared before, containing 93,751 records. The location attributes are then added based on the location that the firm moves to, and discretized as required by the modeling approach. However, differently from what was applied in the earlier location choice model, the relocation choice model comprises some additional attributes and modifications in some of the attributes included before. For example, given a larger number of cases available (i.e., 93.751), the original 15 office firm types are considered instead of a merged number of categories previously adopted. This also applies for the regional effects of the Dutch provinces, in which the 12 provinces are taken instead of aggregating them into a smaller number of categories.

The additional variables regard to: the office firm size at the time of relocation, the relocation distance from the origin location to the destination location, the regional effects represented by the NSR regions, the density of public transportation facilities, and the rent price (these last three are also based on the location that the firm moves to). As for the relocation distance specifically, firms mostly relocate over short distances (e.g., 5 km). Three variables are defined: "relocation distance 1" comprises categories of short distance intervals up to 5 km, in which the last category includes all the observations with distances larger than 5 km; "relocation distance 2" comprises categories of long distance intervals above 5 km, in which the first category includes all the observations with distances smaller than 5 km; and "relocation distance 3" is a binary variable to define whether the relocation distance is lower or higher than 5 km. There is an obvious overlap among these distance variables, but the idea is not to only capture the influences of distances below and above 5 km, but also the influences in between them. Table 10.5 presents an overview of these preparations, describing the variables, the categories obtained for each variable along with their meaning and the frequency of each category.

Table 10.5. Description of the discrete variables

Variable	Cat.	Meaning	Frequency
Office firm type	1	Agriculture	77
	2	Industry	2426
	3	Basic infrastructures (water, electricity and gas)	88
	4	Building industry	504
	5	Retail and horeca	5350
	6	Traffic and communication	355
	7	Financial institutions	8992
	8	Social security	330
	9	Real estate	6474
	10	Business service	54321
	11	Computer and information technology	9787
	12	Research and development	687
	13	Public administration	1190
	14	Education and health	485
	15	Environmental services, culture and recreation	2685
Office firm size	A	1 employee	31828
	B	Between 2 and 10 employees	51896
	C	Between 11 and 50 employees	7882
	D	Between 51 and 100 employees	1187
	E	More than 100 employees	958

Table 10.5. Description of the discrete variables

Variable	Cat.	Meaning	Frequency
Relocation distance 1	A	Less than 0.5 km	15711
	B	Between 0.5 and 1.5 km	22321
	C	Between 1.5 and 3 km	19162
	D	Between 3 and 5 km	13049
	E	More than 5 km	23508
Relocation distance 2	A	Less than 5 km	70243
	B	Between 5 and 10 km	13579
	C	Between 10 and 20 km	6687
	D	More than 20 km	3242
Relocation distance 3	A	If relocation distance is lower than 5 km	70243
	Z	If relocation distance is higher than 5 km	23508
NSR regions	N	North	10444
	R	Randstad	54614
	S	South	28693
Dutch provinces	A	Drenthe	1346
	B	Flevoland	3174
	C	Friesland	2833
	D	Gelderland	9297
	E	Groningen	2405
	F	Limburg	3224
	G	North Brabant	14601
	H	North Holland	26134
	I	Overijssel	3860
	J	Utrecht	9963
	K	Zeeland	1571
	L	South Holland	15343

Table 10.5. Description of the discrete variables

Variable	Cat.	Meaning	Frequency
Urbanization (municipality level)	A	Very highly urbanized area (more than 2500 addresses/km²)	25669
	B	Highly urbanized area (between 1500 and 2500 addresses/km²)	28537
	C	Moderately urbanized area (between 1000 and 1500 addresses/km²)	19084
	D	Lowly urbanized area (between 500 and 1000 addresses/km²)	14140
	E	Non-urbanized area (less than 500 addresses/km²)	6321
Urbanization (post code level)	A	Very highly urbanized area (more than 2500 addresses/km²)	25214
	B	Highly urbanized area (between 1500 and 2500 addresses/km²)	14264
	C	Moderately urbanized area (between 1000 and 1500 addresses/km²)	7582
	D	Lowly urbanized area (between 500 and 1000 addresses/km²)	12105
	E	Non-urbanized area (less than 500 addresses/km²)	34586
Density of public transportation facilities	A	Less than 2	23722
	B	Between 3 and 5	16244
	C	Between 6 and 9	18625
	D	Between 10 and 15	17228
	E	More than 16	17932
Distance to closest airport	A	Less than 10 km	6703
	B	Between 10 and 25 km	32055
	C	Between 25 and 50 km	25124
	D	Between 50 and 75 km	16693
	E	More than 75 km	13176

Table 10.5. Description of the discrete variables

Variable	Cat.	Meaning	Frequency
Distance to Schiphol	A	Less than 15 km	8809
	B	Between 15 and 45 km	25040
	C	Between 45 and 65 km	15613
	D	Between 65 and 100 km	13701
	E	More than 100 km	30588
Distance to closest international train station	A	Less than 5 km	16506
	B	Between 5 and 12 km	13629
	C	Between 12 and 25 km	17546
	D	Between 25 and 50 km	16523
	E	More than 50 km	29547
Distance to closest intercity train station	A	Less than 2 km	10487
	B	Between 2 and 5 km	23444
	C	Between 5 and 12 km	30231
	D	More than 12 km	29589
Distance to closest highway junction	A	Less than 0.5 km	8394
	B	Between 0.5 and 1 km	17556
	C	Between 1 and 1.8 km	28910
	D	Between 1.8 and 3 km	27305
	E	More than 3 km	11586
Distance to closest shopping center	A	Less than 0.5 km	8655
	B	Between 0.5 and 1 km	16862
	C	Between 1 and 2 km	29264
	D	Between 2 and 3.5 km	19044
	E	More than 3.5 km	19926
Population	A	Less than 4,000 inhabitants/postcode area	18029
	B	Between 4,000 and 7,000 inhabitants/postcode area	21982
	C	Between 7,000 and 10,000 inhabitants/postcode area	23951
	D	More than 10,000 inhabitants/postcode area	29789

Table 10.5. Description of the discrete variables

Variable	Cat.	Meaning	Frequency
Number of households	A	Less than 1,500 households/postcode area	15487
	B	Between 1,500 and 3,000 households/postcode area	21037
	C	Between 3,000 and 4,500 households/postcode area	25980
	D	Between 4,500 and 6,000 households/postcode area	17640
	E	More than 6,000 households/postcode area	13607
Labor force	A	Less than 2,000 employees/postcode area	20379
	B	Between 2,000 and 3,500 employees/postcode area	23931
	C	Between 3,500 and 5,000 employees/postcode area	23570
	D	More than 5,000 employees/postcode area	25871
Average income	A	Less than 30,000 Euro/postcode area	10573
	B	Between 30,000 and 40,000 Euro/postcode area	18803
	C	Between 40,000 and 50,000 Euro/postcode area	19545
	D	Between 50,000 and 60,000 Euro/postcode area	21195
	E	More than 60,000 Euro/postcode area	23635
Number of places in schools	A	Less than 250 places/postcode area	13970
	B	Between 250 and 1,000 places/postcode area	24993
	C	Between 1,000 and 2,000 places/postcode area	25092
	D	Between 2,000 and 4,000 places/postcode area	16898
	E	More than 4,000 places/postcode area	12798
Number of parking places	A	Less than 2,000 places/postcode area	23162
	B	Between 2,000 and 3,200 places/postcode area	23434
	C	Between 3,200 and 4,500 places/postcode area	26039
	D	More than 4,500 places/postcode area	21116
Agglomeration economies (1 km)	A	Less than 100 offices	18305
	B	Between 100 and 200 offices	17775
	C	Between 200 and 400 offices	21439
	D	Between 400 and 900 offices	17145
	E	More than 900 offices	19087

Table 10.5. Description of the discrete variables

Variable	Cat.	Meaning	Frequency
Number of shopping centers within 1 km	A	No shopping centers	52233
	B	Just 1 shopping center	20067
	C	Between 2 and 3 shopping centers	9489
	D	4 or more shopping centers	11962
Rent price / m²	A	Less than € 90	9354
	B	Between € 90 and € 120	33902
	C	Between € 120 and € 135	19283
	D	Between € 135 and € 150	11628
	E	More than € 150	19584

10.4.2 Analyses and results

Table 10.6 presents a summary of the goodness-of-fit of the tested BCN structures, based on the probabilistic R (expected hit ratio) obtained from the validation set. It indicates that the structure that fits best is the BAN No-FS (T=2.0) model. Theoretically, it was expected, given that it is the same structure that performed best in the earlier location choice model. The only difference is the (higher) threshold (T) value. However, given the relatively larger amount of observations used here, this threshold value is the smallest that the system can handle computationally. For smaller values of T, the CPTs become too large.

Figure 10.1 shows the structural relationships between office firm types and location characteristics, obtained from firm observations involved in relocation. Similar to what was discussed in chapter 7, this figure shows the general configuration of the network. The probabilities presented by each category of each attribute represent a-priori beliefs, since no hard evidence to a particular firm type (under the node labeled "SBI") is entered. By instantiating this node successively across the office firm types considered, the updated probabilities related to each attribute become apparent. These findings will not be, however, entirely presented here. Instead, only the differences compared to the earlier location choice model are discussed. Table 10.7 presents the selected results that are of interest in this discussion. For a full table containing all the results obtained in the analysis, please refer to the Appendix.

Table 10.6. Results of the explored models

Model			Probabilistic R (validation set)
Structure	**FS**	**T**	
Naïve	BN-FS	0.4	0.350
		0.6	0.370
		0.8	0.374
	No-FS	Any	0.319
BAN	BN-FS (sample of 10 %)*	1.0	0.371
	BN-FS (sample of 40 % - max.)*	1.0	0.399
	No-FS	2.0	0.470
		3.0	0.424
		4.0	0.416

** The large amount of observations also limited the test of BAN BN-FS structures. Therefore, a random sample of 10 % and 40 % (maximum that can be handled computationally) of observations were initially taken to test this specific structure. However, the tests were not carried out further because of their relatively lower performance compared to the BAN No-FS models*

The additional location attributes considered include office firm size, density of public transportation, rent price, and relocation distance. The effect of NSR regions was also taken into account, but they are not discussed here. As most firms relocate over relatively short distances (e.g., within 5 km), the effects of these NSR regions are not a relevant attribute in the sense that in most cases there is not a change in the current region upon a relocation.

Office firm size is not in fact a spatial attribute that firms would be looking for, when selecting a particular location to relocate their business. However, it is included here as an indication of what is the average firm size (across firm types) when a relocation occurs. It suggests that most office firm types have a preference to relocate when their size is rather small, between 2 and 10 employees. The only exception are office firm types 3 (basic infrastructures) and 13 (public administration). The first has a preference to relocate when its size is more than 100 employees, whereas the latter is likely to relocate when its size is between 11 and 50 employees.

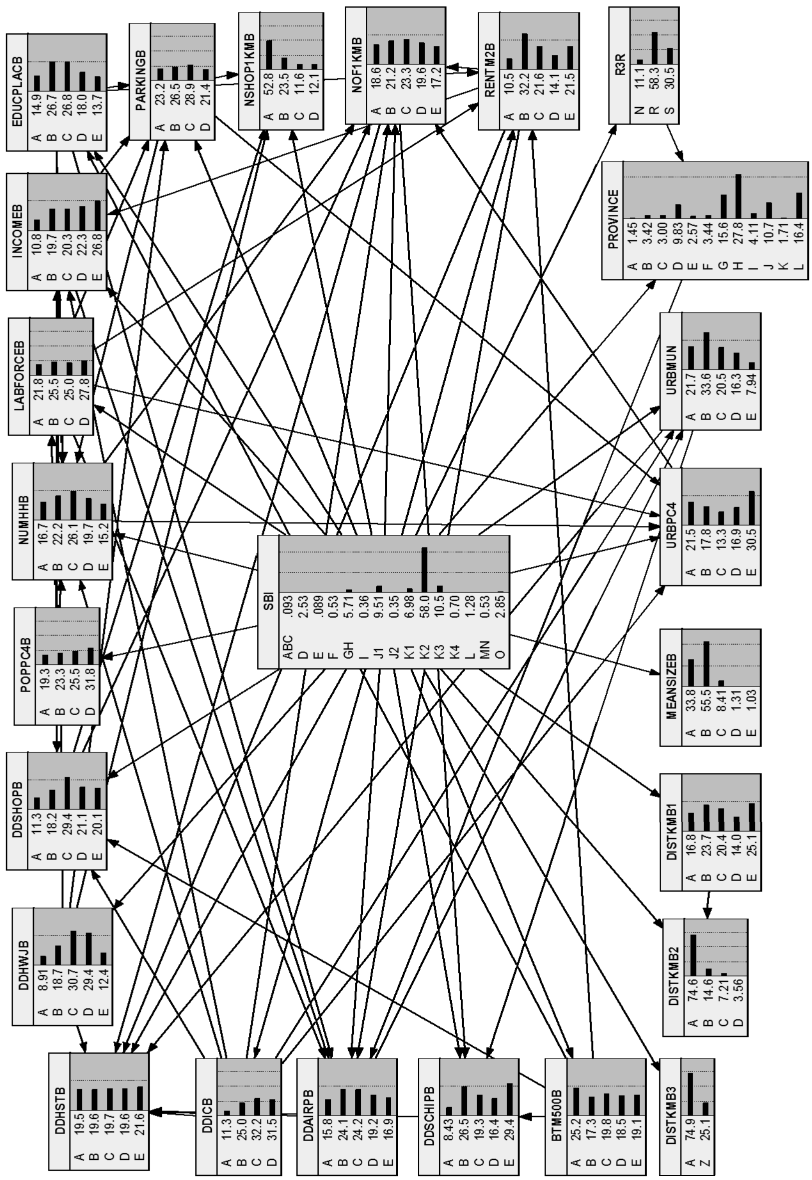


Figure 10.1. Representation of the BAN No-FS (T=2.0) model

Table 10.7. Predicted probabilities of location characteristics across office firm types for selected attributes in the location choice model for relocating office firms

Variable	**Cat.**	**Office firm type class (rounded up values in %)**														
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Firm size	A	20	25	7	26	33	25	20	18	33	37	37	32	3	25	33
	B	64	60	25	59	61	60	69	48	59	54	52	46	19	54	54
	C	9	12	16	13	5	10	9	20	6	8	9	16	35	16	10
	D	4	1	18	2	0	2	2	6	1	1	1	3	17	3	2
	E	3	2	33	0	0	3	1	8	0	1	1	3	27	1	1
Density of public transportation	A	33	25	19	28	28	22	24	20	25	26	24	26	21	20	20
	B	16	17	27	20	17	14	18	16	19	17	17	17	18	16	14
	C	29	21	15	24	20	26	21	18	19	19	21	19	20	22	20
	D	19	19	21	17	18	20	19	21	19	18	19	13	17	19	20
	E	4	18	18	11	16	19	18	25	19	19	19	25	24	23	25
Rent price / m²	A	19	15	19	18	13	18	11	18	12	9	10	15	13	16	10
	B	24	31	21	30	33	24	34	23	33	32	32	26	29	24	27
	C	21	20	21	19	20	20	20	19	20	22	23	18	22	19	19
	D	18	14	21	17	13	17	12	17	12	15	14	17	16	17	16
	E	19	21	19	17	21	22	22	23	22	21	21	24	19	23	27
Reloc. dist. 1	A	10	17	21	12	14	12	21	22	21	16	13	16	29	20	21
	B	29	23	22	19	20	21	25	22	25	24	22	25	32	24	23
	C	9	19	21	17	20	23	18	21	19	21	22	21	18	23	22
	D	21	15	22	14	14	15	13	17	13	14	15	14	8	14	14
	E	31	26	13	38	32	29	23	19	24	25	28	24	13	18	21
Reloc. dist. 2	A	55	74	70	59	67	66	77	76	76	75	72	73	86	78	78
	B	19	13	10	14	18	18	14	15	14	14	16	14	7	11	12
	C	13	8	10	14	10	11	6	4	7	7	8	8	5	6	6
	D	14	4	9	13	5	5	3	5	3	3	4	5	2	5	3
Reloc. dist. 3	A	66	74	84	62	68	70	77	80	76	75	72	76	87	81	79
	Z	34	26	16	38	32	30	23	20	24	25	28	24	13	19	21

Regarding the density of public transportation, office firm types 1 (agriculture), 2 (industry), 4 (building industry), 5 (retail and horeca), 7 (financial institutions), 9 (real estate), 10 (business service), 11 (computer and information technology), and 12 (research and development) relocate to areas where the number of bus, tram, and metro stops is relatively low. If higher density levels of public transportation were related to higher urbanization levels, the quality of the location is suggested to be high, and so are the costs. Therefore, these office firm types also relocate to areas where rent price tend to be lower. As the results indicate, these office firm types search for locations in which rent price levels are between € 90 and € 120 per m².

Office firm types 3 (basic infrastructures) and 6 (traffic and communication) relocate to areas where the density of public transportation is low to moderate: for firm type 3, between 3 and 5 stops; and for firm type 6, between 6 and 9 stops or stations. Regarding rent price levels, firm type 3 has an even preference for low to high costs, between € 90 and € 150 per m², while firm type 6 relocates to areas where rent price levels are still lower (e.g., € 90 and € 120 per m²). Finally, office firm types 8 (social security), 13 (public administration), 14 (education and health), and 15 (environmental services, culture and recreation) have a preference for areas where the density of public transportation is very high. However, only firm type 8 and 15 are willing to relocate where rent price levels are also very high. Firm types 13 and 14 still relocate to areas where rent price is low.

In sum, these findings presented the differences in terms of additional location attributes that were considered for the relocation choice model, which were not included in the earlier location choice model. Although the remaining results are not discussed in detail here, one would expect that largely the same tendencies as described in the location choice model occur.

Notwithstanding the above, an important attribute that is intrinsic to the relocation event and obviously could not have been considered before regards the relocation distance. It gives an indication of the distance that office firms tend to consider when relocating their business. The variables related to relocation distance suggest that office firms have a preference to relocate over distances up to 5 km. This is found for all office firm types. Looking at the individual results, specifically at the "relocation distance 1" variable, most office firm types relocate over a distance between 0.5 and 1.5 km. Exceptions are firm types 3 (basic infrastructure), 5 (retail and horeca), 6 (traffic and communication), and 11 (computer and information technology). Firm type 3 has an even preference for any relocation distance up to 5 km; firm types 5 and

11 also have an even preference, but for relocation distances between 0.5 and 3 km; and firm type 6 has a preference for relocation distances between 1.5 and 3 km. These findings indicate that office firms tend to stay in the same area where they started, and probably try to adjust some minor deficiency related to their current location when, in fact, a relocation occurs.

10.5 Conclusions and discussion

The findings discussed in this chapter allow the following conclusions to be drawn. Firm relocation can be viewed as a two-step process, consisting of the firm's propensity to relocate, followed by the choice of a new location. Forces driving the decision to leave the current location are referred to as push factors and often comprise firm internal characteristics as well as external location attributes. These push factors are seen as deficiencies of the present location that firms want to resolve in order to better adjust their needs for a more optimal production environment. On the other hand, the forces attracting firms to a specific location are seen as pull factors, which comprise the desired location characteristics underlying the firm's location choice process.

The literature shows a long tradition on examining the firm's propensity to relocate in terms of its internal characteristics. Behavioral principles explain that firm size and age, as well as stage of lifecycle and economic sector are influencing factors on the firm's decision to relocate. The influences of location aspects on this decision have been investigated in urban and transportation planning, specifically in LUTI modeling practices. But the efforts have not given very detailed and conclusive responses. The analyses carried out here tried to bring better understanding on the relationship between the location characteristics and the relocation decision process. The findings show significant influences across the location attributes considered, and some of them are in line with previous studies. Overall, access to transportation infrastructure, effects of agglomeration economies, socioeconomic aspects, urbanization levels and regional characteristics seem to contribute to the firm's decision to relocate.

As for the choice of the new location when a relocation decision is made, the influencing factors comprise basically spatial characteristics. There is not much difference in terms of modeling approaches compared to what is done for the location choice process when a firm starts-up. Once a firm decides to relocate, the search process can be seen as a location choice problem. Thus, location theories and modeling practices, as discussed before in chapter 7, can be applied. To a large extent, the same spatial tendencies are expected, as described before. However, there is an important and new factor here: the distance over which firms relocated. It regards an influential factor in the relocation process that should be taken into account. The findings suggest that relocations are over short distances, up to 5 km, indicating that office firms may relocate to adjust minor deficiencies in their current location, but they would still remain in the same area where they were firstly established.

Overall, the findings obtained with the analyses presented here can be used in a component of a multi-agent system to simulate the office firm relocation process. The methods explored seem to be a valuable approach to this end. However, it should not go without saying that these findings do not put an end to the investigations, especially regarding the examination of external location influences on the firm relocation decision. The investigations carried out here confirm that location attributes do need to be considered along with firm internal factors, but the results are to be confronted with further studies; since some specific findings obtained with related location attributes included here are relatively new. Next chapter will present the final conclusions of the work carried out in this research project, as well as bring some directions for future investigations.

11

Conclusions and future work

11.1 Concluding remarks

The focus of this research project was on investigating office firm dynamics. Nowadays, a large share of economic activities is represented by office firms, especially in the sector of professional services. Increasing urbanization and specialization of employment were processes observed in the last decades, which have been contributing to a shift in the bases of the economy to the service sector. Consequently, an intense activity among people's behavior, services and infrastructure systems is observed in cities. This generates land use and travel patterns that influence especially the transportation infrastructure systems. From an urban and transportation planning perspective, the proper examination of these interactions constitutes a major concern to establish policies of land use and organize measures to plan, build and manage infrastructure systems. In line with this, LUTI models have been traditionally employed to examine these interactions.

The contribution aimed at this research was modeling office firm dynamics within a LUTI framework. More specifically, it was based on a firm demography approach that analyzes the events of firm start-up, growth, decline, and closure as well as location choice and relocation decision patterns. The current trends in LUTI models, which are based on agent-based and micro simulation models, have found strong linkages with the firm demographic approach, capable to accommodate behaviorally richer concepts of firm dynamics. While several applications dealing with firms in general can be found, the specific interest aimed here was in the behavior of office firms. This comes along with the influences underlain by the urban environment. In LUTI modeling, the specification in detail of the factors related to urban processes is required in order to obtain as much realistic models as possible. This was explored here, by including and analyzing a large set of location variables as well as trying to define them in a high level of detail, using nationwide data from The Netherlands.

The approach taken to investigate office firm dynamics was based on a set of econometric/statistical models that were employed to examine the related demographic events as a function of both firm internal attributes and location characteristics. The start-up model was based on the idea that firms must exist before it can be simulated and it can only be modeled by taking an area as a decision unit. In fact, the model does not deal with the actual start-up behavior, but simulates an environment where firms come into existence every year as a result of certain urban conditions. Therefore, the frequency of firms starting in these areas is modeled using the Poisson regression model. The main factors influencing firm start-up include effects of agglomeration economies, high population levels, and some accessibility to transportation infrastructure such as high density of public transportation.

Given that the firm dynamics investigated here operate in an (urban) environment, it is expected that some constraints apply as a result of existing available resources and market demands. Therefore, the notion of carrying capacity was introduced to control the firm demographic processes in relation to the urban conditions. A stochastic frontier model was investigated and its production form was associated to the carrying capacity notion. The approach considered that the carrying capacity of a region is measured in terms of a maximum number of jobs that the region can support, based on the size of the population in the region. Population imposes a maximum to the production either as a resource of labor (in case of basic industries) or as limiting the demand for products (in case of non-basic industries). However, the stochastic frontier model also presents a cost form and this form fitted best most office firm types according to the data used. In fact, the case is that either a production or a cost function usually fits a data distributional structure. Therefore, in cases where a cost function rather than a production function fitted the data, the conclusion was that the notion of carrying capacity is not supported by the data. Such a finding indicates that the population defines a minimum rather than a maximum frontier. As a source of demand or supplier of labor, population is not a limiting factor for the sector's size: the sector could grow further even if the population stays constant.

As a critical part of firm demographic models, the spatial distribution of economic activities was investigated by an approach that views the location choice decision as a matching process of, on the one hand, a set of firm requirements and, on the other, a set of location characteristics. More specifically, office firm types (the class labels or dependent variable) are matched with the location attributes (the independent variables) using a classification task approach based on Bayesian classifier networks. In fact, the model proposed here does not actually describe the choice of discrete locations, but the probability of having certain location characteristics related to each office firm type. In order to find which (concrete) locations firms would be choosing to locate their business, a spatial component is needed. If those selected location characteristics could be matched, for example, with a geographical database of available locations, such location decision mechanism could be derived and an allocation process between firm and location could be established.

Following the course of lifecycle, firms will grow and decline. The approach developed here to investigate this process is derived from the well-known logistic growth function. In fact, it consists of an approximation based on a quadratic function, in which growth can be described as a function of firm size. Both quadratic and linear firm size terms are included in the right-hand side of the equation to make sure that growth is an increasing function for initial values of current size, followed by a decreasing function for larger values of current size, which can accommodate the firm decline process. The examination of external factors influencing firm growth/decline is dealt with the inclusion of location attributes as interaction terms of firm size. As opposed to the logistic growth function, this approximation turned out to be a more flexible approach, capable to accommodate different sets of firm size-development patterns as well as to account for external factors on the firm growth/decline process. The findings suggested that location attributes do play a role and they should be considered along with firm internal characteristics when investigating firm growth/decline processes in the perspective of LUTI models.

Upon establishing a business, firms may consider relocating during its lifecycle. Although they expect to minimize this event as much as possible, due to large costs involved and risks of losing established markets, a relocation decision is often associated with an improvement in the quality of the location in order to maintain the firm's best level of service. The assumption is that both firm internal characteristics and external (location) factors play a role in this decision. A modeling approach to examine the determinants of this decision is rather straightforward as a binary response model can be easily applied. However, the longitudinal nature of the data available here allowed the exploration of duration models. They consider not only whether or not a firm decides to relocate, but also take into account the elapsed time before this decision. Among the push factors (e.g., the factors that motivate a relocation decision) investigated here, firm internal characteristics such as firm size and whether a firm was involved in a relocation before suggest, respectively, that larger firms tend to avoid relocating and firms that relocated before tend to avoid relocating again. The location attributes also indicated significant influences, and some of them were in line with previous studies. Overall, access to transportation infrastructure, effects of agglomeration economies, socioeconomic aspects, urbanization levels and regional characteristics seem to contribute to the firm's decision to relocate.

Once a firm decides to relocate, a process of finding a new location begins. Usually, there is not much difference in terms of modeling approaches compared to what is done for the location choice process when a firm starts-up. Thus, the search process can be seen as a location choice problem. The same modeling approach employed for the location choice based on the Bayesian classifier networks were applied to examine the pull factors (e.g., the factors attracting firms to a specific location). To a large extent, the same spatial tendencies compared to a first location decision are expected. However, there is an important and new factor here: the distance over which firms relocated. It regards an influential factor in the relocation process that should be taken into account. The findings obtained here suggest that relocations are over short distances, up to 5 km, indicating that office firms may relocate to adjust minor deficiencies in their current location, but they would still remain in the same area where they firstly were established.

At the end of a lifecycle, firms will go out of business either because of bankruptcy or voluntary exit. In addition to these, the determinants of firms' closing patterns are related to firm internal characteristics, such as: age and size. But location characteristics may also influence the firm closure process. The focus of the closure model was on the investigation of the influence of these external factors on firm closure. Duration models were also applied. Similar to the relocation decision event, given the longitudinal data available, the model is capable of not only taking into account whether a firm is closed, but also to consider the duration that a firm existed before going out of business. The findings suggest that firm closure is highly influenced by location attributes, as showed by most of the variables included. This is also in line with the aim of the analysis proposed, suggesting an evident relationship between firm closure and location attributes. The analysis carried out here brought a better understanding about the firm closure process by using an extended set of location attributes.

Overall, from a LUTI modeling perspective, the findings obtained across the several location attributes considered brought a better understanding in relation to the influences of these attributes on firm dynamics, more specifically, on the behavior of office firms. Compared to earlier efforts found in the literature, the set of location attributes appears to be more comprehensive and some factors that were not explored in detail before could be examined here. This comes along with the development of the analyses using nationwide data for The Netherlands. Therefore, one of the objectives of this research project could be fulfilled. The examination of office firm dynamics using behaviorally richer concepts of firm demography along with the current trends in LUTI models, which is based on agent-based modeling, is also another achievement. The methods and empirical analyses carried out introduce the keystones for a modeling framework to simulate office firm dynamics.

Despite the above, some minor comments to improve these methods and empirical analyses comprise a refining process across the several models in terms of including attributes that were not considered previously. For example, the effects of agglomeration economies were basically derived in terms of firms of the same type. However, a more detailed specification could be observed in the start-up model, where these effects were derived in terms of both firms of same type and firms of different type. These could be interesting to be examined in the other models as well. Another example is the case of including rent price in the location choice model. Certainly a list of such details can be formulated when checking the specification of the models across this thesis, but it would not have major impacts on the findings depicted here. Instead, it is important to think of the contributions that these findings can provide in terms of an overall simulation system. They will be discussed in the next section.

11.2 Toward an integrated model

One of the main questions at the end of this thesis regards to the integration of the several models developed here into a simulation framework of office firm dynamics. The final product of this research does not result in such a fully operational simulation, but the methods developed and explored along with the parameters obtained are intended to build it. In short, a simulation system begins with a micro data of the population of office firms that have predefined internal characteristics. Based on these characteristics, location preferences to establish their business are derived. A component with available locations and their characteristics is used to match with the derived firm's location preferences. Given that one firm must be assigned to one location, a one-to-one allocation process must be carried out based on some kind of decision rules that ranks locations across firms. The simulation system then evaluates firm internal characteristics as well as location attributes. Based on these factors, firms will grow or decline and the firm size is updated. The carrying capacity component checks the conditions for this. For some office firm types, there will be limiting conditions for start-up and growth processes in terms of generating more jobs. On the other hand, for other sectors, the simulation system will have to keep a minimum level in the number of jobs to attend the demand. In turn, the relocation component is activated and firms are classified in either moving or non-moving firms, also based on both internal characteristics and location attributes. The location component derives location preferences for these moving firms, the allocation process is carried out, and the location ranks across firms are recalculated. At the end, the closure component evaluates firms that will go out business and classifies them in either surviving or closed firms. The component with available locations is updated, due to firms that went out business. The system restarts and, based on its characteristics regarding both firm spatial distributions and location attributes, new firms start-up and the process described above is repeated.

The above indicates that some additional components have not been described before, but they appear to be important within the overall simulation framework and must be, thus, developed. The first regards to a real estate component that controls the available office locations. It operates along with the location choice model and is updated every time that a relocation decision is taken and when a firm goes out of business. Secondly, the location choice component implies the use of an allocation procedure that assigns a firm to a location on a one-to-one allocation process. As there may be more than one firm that best fits one location, the allocation algorithm could be then described in terms of a price mechanism and some form of auction or bidding process to determine which firm gets which location. It assumes a utility function and a process where firms make bids that reflect the utility of a location for the firm. The highest bidder gets the location. Thirdly, location aspects may change over time as a result of population changes, investments in infrastructure, land policies, etc. Therefore, a component to account for these changes and update the system accordingly may be also an important feature in such a simulation. Lastly, economic activities are highly sensitive to the prospects of the global economy. This issue is difficult to predict, and so is its incorporation into the simulation system. An alternative could be the simulation of different scenarios where the overall framework can deal with the impact of global crisis, or economy expansion, for example. The parameters obtained could be adjusted manually in such a way that, respectively, firms decline and close more often, or more firms will be born and live longer. However, more research is needed at this level as this would probably require extensions of the existing databases and models developed here.

References

Abraham, J.E., Garry, G., & Hunt, J.D. (2005) The Sacremento PECAS model. In *Proceedings of the 84th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Abraham, J.E. & Hunt, J.D. (1999) Firm location in the MEPLAN model of Sacramento. *Transportation Research Record: Journal of the Transportation Research Board*, **1685**, 187–198.

Aigner, D., Lovell, C.A.K., & Schmidt, P. (1977) Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, **6**, 21–37.

Alexander, I. (1979) *Office Location and Public Policy*. Longman, London.

Allaman, P.A. & Birch, D.L. (1975) *Components of Employment Change for States by Industry Group 1970-1972*, Working paper No. 5. Harvard University, MIT Joint Center for Urban Studies, Cambridge Mass.

Allison, P.D. (2004) Event History Analysis. In Bryman, A.E. & Hardy, M. (eds), *Handbook of Data Analysis*. Sage, Thousand Oaks, pp. 369–386.

Anas, A. (1982) *Residential Location Markets and Urban Transportation*. Academic Press, New York.

Anas, A. (1983a) Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B: Methodological*, **17**, 13–23.

Anas, A. (1983b) The Chicago Area Transportation-Land Use Analysis System, Final report to US Dept. of Transportation. Washington, D.C.

Anas, A. (1994) METROSIM, A Unified Economic Model of Transportation and Land Use. In *Brochure Transportation Model Improvement Program's Land Use Modelling Conference*. Dallas, Texas.

Andersen, S.K., Olesen, K.G., Jensen, F.V., & Jensen, F. (1989) Hugin - a Shell for Building Bayesian Belief Universes for Expert Systems. In *11th International Joint Conference on Artificial Intelligence*. Detroit, pp. 1080–1085.

Anderstig, C. & Mattsson, L. (1991) An Integrated Model of Residential and Employment Location in a Metropolitan Region. *Papers in Regional Science*, **70**, 167–184.

Anderstig, C. & Mattsson, L. (1992) Appraising large-scale investments in a metropolitan transportation system. *Transportation*, **19**, 267–283.

Anderstig, C. & Mattsson, L. (1998) Modelling land use and transport interaction: policy analyses using the IMREL model. In Lundqvist, L., Mattsson, L., & Kim, T.J. (eds), *Network Infrastructure and the Urban Environment*. Springer, pp. 308–328.

Anggraini, R. (2009) Household activity-travel behavior: implementation of within-household interactions.

Appel-Meulenbroek, R. (2008) Managing "keep" factors of office tenants to raise satisfaction and loyalty. *Property Management*, **26**, 43–55.

Arentze, T.A., Borgers, A.W.J., Ma, L., & Timmermans, H.J.P. (2010) An agent-based heuristic method for generating land-use plans in urban planning. *Environment and Planning B: Planning and Design*, **37**, 463–482.

Arentze, T.A., Hofman, F., Mourik, H., Timmermans, H.J.P., & Wets, G. (2000) Using decision tree induction systems for modeling space-time behavior. *Geographical Analysis*, **32**, 330–350.

Arentze, T.A. & Timmermans, H.J.P. (2003) A multiagent model of negotiation processes between multiple actors in urban developments: a framework for and results of numerical experiments. *Environment and Planning B: Planning and Design*, **30**, 391–410.

Arentze, T.A. & Timmermans, H.J.P. (2006) A new theory of dynamic activity generation. In *Proceedings of the 85th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Arentze, T.A. & Timmermans, H.J.P. (2009) Regimes in social-cultural events-driven activity sequences: Modelling approach and empirical application. *Transportation Research Part A: Policy and Practice*, **43**, 311–322.

Audretsch, D.B. (1995) Innovation, growth and survival. *International Journal of Industrial Organization*, **13**, 441–457.

Audretsch, D.B. & Mahmood, T. (1995) New Firm Survival: New Results Using a Hazard Function. *The Review of Economics and Statistics*, **77**, 97–103.

Axelrod, R.M. (1997) *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration*. Princeton University Press, Princeton.

Bade, F.J. (1983) Locational Behaviour and the Mobility of Firms in West Germany. *Urban Studies*, **20**, 279–297.

Baldwin, J., Bian, L., Dupuy, R., & Gellatly, G. (2000) Failure rates for new canadian firms: new perspectives on entry and exit (No. Catalogue no. 61-526-XIE). Statistics Canada, Ottawa.

de la Barra, T., Pérez, B., & Vera, N. (1984) TRANUS-J: putting large models into small computers. *Environment and Planning B: Planning and Design*, **11**, 87–101.

Batty, M. (2005) Agents, cells, and cities: new representational models for simulating multiscale urban dynamics. *Environment and Planning A*, **37**, 1373–1394.

Batty, M., Couclelis, H., & Eichen, M. (1997) Urban systems as cellular automata. *Environment and Planning B: Planning and Design*, **24**, 159–164.

Batty, M. & Xie, Y. (1994) From cells to cities. *Environment and Planning B: Planning and Design*, **21**, s31–s48.

Baum, J.A.C. & Singh, J.V. (1994) Organizational Niches and the Dynamics of Organizational Mortality. *American Journal of Sociology*, **100**, 346–380.

Berglund, E. & Brännäs, K. (2001) Plants' entry and exit in Swedish municipalities. *The Annals of Regional Science*, **35**, 431–448.

Birch, D.L. (1979) *The Job Generation Process*. Cambridge University Press, Cambridge Mass.

Bodenmann, B.R. (2011a) Modelling firm (re-)location choice in UrbanSim. In *51st European Congress of the Regional Science Association*. Barcelona.

Bodenmann, B.R. (2011b) Location choice of firms with special emphasis on spatial accessibility.

de Bok, M. (2007) Infrastructure and firm dynamics: a micro-simulation approach.

de Bok, M. & Bliemer, M.C.J. (2006a) Land use and transportation interaction: calibration of a micro simulation model for firms in the Netherlands. In *Proceedings of the 85th. Annual Meeting of the Transportation Reseach Board*. Washington, D.C.

de Bok, M. & Bliemer, M.C.J. (2006b) Infrastructure and Firm Dynamics: Calibration of Microsimulation Model for Firms in the Netherlands. *Transportation Research Record: Journal of the Transportation Research Board*, **1977**, 132–144.

de Bok, M. & Sanders, F. (2005a) Firm location and the accessibility of locations: empirical results from The Netherlands. In *Proceedings of the 84th Annual Meeting of the Transportation Research Board*. Washington, D.C.

de Bok, M. & Sanders, F. (2005b) Firm Relocation and Accessibility of Locations Empirical Results from the Netherlands. *Transportation Research Record: Journal of the Transportation Research Board*, **1902**, 35–43.

Bonabeau, E. (2002) Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 7280–7287.

Boschma, R.A. (2009) Evolutionary economic geography and its implications for regional innovation policy. OECD.

Boschma, R.A. & Frenken, K. (2006a) Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography*, **6**, 273–302.

Boschma, R.A. & Frenken, K. (2006b) Applications of Evolutionary Economic Geography (No. 06-26), DRUID Working Paper. Danish Research Unit for Industrial Dynamics.

Boschma, R.A. & Frenken, K. (2011) The emerging empirics of evolutionary economic geography. *Journal of Economic Geography*, **11**, 295–307.

Boschma, R.A. & Lambooy, J.G. (1999) Evolutionary economics and economic geography. *Journal of Evolutionary Economics*, **9**, 411–429.

Boschma, R.A. & Martin, R. (2007) Editorial: Constructing an evolutionary economic geography. *Journal of Economic Geography*, **7**, 537–548.

Boyce, D. & Mattsson, L. (1999) Modeling residential location choice in relation to housing location and road tolls on congested urban highway networks. *Transportation Research Part B: Methodological*, **33**, 581–591.

Breschi, S., Malerba, F., & Orsenigo, L. (2000) Technological Regimes and Schumpeterian Patterns of Innovation. *The Economic Journal*, **110**, 388–410.

Brouwer, A.E., Mariotti, I., & van Ommeren, J.N. (2004) The firm relocation decision: An empirical investigation. *The Annals of Regional Science*, **38**, 335–347.

Brüderl, J. & Schüssler, R. (1990) Organizational Mortality: The Liabilities of Newness and Adolescence. *Administrative Science Quarterly*, **35**, 530–547.

Cameron, G.C. & Clark, B.D. (1966) *Industrial Movement and the Regional Problem*, University of Glasgow Social and Economic Studies, Occasional papers No. 5. Oliver & Boyd, Edinburgh and London.

Carroll, G.R. (1984) Organizational Ecology. *Annual Review of Sociology*, **10**, 71–93.

Carroll, G.R. & Hannan, M.T. (2000) *The Demography of Corporations and Industries*. Princeton University Press.

Cecchini, A. (1996) Urban modelling by means of cellular automata: generalised urban automata with the help on-line (AUGH) model. *Environment and Planning B: Planning and Design*, **23**, 721–732.

Chandra, A. & Thompson, E. (2000) Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, **30**, 457–490.

Cheng, J. & Greiner, R. (1999) Comparing Bayesian Network Classifiers. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*.

Cheng, J. & Greiner, R. (2001) Learning Bayesian Belief Network Classifiers: Algorithms and System. In *Proceedings of the Fourteenth Canadian Conference on Artificial Intelligence*.

Cheng, J., Greiner, R., Kelly, J., Bell, D., & Liu, W. (2002) Learning Bayesian networks from data: An information-theory based approach. *Artificial Intelligence*, **137**, 43–90.

Christaller, W. (1933) *Die Zentralen Orte in Suddeutschland*. Gustav Fischer, Jena.

Clarke, K.C., Hoppen, S., & Gaydos, L. (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design*, **24**, 247–261.

Collett, D. (1994) *Modelling Survival Data in Medical Research*. Chapman & Hall, London.

Cooper, M.J.M. (1975) *The Industrial Location Decision Making Process*. Centre for Urban and Regional Studies, Birmingham.

Cox, W.E. (1967) Product Life Cycles as Marketing Models. *The Journal of Business*, **40**, 375–384.

Crecine, J.P. (1964) TOMM (No. 6), CRP Technical Bulletin. Department of City and Regional Planning, Pittsburgh.

Crecine, J.P. (1968) A Dynamic Model of Urban Structure (No. P-3803). RAND Corporation, Santa Monica.

Daniels, P.W. (1975) *Office Location: An Urban and Regional Study*, Urban and social geography series. G. Bell, London.

Dean, J. (1950) Pricing policies for new products. *Harvard Business Review*, **28**, 45–53.

Devisch, O.T.J. (2008) In search of a complex system model: the case of residential mobility.

Devisch, O.T.J., Arentze, T.A., Borgers, A.W.J., & Timmermans, H.J.P. (2006) Bilevel Negotiation Protocol for Multiagent Simulation of Housing Transactions and Market Clearing Processes. *Transportation Research Record: Journal of the Transportation Research Board*, **1977**, 84–92.

van Dijk, J. & Pellenbarg, P.H. (1999a) *Demography of Firms: Spatial Dynamics of Firm Behaviour*. Koninklijk Nederlands Aardrijkskundig Genootschap; Faculteit Ruimtelijke Wetenschappen Rijksuniv, Utrecht; Groningen.

van Dijk, J. & Pellenbarg, P.H. (1999b) The demography of firms: progress and problems in empirical research. In van Dijk, J. & Pellenbarg, P.H. (eds), *Demography of Firms: Spatial Dynamics of Firm Behaviour*. Rijksuniversiteit Groningen, Nederlandse geografische studies, pp. 325–337.

van Dijk, J. & Pellenbarg, P.H. (2000a) Firm relocation decisions in The Netherlands: An ordered logit approach. *Papers in Regional Science*, **79**, 191–219.

van Dijk, J. & Pellenbarg, P.H. (2000b) Spatial perspectives on firm demography. *Papers in Regional Science*, **79**, 107–110.

van Dijk, J., Pellenbarg, P.H., & van Steen, P.J.M. (1999) Determinants of firm migration in the Netherlands: An exercise in the demography of firms approach. In van Dijk, J. & Pellenbarg, P.H. (eds), *Demography of Firms: Spatial Dynamics of Firm Behaviour*. Netherlands Geographical Studies, KNAG, RUG, pp. 87–121.

Echenique, M.H. (1994) Urban and regional studies at the Martin Centre: its origins, its present, its future. *Environment and Planning B: Planning and Design*, **21**, 517–533.

Echenique, M.H., Flowerdew, A.D.J., Hunt, J.D., Mayo, T.R., Skidmore, I.J., & Simmonds, D.C. (1990) The MEPLAN models of Bilbao, Leeds and Dortmund. *Transport Reviews*, **10**, 309–322.

Ekamper, P. (1996) Opheffing van bedrijfsvestigingen: een sterftetafelbenadering. *Planning, Methodiek en Toepassing*, **48**, 12–21.

Elgar, I., Farooq, B., & Miller, E.J. (2009) Modeling location Decisions of Office Firms: Introducing Anchor Points and Constructing Choice-Sets into the Model System. In *Proceedings of the 88th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Elgar, I., Miller, E.J., & Nurul Habib, K.M. (2008) Office Decisions to Change Location: A Stress Triggered Approach. In *Proceedings of the 87th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Elgar, I. & Miller, E.J. (2006) Conceptual Model of Location of Small Office Firms. *Transportation Research Record: Journal of the Transportation Research Board*, **1977**, 190–196.

Elgar, I. & Miller, E.J. (2007) Office Location Decisions: Analysis of the Results of SOLD. In *Proceedings of the 86th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Elgar, I. & Miller, E.J. (2010) How Office Firms Conduct Their Location Search Process?: An Analysis of a Survey from the Greater Toronto Area. *International Regional Science Review*, **33**, 60–85.

Eliasson, J. (2000) Integrated travel pattern modelling: theory and estimation of an integrated land use-transportation model. In *Proceedings of the IATBR Meeting: International Association for Travel Behaviour Research*.

Eliasson, J. & Mattsson, L. (2001) Transport and Location Effects of Road Pricing: A Simulation Approach. *Journal of Transport Economics and Policy*, **35**, 417–456.

Epstein, J.M. & Axtell, R. (1996) *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press, Cambridge.

Ettema, D.F., de Jong, K., Timmermans, H.J.P., & Bakema, A. (2007) PUMA: Multi-Agent Modelling of Urban Systems. In Koomen, E., Stillwell, J., Bakema, A., & Scholten, H.J. (eds), *Modelling Land-Use Change*. Springer Netherlands, Dordrecht, pp. 237–258.

Evans, D.S. (1987) The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries. *The Journal of Industrial Economics*, **35**, 567–581.

Fetter, F. (1924) The economic law of market areas. *The Quarterly Journal of Economics*, **XXXVIII**, 520–529.

Fotopoulos, G. & Louri, H. (2000) Location and Survival of New Entry. *Small Business Economics*, **14**, 311–321.

Freeman, J. (1982) Organizational life cycles and natural selection processes. In Staw, B. & Cummings, L. (eds), *Research in Organizational Behavior*. JAI, Greenwich, Conn., pp. 1–32.

Frenkel, A. (2001) Why High-technology Firms Choose to Locate in or near Metropolitan Areas. *Urban Studies*, **38**, 1083–1101.

Frenken, K. & Boschma, R.A. (2007) A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process. *Journal of Economic Geography*, **7**, 635–649.

Garofoli, G. (1994) New Firm Formation and Regional Development: The Italian Case. *Regional Studies*, **28**, 381–393.

Geroski, P.A. (1995) What do we know about entry? *International Journal of Industrial Organization*, **13**, 421–440.

Gibrat, R. (1931) *Les Inégalités Économiques*. Paris, France.

Goldner, W. (1971) The Lowry Model Heritage. *Journal of the American Institute of Planners*, **37**, 100–110.

Goldner, W., Rosenthall, S., & Meredith, J. (1972) Projective Land Use Model-PLUM: Theory and Application. Institute of Transportation and Traffic Engineering, University of California, Berkeley.

Greene, W.H. (2007) *NLOGIT Version 4.0: Reference Guide*. Econometric Software Inc., Plainview, USA and Castle Hill, Australia.

Hannan, M.T., Carroll, G.R., Dobrev, S.D., & Han, J. (1998a) Organizational Mortality in European and American Automobile Industries Part I: Revisiting the Effects of Age and Size. *European Sociological Review*, **14**, 279–302.

Hannan, M.T., Carroll, G.R., Dobrev, S.D., Han, J., & Torres, J.C. (1998b) Organizational Mortality in European and American Automobile Industries Part II: Coupled Clocks. *European Sociological Review*, **14**, 303–313.

Hannan, M.T. & Freeman, J. (1977) The Population Ecology of Organizations. *American Journal of Sociology*, **82**, 929–964.

Hannan, M.T. & Freeman, J. (1987) The Ecology of Organizational Founding: American Labor Unions, 1836-1985. *American Journal of Sociology*, **92**, 910–943.

Hannan, M.T. & Freeman, J. (1989) *Organizational Ecology*. Harvard University Press, Cambridge.

Harhoff, D., Stahl, K., & Woywode, M. (1998) Legal Form, Growth and Exit of West German Firms - Empirical Results for Manufacturing, Construction, Trade and Service Industries. *The Journal of Industrial Economics*, **46**, 453–488.

Hart, M. & Scott, R. (1994) Measuring the Effectiveness of Small Firm Policy: Some Lessons from Northern Ireland. *Regional Studies*, **28**, 849–858.

Hart, P.E. & Oulton, N. (1996) Growth and Size of Firms. *The Economic Journal*, **106**, 1242–1252.

Healey, M.L. & Ilbery, B.W. (1990) *Location and Change: Perspectives on Economic Geography*. Oxford University Press, Oxford.

Heckerman, D., Mandani, A., & Wellman, M.P. (1995) Real-world Applications of Bayesian Networks. *Communications of the ACM*, **38**, 24–26.

Hilbers, H., Jorritsma, P., & Louter, P. (1994) The relationship between accessibility and regional development explored for the region of Amersfoort. In *Proceedings of the 21st Colloquium Vervoerplanologisch Speurwerk*.

Hill, E.J., Ferris, M., & Märtinson, V. (2003) Does it matter where you work? A comparison of how three work venues (traditional office, virtual office, and home office) influence aspects of work and personal/family life. *Journal of Vocational Behavior*, **63**, 220–241.

Holl, A. (2004) Transport Infrastructure, Agglomeration Economies, and Firm Birth: Empirical Evidence from Portugal. *Journal of Regional Science*, **44**, 693–712.

Hoogstra, G.J. & van Dijk, J. (2004) Explaining Firm Employment Growth: Does Location Matter? *Small Business Economics*, **22**, 179–192.

Hosmer, D.W. & Lemeshow, S. (1999) *Applied Survival Analysis: Regression Modeling of Time-to-event Data*. Wiley & Sons Inc., USA.

Huhns, M.N. & Stephens, L.M. (1999) Multiagent systems and societies of agents. In Weiss, G. (ed), *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press, Cambridge, pp. 79–120.

Huisman, C. & van Wissen, L.J.G. (2004) Localization effects of firm startups and closures in the Netherlands. *The Annals of Regional Science*, **38**, 291–310.

Hunt, J.D. (1994) Calibrating the Naples land-use and transport model. *Environment and Planning B: Planning and Design*, **21**, 569–590.

Hunt, J.D. (2002) Agent Behaviour Issues Arising with Urban System Micro-Simulation. *European Journal of Transport and Infrastructure Research*, **2**, 233–254.

Iacono, M., Levinson, D., & El-Geneidy, A. (2008) Models of Transportation and Land Use Change: A Guide to the Territory. *Journal of Planning Literature*, **22**, 323–340.

Janssens, D., Wets, G., Brijs, T., Vanhoof, K., Arentze, T.A., & Timmermans, H.J.P. (2004) Improving Performance of Multiagent Rule-Based Model for Activity Pattern Decisions with Bayesian Networks. *Transportation Research Record: Journal of the Transportation Research Board*, **1894**, 75–83.

Janssens, D., Wets, G., Brijs, T., Vanhoof, K., Arentze, T.A., & Timmermans, H.J.P. (2006) Integrating Bayesian networks and decision trees in a sequential rule-based transportation model. *European Journal of Operational Research*, **175**, 16–34.

Johnston, R.A., Gao, S., & Shabazian, D.R. (2003) Uplan: a versatile urban growth model for transportation planning. In *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*. Washington, D.C.

Jovanovic, B. (1982) Selection and the Evolution of Industry. *Econometrica*, **50**, 649–670.

Katoshevski-Cavari, R., Arentze, T.A., & Timmermans, H.J.P. (2010) A Multi-Agent Planning Support-System for Assessing Externalities of Urban Form Scenarios: Results of Case Studies. *Sustainability*, **2**, 2253–2278.

Kemper, N.J. & Pellenbarg, P.H. (1993) Bedrijfsverplaatsing in Nederland 1990, 1991. *Economisch Statistische Berichten*, **78**, 380–384.

Kemper, N.J. & Pellenbarg, P.H. (1995) Een vlucht uit de Randstad? *Economisch Statistische Berichten*, **80**, 465–469.

Kemper, N.J. & Pellenbarg, P.H. (1997) De Randstad een hogedrukpan. *Economisch Statistische Berichten*, **82**, 508–512.

Khan, A.S., Abraham, J.E., & Hunt, J.D. (2002a) Agent-based Micro-simulation of Business Establishments. In *Proceedings of the 42nd Congress of the European Regional Science Association*. Dortmund.

Khan, A.S., Abraham, J.E., & Hunt, J.D. (2002b) A system for microsimulating business establishments: analysis, design and results. In *International Colloquium on the Behavioural Foundations of Integrated Land-Use and Transportation Models*. Quebec City.

Kiefer, N.M. (1988) Economic Duration Data and Hazard Functions. *Journal of Economic Literature*, **26**, 646–679.

Kleinbaum, D.G. & Klein, M. (2005) *Survival Analysis: A Self-Learning Text*. Springer, USA.

Klepper, S. (1996) Entry, Exit, Growth, and Innovation over the Product Life Cycle. *The American Economic Review*, **86**, 562–583.

Klepper, S. (1997) Industry Life Cycles. *Industrial and Corporate Change*, **6**, 145–182.

Knoben, J., Ponds, R., & van Oort, F.G. (2011) Employment from new firm formation in the Netherlands: Agglomeration economies and the Knowledge Spillover Theory of Entrepreneurship. *Entrepreneurship & Regional Development*, **23**, 135–157.

Koomen, E. & Stillwell, J. (2007) Modelling Land-Use Change: Theories and methods. In Koomen, E., Stillwell, J., Bakema, A., & Scholten, H.J. (eds), *Modelling Land-Use Change*, *The GeoJournal Library*. Springer Netherlands, pp. 1–21.

Krugman, P. (1998) What's new about the new economic geography? *Oxford Review of Economic Policy*, **14**, 7–17.

Landis, J.D. (1994) The California Urban Futures Model: a new generation of metropolitan simulation models. *Environment and Planning B: Planning and Design*, **21**, 399–420.

Landis, J.D. & Zhang, M. (1998a) The second generation of the California urban futures model. Part 1: Model logic and theory. *Environment and Planning B: Planning and Design*, **25**, 657–666.

Landis, J.D. & Zhang, M. (1998b) The second generation of the California urban futures model. Part 2: Specification and calibration results of the land-use change submodel. *Environment and Planning B: Planning and Design*, **25**, 795–824.

Lane, M. (2010) The carrying capacity imperative: Assessing regional carrying capacity methodologies for sustainable land-use planning. *Land Use Policy*, **27**, 1038–1045.

Launhardt, W. (1882) Die Bestimmung des Zweckmässigsten Standortes einer gewerblichen Anlage. *Zeitschrift des Vereins Deutscher Ingenieure*, **26**, 106–115.

Levitt, T. (1965) Exploit the product life cycle. *Harvard Business Review*, **18**, 81–94.

Lösch, A. (1940) *Die Räumliche Ordnung Der Wirtschaft. Eine Untersuchung Über Standort, Wirtschaftsgebiete Und Internationalen Handel*. Gustav Fischer, Jena.

Mackett, R.L. (1983) The Leeds Integrated Transport Model (LILT), Supplementary Report 805. Transport and Road Research Laboratory, Crowthorne.

Mackett, R.L. (1985) Integrated land use-transport models. *Transportation Reviews*, **5**, 325–343.

Maoh, H.F. & Kanaroglou, P.S. (2005) Agent-based firmographic models: a simulation framework for the city of Hamilton. In *Proceedings of the PROCESSUS Second International Colloquium on the Behavioural Foundations of Integrated Land-use and Transportation Models: Frameworks, Models and Applications*. Toronto.

Maoh, H.F. & Kanaroglou, P.S. (2007a) Business establishment mobility behavior in urban areas: a microanalytical model for the City of Hamilton in Ontario, Canada. *Journal of Geographical Systems*, **9**, 229–252.

Maoh, H.F. & Kanaroglou, P.S. (2007b) Modeling the failure of small and medium size business establishments in urban areas: an application to Hamilton, Ontario. Centre for Spatial Analysis, McMaster University, Hamilton.

Maoh, H.F. & Kanaroglou, P.S. (2009) Intrametropolitan Location of Business Establishments. *Transportation Research Record: Journal of the Transportation Research Board*, **2133**, 33–45.

Maoh, H.F., Kanaroglou, P.S., & Buliung, R.N. (2002) Modelling the location of firms within an Integrated Transport and Land-Use Model for Hamilton, Ontario. In *International Colloquium on the Behavioural Foundations of Integrated Land-Use and Transportation Models*. Quebec City.

Maoh, H.F., Kanaroglou, P.S., & Buliung, R.N. (2005) Modelling the location of firms within an Integrated Transport and Land-Use Model for Hamilton, Ontario. Centre for Spatial Analysis, McMaster University, Hamilton.

Martin, R. & Sunley, P. (2007) Complexity thinking and evolutionary economic geography. *Journal of Economic Geography*, **7**, 573–601.

Martinez, F.J. (1997) MUSSA: A land Use Model for Santiago City. University of Chile, Department of Civil Engineering.

Mata, J. & Portugal, P. (1994) Life Duration of New Firms. *The Journal of Industrial Economics*, **42**, 227–245.

McFadden, D. (1974) Conditional Logit Analysis and Subjective Probability. In Zarembda, P. (ed), *Frontiers in Econometrics*. Academic Press, New York, pp. 105–142.

Meeusen, W. & van den Broeck, J. (1977) Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*, **18**, 435–444.

Melo, P.C., Graham, D.J., & Noland, R.B. (2008) Firm Formation and Transport Infrastructure: A Study of Portugal. In *Proceedings of the 87th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Miller, E.J., Hunt, J.D., Abraham, J.E., & Salvini, P.A. (2004) Microsimulating urban systems. *Computers, Environment and Urban Systems*, **28**, 9–44.

Miller, E.J. & Salvini, P.A. (1998) The Integrated Land Use, Transportation, Environment (ILUTE) Modelling System: A Framework. In *Proceedings of the 77th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Moeckel, R. (2005) Microsimulation of Firm Location Decisions. In *Proceedings of the 9th. CUPUM - International Conference on Computers in Urban Planning and Urban Management*. London.

Moeckel, R. (2009) Simulation of firms as a planning support system to limit urban sprawl of jobs. *Environment and Planning B: Planning and Design*, **36**, 883–905.

Moeckel, R., Schurmann, C., & Wegener, M. (2002) Microsimulation of urban land use. In *Proceedings of the 42nd European Congress of the Regional Science Association*. Dortmund.

Mueller, D.C. (1972) A Life Cycle Theory of the Firm. *The Journal of Industrial Economics*, **20**, 199–219.

Norušis, M.J. (2008) *SPSS 16.0 Advanced Statistical Procedures Companion*. Prentice Hall Inc., Upper Saddle River.

O'Rand, A.M. & Krecker, M.L. (1990) Concepts of the Life Cycle: Their History, Meanings, and Uses in the Social Sciences. *Annual Review of Sociology*, **16**, 241–262.

Oh, K., Jeong, Y., Lee, D., Lee, W., & Choi, J. (2005) Determining development density using the Urban Carrying Capacity Assessment System. *Landscape and Urban Planning*, **73**, 1–15.

van Oort, F.G. & Atzema, O.A.L.C. (2004) On the conceptualization of agglomeration economies: The case of new firm formation in the Dutch ICT sector. *The Annals of Regional Science*, **38**, 263–290.

Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Francisco, CA.

Pellenbarg, P.H. & van Steen, P.J.M. (2003) Spatial Perspectives on Firm Dynamics in The Netherlands. *Tijdschrift voor Economische en Sociale Geografie*, Tijdschrift voor Economische en Sociale Geografie, **94**, 620–630.

Pellenbarg, P.H., van Wissen, L.J.G., & van Dijk, J. (2002) Firm Relocation State of the Art and Research Prospects. University of Groningen, Groningen.

Ponds, R., van Oort, F.G., & Frenken, K. (2010) Innovation, spillovers and university-industry collaboration: an extended knowledge production function approach. *Journal of Economic Geography*, **10**, 231–255.

Pred, A.R. (1967) *Behavior and Location : Foundations for a Geographic and Dynamic Location Theory, Part 1*. The Royal University of Lund, Lund.

Pred, A.R. (1969) *Behavior and Location : Foundations for a Geographic and Dynamic Location Theory, Part 2*. The Royal University of Lund, Lund.

Putman, S.H. (1983) *Integrated Urban Models*. Pion Limited, London, England.

Raspe, O. & van Oort, F.G. (2011) Growth of new firms and spatially bounded knowledge externalities. *The Annals of Regional Science*, **46**, 495–518.

Reynolds, P.D. (1997) Who Starts New Firms? – Preliminary Explorations of Firms-in-Gestation. *Small Business Economics*, **9**, 449–462.

Reynolds, P.D., Storey, D.J., & Westhead, P. (1994) Cross-national Comparisons of the Variation in New Firm Formation Rates. *Regional Studies*, **28**, 443–456.

Rietveld, P. (1994) Spatial economic impacts of transport infrastructure supply. *Transportation Research Part A: Policy and Practice*, **28**, 329–341.

Saarloos, D.J.M., Arentze, T.A., Borgers, A.W.J., & Timmermans, H.J.P. (2005) A multiagent model for alternative plan generation. *Environment and Planning B: Planning and Design*, **32**, 505–522.

Saarloos, D.J.M., Arentze, T.A., Borgers, A.W.J., & Timmermans, H.J.P. (2008) A multi-agent paradigm as structuring principle for planning support systems. *Computers, Environment and Urban Systems*, **32**, 29–40.

Salvini, P.A. & Miller, E.J. (2005) ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems. *Networks and Spatial Economics*, **5**, 217–234.

Sandholm, T.W. (1999) Distributed rational decision making. In Weiss, G. (ed), *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press, Cambridge, pp. 201–258.

Schirmer, P., Zöllig, C., Müller, K., Bodenmann, B.R., & Axhausen, K.W. (2011) The Zurich Case Study of UrbanSim. In *51st European Congress of the Regional Science Association*. Barcelona.

Simmonds, D.C. (2005) Land-use modelling with DELTA: update and experience. In *Proceedings of the 9th. CUPUM - International Conference on Computers in Urban Planning and Urban Management*. London.

Sivakumar, A. & Bhat, C.R. (2007) A Comprehensive, Unified, Framework for Analyzing Spatial Location Choice. In *Proceedings of the 86th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Smith, P.J. (2002) *Analysis of Failure and Survival Data*. Chapman & Hall, USA.

Spiegelhalter, D.J., Philip, A., Lauritzen, S.L., & Cowell, R.G. (1993) Bayesian Analysis in Expert Systems. *Statistical Science*, **8**, 219–283.

Stanback, T.M. & Knight, R.V. (1970) *The Metropolitan Economy; the Process of Employment Expansion*. Columbia University Press, New York,.

Storper, M. & Walker, R. (1983) The theory of labour and the theory of location. *International Journal of Urban and Regional Research*, **7**, 1–43.

Taylor, M.J. (1975) Organizational growth, spatial interaction and location decision-making. *Regional Studies*, **9**, 313–323.

Teece, D.J. (1986) Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, **15**, 285–305.

Thill, J.C. & Wheeler, A. (2000) Tree Induction of Spatial Choice Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, **1719**, 250–258.

Timmermans, H.J.P. (2003) The Saga of Integrated Land Use-Transport Modeling: How Many More Dreams Before We Wake Up? In *Proceedings of the 10th International Conference on Travel Behaviour Research*. Lucerne.

Timmermans, H.J.P., Arentze, T.A., & Joh, C. (2002) Analysing space-time behaviour: new approaches to old problems. *Progress in Human Geography*, **26**, 175–190.

Townroe, P.M. (1969) Locational choice and the individual firm. *Regional Studies*, **3**, 15–24.

Townroe, P.M. (1972) Some behavioural considerations in the industrial location decision. *Regional Studies*, **6**, 261–272.

Townroe, P.M. (1975) Approaches to the Study of Industrial Location. In Massey, D.B. & Morrison, W.I. (eds), *Industrial Location: Alternative Frameworks*. Centre for Environmental Studies, London, pp. 32–40.

Townroe, P.M. (1991) Rationality in Industrial Location Decisions. *Urban Studies*, **28**, 383–392.

Tsoularis, A. & Wallace, J. (2002) Analysis of logistic growth models. *Mathematical Biosciences*, **179**, 21–55.

Veldhuisen, J., Timmermans, H.J.P., & Kapoen, L. (2000) RAMBLAS: a regional planning model based on the microsimulation of daily activity travel patterns. *Environment and Planning A*, **32**, 427–443.

Verhoeven, M., Arentze, T.A., Timmermans, H.J.P., & van der Waerden, P. (2006) Modeling the Influence of Structural Lifecycle Events on Activity-Travel Decisions Using a Structure Learning Algorithm. In *11th International Conference on Travel Behaviour Research*. Kyoto.

Verhulst, P.F. (1838) Notice sur la loi que la population poursuit dans son accroissement. *Correspondance mathématique et physique*, **10**, 113–121.

Vernon, R. (1966) International Investment and International Trade in the Product Cycle. *The Quarterly Journal of Economics*, **80**, 190–207.

Vernon, R. (1979) The Product Cycle Hypothesis in a New International Environment. *Oxford Bulletin of Economics and Statistics*, **41**, 255–267.

Vittinghoff, E., Glidden, D., Shiboski, S., & McCulloch, C.E. (2005) *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*, 1st edn. Springer, USA.

Waddell, P. (2000) A behavioral simulation model for metropolitan policy analysis and planning: residential location and housing market components of UrbanSim. *Environment and Planning B: Planning and Design*, **27**, 247–263.

Waddell, P. (2002) Modeling Urban Development for Land Use, Transportation, and Environmental Planning. *Journal of the American Planning Association*, **68**, 297–314.

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M., & Ulfarsson, G.F. (2003) Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. *Networks and Spatial Economics*, **3**, 43–67.

Waddell, P. & Ulfarsson, G.F. (2003) Accessibility and Agglomeration: Discrete-Choice Models of Employment Location by Industry Sector. In *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*. Washington, D.C.

Waddell, P., Ulfarsson, G.F., Franklin, J.P., & Lobb, J. (2007) Incorporating land use in metropolitan transportation planning. *Transportation Research Part A: Policy and Practice*, **41**, 382–410.

Wagner, P. & Wegener, M. (2007) Urban Land Use, Transport and Environment Models: Experiences with an Integrated Microscopic Approach. *DISP*, **170**, 45–56.

Walker, R. & Storper, M. (1981) Capital and Industrial Location. *Progress in Human Geography*, **5**, 473–509.

Weber, A. (1909) *Über Den Standort Der Industrien. Teil I: Reine Theorie Des Standorts*. Verlag J.C.B. Mohr, Tübingen.

Wegener, M. (1982a) A multilevel economic-demographic model for the Dortmund region. *Sistemi Urbani*, **3**, 371–401.

Wegener, M. (1982b) Modeling Urban Decline: A Multilevel Economic-Demographic Model for the Dortmund Region. *International Regional Science Review*, **7**, 217–241.

Wegener, M. (1983) Description of the Dortmund Region Model, Working Paper 8. Institut für Raumplanung, Dortmund.

Wenting, R., Atzema, O.A.L.C., & Frenken, K. (2011) Urban Amenities and Agglomeration Economies? The Locational Behaviour and Economic Success of Dutch Fashion Design Entrepreneurs. *Urban Studies*, **48**, 1333–1352.

Wets, G., Vanhoof, K., Arentze, T.A., & Timmermans, H.J.P. (2000) Identifying Decision Structures Underlying Activity Patterns: An Exploration of Data Mining Algorithms. *Transportation Research Record: Journal of the Transportation Research Board*, **1718**, 1–9.

Wever, E. (1984) *Nieuwe Bedrijven in Nederland*, Serie Mens en ruimte. Van Gorcum.

White, R. & Engelen, G. (1993a) Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A*, **25**, 1175–1199.

White, R. & Engelen, G. (1993b) Cellular dynamics and GIS: modelling spatial complexity. *Geographical Systems*, **1**, 237–253.

van Wissen, L.J.G. (1997) Demography of the firm. Modelling birth and death of firms using the concept of carrying capacity. In van den Brekel, H. & Deven, F. (eds), *Population and Families in the Low Countries 1996/1997*, *NIDI/CBGS*. The Hague/Brussels.

van Wissen, L.J.G. (2000) A micro-simulation model of firms: Applications of concepts of the demography of the firm. *Papers in Regional Science*, **79**, 111–134.

van Wissen, L.J.G. (2002) Demography of the Firm: A Useful Metaphor? *European Journal of Population/Revue européenne de Démographie*, **18**, 263–279.

van Wissen, L.J.G. (2003) Modelling Regional Economic Growth by Means of Carrying Capacity. In Stillwell, J. & Clarke, G. (eds), *Applied GIS and Spatial Analysis*. John Wiley & Sons, Ltd, pp. 297–313.

Witlox, F., Borgers, A.W.J., & Timmermans, H.J.P. (2004) Modeling locational decision making of firms using multidimensional fuzzy decision tables: an illustration. *Eletronic Journal of Mathematics*, **15**, 1–17.

Witten, I.H. & Frank, E. (2005) *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufman, Amsterdam.

Wood, P.A. (1978) Industrial organisation, location and planning. *Regional Studies*, **12**, 143–152.

Wu, N. & Silva, E.A. (2010) Artificial Intelligence Solutions for Urban Land Dynamics: A Review. *Journal of Planning Literature*, **24**, 246–265.

Appendix

Summary of the results of the location choice model for firms involved in relocation, based on the BCN model BAN No-FS with T = 2.0.

Variable	**Cat.**	**Office firm type class (rounded up values in %)**														
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
NSR areas	N	8	13	25	9	11	8	13	17	13	11	9	11	16	8	10
	R	67	59	47	68	56	73	58	60	56	57	64	56	51	73	69
	S	25	28	28	23	33	20	29	23	30	32	27	32	33	19	21
Province	A	2	1	4	1	1	1	2	2	2	1	1	1	4	1	2
	B	3	4	3	6	5	8	4	2	2	3	5	3	3	9	4
	C	2	3	4	4	4	2	3	3	4	3	3	2	2	1	2
	D	5	10	10	7	11	7	9	6	9	10	9	12	12	7	8
	E	2	4	2	3	2	2	3	5	3	2	2	4	5	5	3
	F	4	3	5	3	3	2	4	5	4	3	3	4	6	2	3
	G	12	13	10	10	17	7	15	9	14	17	14	13	12	8	8
	H	35	36	15	42	40	51	30	30	30	25	27	27	11	40	36
	I	2	4	15	1	4	2	5	7	4	4	3	4	5	1	3
	J	3	8	14	5	5	3	7	11	6	12	14	11	5	6	12
	K	4	2	2	4	3	5	2	4	3	1	1	3	3	2	2
	L	25	12	15	15	6	10	17	17	18	17	18	15	32	17	16
Urbanization (municipality level)	A	20	23	20	18	16	22	18	22	19	22	24	27	26	29	31
	B	18	33	29	27	34	32	32	29	31	34	37	29	32	27	33
	C	23	19	23	25	21	17	21	19	21	21	19	19	18	20	17
	D	21	16	14	18	19	16	19	15	18	16	14	15	14	14	12
	E	17	9	14	13	11	13	10	16	11	7	6	10	10	10	7

Variable	Cat.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
		Office firm type class (rounded up values in %)														
Urbanization (postcode level)	A	20	22	20	19	19	21	20	20	20	22	23	21	22	24	26
	B	19	18	20	19	18	19	19	20	18	17	19	19	20	19	19
	C	19	16	19	18	15	19	14	19	14	12	15	18	17	18	16
	D	20	18	20	20	17	19	18	19	17	16	17	19	18	19	18
	E	22	26	22	24	31	22	30	22	30	33	26	23	23	20	21
Firm size	A	20	25	7	26	33	25	20	18	33	37	37	32	3	25	33
	B	64	60	25	59	61	60	69	48	59	54	52	46	19	54	54
	C	9	12	16	13	5	10	9	20	6	8	9	16	35	16	10
	D	4	1	18	2	0	2	2	6	1	1	1	3	17	3	2
	E	3	2	33	0	0	3	1	8	0	1	1	3	27	1	1
Reloc. dist. 1	A	10	17	21	12	14	12	21	22	21	16	13	16	29	20	21
	B	29	23	22	19	20	21	25	22	25	24	22	25	32	24	23
	C	9	19	21	17	20	23	18	21	19	21	22	21	18	23	22
	D	21	15	22	14	14	15	13	17	13	14	15	14	8	14	14
	E	31	26	13	38	32	29	23	19	24	25	28	24	13	18	21
Reloc. dist. 2	A	55	74	70	59	67	66	77	76	76	75	72	73	86	78	78
	B	19	13	10	14	18	18	14	15	14	14	16	14	7	11	12
	C	13	8	10	14	10	11	6	4	7	7	8	8	5	6	6
	D	14	4	9	13	5	5	3	5	3	3	4	5	2	5	3
Reloc. dist. 3	A	66	74	84	62	68	70	77	80	76	75	72	76	87	81	79
	Z	34	26	16	38	32	30	23	20	24	25	28	24	13	19	21
Density of public transportation	A	33	25	19	28	28	22	24	20	25	26	24	26	21	20	20
	B	16	17	27	20	17	14	18	16	19	17	17	17	18	16	14
	C	29	21	15	24	20	26	21	18	19	19	21	19	20	22	20
	D	19	19	21	17	18	20	19	21	19	18	19	13	17	19	20
	E	4	18	18	11	16	19	18	25	19	19	19	25	24	23	25

Variable	Cat.	Office firm type class (rounded up values in %)														
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Dist. to Schiphol	A	19	12	19	16	12	18	9	18	10	7	9	15	12	18	12
	B	21	26	20	25	27	26	25	21	23	27	28	24	24	27	30
	C	20	19	20	19	14	18	18	19	18	20	21	19	17	18	19
	D	19	16	20	18	17	17	17	18	17	16	16	17	19	17	15
	E	21	28	21	21	30	20	30	23	32	30	26	25	28	20	23
Dist. to closest airport	A	20	18	20	19	17	19	17	20	17	14	17	19	19	19	18
	B	20	22	20	21	23	21	24	20	23	25	23	21	21	22	24
	C	20	21	20	21	22	20	23	20	22	26	23	20	20	20	21
	D	20	20	20	20	19	20	19	20	19	19	19	20	20	20	19
	E	20	19	20	19	18	19	18	20	18	16	17	20	20	19	19
Dist. to IC train	A	10	11	15	6	9	10	11	13	10	11	11	13	19	10	15
	B	13	26	24	19	21	27	21	21	26	25	26	29	26	38	33
	C	33	32	41	38	32	29	33	33	29	32	34	34	28	29	25
	D	43	31	20	37	38	34	35	33	35	31	29	25	27	23	27
Dist. to international train	A	20	20	20	20	20	20	20	20	20	19	20	20	20	20	21
	B	20	20	20	20	20	20	20	20	20	19	20	20	20	20	20
	C	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
	D	20	20	20	20	20	20	20	20	20	19	20	20	20	20	20
	E	20	20	20	20	21	20	21	20	21	22	20	20	20	20	20
Dist. to highway junction	A	13	9	15	8	9	7	10	13	10	9	8	9	11	5	8
	B	10	20	12	20	21	22	20	16	20	18	19	22	18	18	16
	C	31	30	24	33	30	30	30	33	29	31	31	30	30	30	29
	D	27	27	40	26	28	28	27	29	29	30	29	26	29	29	32
	E	19	14	9	12	12	13	12	9	12	12	13	14	13	17	15
Dist. to shopping centers	A	19	14	19	17	12	18	14	19	14	10	10	18	18	18	15
	B	19	19	20	18	16	19	19	20	18	18	18	19	21	19	21
	C	20	25	20	21	25	21	25	21	25	32	31	23	21	22	26
	D	20	20	21	21	21	21	19	19	19	21	23	20	19	21	20
	E	22	21	20	23	24	21	23	21	23	19	18	20	20	19	18

Variable	Cat.	Office firm type class (rounded up values in %)														
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Population	A	16	21	42	22	22	27	20	20	20	19	19	24	22	15	16
	B	35	24	17	28	24	22	23	28	25	24	22	21	25	25	21
	C	12	21	14	25	26	17	26	25	27	26	24	24	28	22	28
	D	38	34	27	25	29	33	32	27	28	32	35	32	25	37	35
Num. of households	A	20	19	21	20	19	20	17	20	18	16	17	20	19	19	17
	B	21	21	20	21	22	20	22	20	22	23	21	20	20	20	19
	C	20	21	20	20	24	20	25	20	25	28	25	20	21	20	22
	D	20	20	20	19	19	20	20	20	19	20	20	20	20	20	21
	E	20	19	20	19	16	20	16	20	16	14	18	20	19	21	20
Labor force	A	21	23	43	25	25	30	23	22	23	21	20	26	22	17	17
	B	32	25	17	30	26	23	24	33	28	26	24	21	32	26	24
	C	15	21	16	22	25	16	25	22	25	25	24	25	23	24	26
	D	32	31	24	22	24	31	28	24	25	28	31	28	24	34	33
Average income	A	18	14	18	15	10	19	10	17	10	10	13	18	17	21	17
	B	19	20	20	19	18	21	17	19	18	20	21	21	24	21	23
	C	21	22	20	20	22	23	22	22	22	20	20	20	20	19	19
	D	20	21	20	23	24	18	24	20	22	23	22	20	20	18	18
	E	22	23	23	24	26	20	27	22	28	28	24	21	18	21	23
Num. of places in schools	A	18	17	22	19	17	24	14	16	15	15	15	15	16	13	12
	B	33	27	24	36	31	25	27	25	28	27	25	23	21	25	24
	C	24	28	18	23	25	25	28	24	26	27	28	27	23	26	29
	D	12	16	17	13	15	15	18	16	17	18	18	15	22	20	19
	E	13	13	18	9	11	11	14	18	14	14	14	20	18	16	17
Num. of parking places	A	25	24	26	27	26	26	23	25	23	23	22	26	24	23	21
	B	25	26	25	26	26	25	25	25	27	27	26	25	25	26	26
	C	25	27	25	25	27	25	28	25	27	30	30	26	26	26	28
	D	25	24	25	23	21	25	23	25	23	20	22	24	25	25	25

Variable	Cat.	Office firm type class (rounded up values in %)														
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Shopping centers 1 km	A	30	48	29	41	54	37	51	31	51	55	53	39	38	37	45
	B	23	23	24	23	23	22	23	24	22	24	25	23	20	23	23
	C	24	13	24	17	13	20	13	22	12	11	12	18	18	19	14
	D	23	16	24	18	10	21	13	23	14	11	11	19	24	21	18
Agglomeration economies 1 km	A	20	20	20	21	21	20	20	20	20	18	18	20	20	20	19
	B	20	20	20	20	20	20	21	20	20	22	21	20	20	20	19
	C	20	20	20	20	21	20	22	20	21	25	23	20	20	20	20
	D	20	20	20	20	19	20	19	20	20	20	20	20	20	20	20
	E	20	20	20	19	18	20	18	20	18	16	18	20	20	20	22
Rent price / m^2	A	19	15	19	18	13	18	11	18	12	9	10	15	13	16	10
	B	24	31	21	30	33	24	34	23	33	32	32	26	29	24	27
	C	21	20	21	19	20	20	20	19	20	22	23	18	22	19	19
	D	18	14	21	17	13	17	12	17	12	15	14	17	16	17	16
	E	19	21	19	17	21	22	22	23	22	21	21	24	19	23	27

Author index

I

J

K

L

X

Z

Subject index

D

E

F

G

H

I

N

O

P

Q

R

S

T

U

W

Z

Summary

Modeling Office Firm Dynamics in an Agent-Based Micro Simulation Framework: Methods and Empirical Analyses

The focus of this research project was on investigating office firm dynamics. Nowadays, a large share of economic activities is represented by office firms, especially in the sector of professional services. Increasing urbanization and specialization of employment were processes observed in the last decades, which have been contributing to a shift in the bases of the economy to the service sector. Consequently, an intense activity among people's behavior, services and infrastructure systems is observed in cities. This generates land use and travel patterns that influence especially the transportation infrastructure systems. From an urban and transportation planning perspective, the proper examination of these interactions constitutes a major concern to establish policies of land use and organize measures to plan, build and manage infrastructure systems. In line with this, LUTI models have been traditionally employed to examine these interactions.

The contribution aimed at this research was modeling office firm dynamics within a LUTI framework. More specifically, it was based on a firm demography approach that analyzes the events of firm start-up, growth, decline, and closure as well as location choice and relocation decision patterns. The current trends in LUTI models, which are based on agent-based and micro simulation models, have found strong linkages with the firm demographic approach, capable to accommodate behaviorally richer concepts of firm dynamics. While several applications dealing with firms in general can be found, the specific interest aimed here was in the behavior of office firms. This comes along with the influences underlain by the urban environment. In LUTI modeling, the specification in detail of the factors related to urban processes is required in order to obtain as much realistic models as possible. This was explored here, by including and analyzing a large set of location variables as well as trying to define them in a high level of detail, using nationwide data from The Netherlands.

The approach taken to investigate office firm dynamics was based on a set of econometric/statistical models that were employed to examine the related demographic events as a function of both firm internal attributes and location characteristics. The start-up model was based on the idea that firms must exist before it can be simulated and it can only be modeled by taking an area as a decision unit. In fact, the model does

not deal with the actual start-up behavior, but simulates an environment where firms come into existence every year as a result of certain urban conditions. Therefore, the frequency of firms starting in these areas is modeled using the Poisson regression model. The main factors influencing firm start-up include effects of agglomeration economies, high population levels, and some accessibility to transportation infrastructure such as high density of public transportation.

Given that the firm dynamics investigated here operate in an (urban) environment, it is expected that some constraints apply as a result of existing available resources and market demands. Therefore, the notion of carrying capacity was introduced to control the firm demographic processes in relation to the urban conditions. A stochastic frontier model was investigated and its production form was associated to the carrying capacity notion. The approach considered that the carrying capacity of a region is measured in terms of a maximum number of jobs that the region can support, based on the size of the population in the region. Population imposes a maximum to the production either as a resource of labor (in case of basic industries) or as limiting the demand for products (in case of non-basic industries). However, the stochastic frontier model also presents a cost form and this form fitted best most office firm types according to the data used. In fact, the case is that either a production or a cost function usually fits a data distributional structure. Therefore, in cases where a cost function rather than a production function fitted the data, the conclusion was that the notion of carrying capacity is not supported by the data. Such a finding indicates that the population defines a minimum rather than a maximum frontier. As a source of demand or supplier of labor, population is not a limiting factor for the sector's size: the sector could grow further even if the population stays constant.

As a critical part of firm demographic models, the spatial distribution of economic activities was investigated by an approach that views the location choice decision as a matching process of, on the one hand, a set of firm requirements and, on the other, a set of location characteristics. More specifically, office firm types (the class labels or dependent variable) are matched with the location attributes (the independent variables) using a classification task approach based on Bayesian classifier networks. In fact, the model proposed here does not actually describe the choice of discrete locations, but the probability of having certain location characteristics related to each office firm type. In order to find which (concrete) locations firms would be choosing to locate their business, a spatial component is needed. If those selected location characteristics could be matched, for example, with a geographical database of available locations, such location decision mechanism could be derived and an allocation process between firm and location could be established.

Following the course of lifecycle, firms will grow and decline. The approach developed here to investigate this process is derived from the well-known logistic growth function. In fact, it consists of an approximation based on a quadratic function, in which growth can be described as a function of firm size. Both quadratic and linear firm size terms are included in the right-hand side of the equation to make sure that growth is an increasing function for initial values of current size, followed by a decreasing function for larger values of current size, which can accommodate the firm decline process. The examination of external factors influencing firm growth/decline is dealt with the inclusion of location attributes as interaction terms of firm size. As opposed to the logistic growth function, this approximation turned out to be a more flexible approach, capable to accommodate different sets of firm size-development patterns as well as to account for external factors on the firm growth/decline process. The findings suggested that location attributes do play a role and they should be considered along with firm internal characteristics when investigating firm growth/decline processes in the perspective of LUTI models.

Upon establishing a business, firms may consider relocating during its lifecycle. Although they expect to minimize this event as much as possible, due to large costs involved and risks of losing established markets, a relocation decision is often associated with an improvement in the quality of the location in order to maintain the firm's best level of service. The assumption is that both firm internal characteristics and external (location) factors play a role in this decision. A modeling approach to examine the determinants of this decision is rather straightforward as a binary response model can be easily applied. However, the longitudinal nature of the data available here allowed the exploration of duration models. They consider not only whether or not a firm decides to relocate, but also take into account the elapsed time before this decision. Among the push factors (e.g., the factors that motivate a relocation decision) investigated here, firm internal characteristics such as firm size and whether a firm was involved in a relocation before suggest, respectively, that larger firms tend to avoid relocating and firms that relocated before tend to avoid relocating again. The location attributes also indicated significant influences, and some of them were in line with previous studies. Overall, access to transportation infrastructure, effects of agglomeration economies, socioeconomic aspects, urbanization levels and regional characteristics seem to contribute to the firm's decision to relocate.

Once a firm decides to relocate, a process of finding a new location begins. Usually, there is not much difference in terms of modeling approaches compared to what is done for the location choice process when a firm starts-up. Thus, the search process can be seen as a location choice problem. The same modeling approach employed for the location choice based on the Bayesian classifier networks were applied to examine the pull factors (e.g., the factors attracting firms to a specific location). To a large extent, the same spatial tendencies compared to a first location decision are expected. However, there is an important and new factor here: the distance over which firms relocated. It regards an influential factor in the relocation process that should be taken into account. The findings obtained here suggest that relocations are over short distances, up to 5 km, indicating that office firms may relocate to adjust minor deficiencies in their current location, but they would still remain in the same area where they firstly were established.

At the end of a lifecycle, firms will go out of business either because of bankruptcy or voluntary exit. In addition to these, the determinants of firms' closing patterns are related to firm internal characteristics, such as: age and size. But location characteristics may also influence the firm closure process. The focus of the closure model was on the investigation of the influence of these external factors on firm closure. Duration models were also applied. Similar to the relocation decision event, given the longitudinal data available, the model is capable of not only taking into account whether a firm is closed, but also to consider the duration that a firm existed before going out of business. The findings suggest that firm closure is highly influenced by location attributes, as showed by most of the variables included. This is also in line with the aim of the analysis proposed, suggesting an evident relationship between firm closure and location attributes. The analysis carried out here brought a better understanding about the firm closure process by using an extended set of location attributes.

Overall, from a LUTI modeling perspective, the findings obtained across the several location attributes considered brought a better understanding in relation to the influences of these attributes on firm dynamics, more specifically, on the behavior of office firms. Compared to earlier efforts found in the literature, the set of location attributes appears to be more comprehensive and some factors that were not explored in detail before could be examined here. This comes along with the development of the analyses using nationwide data for The Netherlands. Therefore, one of the objectives of this research project could be fulfilled. The examination of office firm dynamics using behaviorally richer concepts of firm demography along with the current trends in LUTI models, which is based on agent-based modeling, is also another achievement. The methods and empirical analyses carried out introduce the keystones for a modeling framework to simulate office firm dynamics.

Curriculum Vitae

Gustavo Garcia Manzato was born on the 8th of September, 1980 in Bauru, Brazil. He studied civil engineering at the Faculty of Engineering of Bauru, São Paulo State University, Brazil. During his studies, he came into contact with geographical information systems (GIS) applied to urban and transportation modeling. This experience led to his graduation project, on the development and analysis of a digital elevation model (DEM) using vector and raster frameworks. During this period, he also participated in teaching assistant (TA) activities. After getting his degree in 2004, he started a Master's in Planning and Operation of Transportation Systems at the Department of Transportation of the School of Engineering of São Carlos, University of São Paulo, Brazil, funded by a grant obtained from the Brazilian government. His research topic was on modeling the interaction of functional urban regions and transportation supply using spatial statistics and spatial modeling techniques along with GIS. He was also involved in some extra activities during this period: he participated again in TA projects; had some experience with distance education, where he coordinated two undergraduate civil engineering courses in a virtual environment and gave support to teachers with the implementation of other courses in an online environment; took part in an origin-destination travel survey carried out in the city of São Carlos, Brazil, in 2007; was a representative of post-graduate students of the Department of Transportation of the School of Engineering of São Carlos, University of São Paulo, Brazil; and participated in the organization of two international conferences held in Brazil. After obtaining his Master's degree in 2007, Gustavo started his PhD at the Urban Planning Group of the Faculty of the Built Environment, Eindhoven University of Technology, The Netherlands. He obtained a special grant from the European Union, named AlBan Program, which provided funding for Latin American students to carry out their graduate studies in Europe. Along with this, he also got another funding from the "*Ruimte voor Geo-Informatie*" (Space for Geo-Information, RGI) program to support his research project. His research topic was on developing a framework for modeling office firm dynamics using agent-based and micro simulation concepts. During his PhD, he received an award of the best article of the International Journal of Urban Sciences in 2010. He was also a board member of the PhD Network Bouwkunde for two years.

Publication list

Journal papers

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2011) Exploration of Location Influences on Firm Survival Rates with Parametric Duration Models. *Transportation Research Record: Journal of the Transportation Research Board*, **2245**, 124–130.

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2011) Matching office firms types and location characteristics: An exploratory analysis using Bayesian classifier networks. *Expert Systems with Applications*, **38**, 9665–9673.

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2010) Location and Accessibility Mediated Influences on Office Firm Closure Rates: A Proportional Hazard Model. *International Journal of Urban Sciences*, **14**, 1–15.

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2010) A support system that delineates location-choice sets for firms seeking office space. *Applied GIS*, **6**, 1–17.

Conference papers

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2011) Exploring location influences on firm survival rates using parametric duration models. In *90th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2011) Spatial attributes mediating regional carrying capacity for office firm sectors: a stochastic frontier approach. In *90th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2010) Matching office firms types and location characteristics: an exploratory analysis using Bayesian classifier networks. In *89th Annual Meeting of the Transportation Research Board*. Washington, D.C.

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2010) Location and accessibility mediated influences on office firm closure rates: a proportional hazard model. In *12th World Conference on Transport Research*. Lisbon.

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2010) Exploring location influences on firm survival rates using parametric duration models. In *10th International Conference on Design and Decision Support Systems in Architecture and Urban Planning*. Eindhoven.

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2009) A support system for delineating location choice sets of a firm seeking office space. In *European Real Estate Society Conference*. Stockholm.

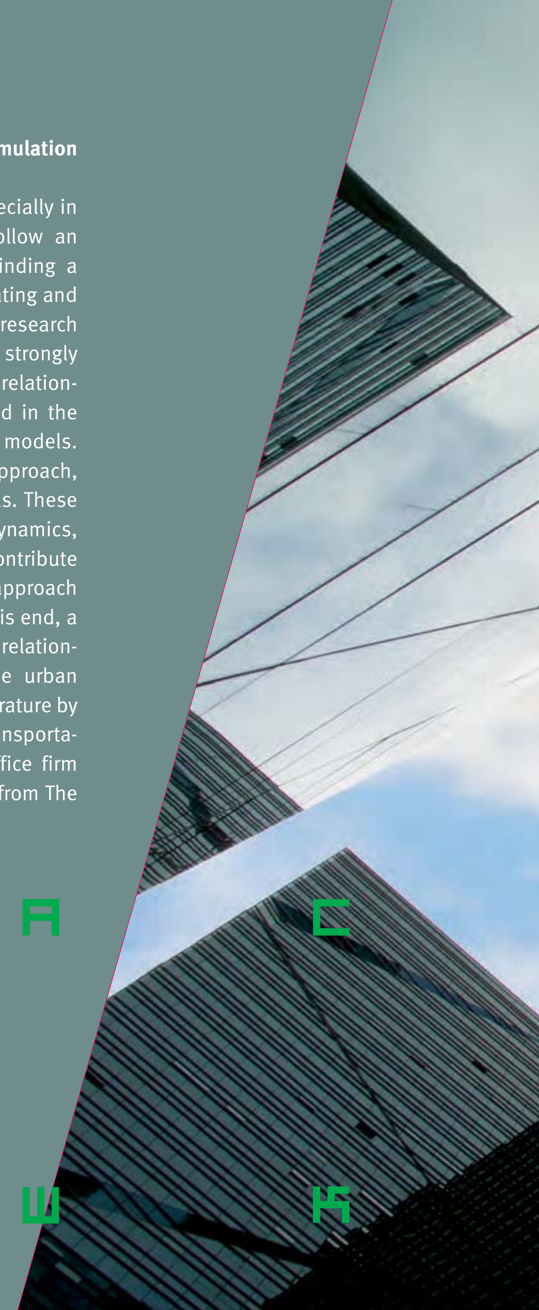
Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2009) A geo-informatics support system for location decisions of firms. In *11th International Conference on Computers in Urban Planning and Urban Management*. Hong Kong.

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2009) A geo-informatics support system for location decisions of office firms. In *3rd AlBan Conference*. Porto.

Manzato, G.G., Arentze, T.A., Timmermans, H.J.P., & Ettema, D.F. (2008) An agent for supporting and simulating location decisions of firms. In *9th International Conference on Design and Decision Support Systems in Architecture and Urban Planning*. Leende.

Modeling Office Firm Dynamics in an Agent-Based Micro Simulation Framework

Office firms represent a large share of economic activities, especially in the sector of professional services. In general, firms will follow an evolutionary cycle comprising the dynamics of starting-up, finding a location to establish their business, growing or declining, relocating and going out of business. The underlying approach taken in this research project relies on the idea that the evolution of office firms is strongly influenced by the urban environment. Traditionally, the specific relationship between transportation and land use has been examined in the framework of aggregate integrated land use-transportation (LUTI) models. However, the field is moving toward a more disaggregate approach, based on concepts of micro simulation and agent-based models. These are built on behaviorally richer concepts for examining firm dynamics, such as firm demography. The aim of this research project is to contribute to this emerging field by developing an agent-based modeling approach to simulate the evolution of office firms in time and space. To this end, a set of statistical/econometric models is used to investigate the relationships between specific firm demographic processes and the urban environment. The research project contributes to the existing literature by focusing on office firm demography and related land use and transportation influences, exploring alternative approaches to model office firm dynamics empirically, and using very detailed nationwide data from The Netherlands.



/ Department of the Built Environment

Technische Universiteit
Eindhoven
University of Technology