DISS. ETH NO. 18911

APPLICATION OF SPATIAL ANALYSIS METHODS FOR UNDERSTANDING GEOGRAPHIC VARIATION OF PRICES, DEMAND AND MARKET SUCCESS

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
MICHAEL LÖCHL

Dipl.-Ing., University of Dortmund
born 29.07.1975
citizen of Germany

accepted on the recommendation of
Prof. Kay W. Axhausen, examiner
Prof. Darren M. Scott, co-examiner

2010

Acknowledgements

First of all, I would like to thank my examiner Professor Kay W. Axhausen from ETH Zürich for all his advise, inspiration and support. Moreover, I would like to thank my co-examiner Professor Darren M. Scott from McMaster University in Hamilton, who gave valuable input at various stages of the dissertation.

Additionally, there are a range of former colleagues and visitors at IVT who supported the disseration in various ways, including discussion, collaboration, data preparation and proof-reading: Michael Balmer, Michaela Bürgle, Francesco Ciari, Alexander Erath, Philipp Fröhlich, Jeremy Hackney, Konrad Meister, Michael van Eggermond and Charlotte Whitehead. Over the years, the group of Prof. Axhausen has changed constantly, but there was always a stimulating athmosphere of companionship and helpfulness and I am thankful to have been a part of it for a while.

Content

List of Figures

List of Tables

Abstract

Spatial analysis is a general term to describe mathematical methods that use locational information in order to better understand processes generating observed attribute values (Fotheringham and Rogerson, 2009, 1). Such techniques are applied in many fields, including biology, epidemiology, ethnology, geography, sociology and statistics. Nevertheless, the nature of the spatial variation of interest is sometimes not well understood and the patterns of spatial dependence and heterogeneity are disregarded. Therefore, this dissertation brings together different applications of spatial analysis methods for understanding geographic variation of certain entities while considering those spatial effects. Additionally, certain spatial analysis, modelling and simulation techniques are examined for solving location problems and informing spatial allocation and deployment of resources. Namely, regression techniques, integrated land use and transport simulation and agent-based modelling approaches are applied in four examples, whereas all those techniques are grounded in spatial analysis.

In the first application, cross-sectional business data of a cooperating major insurance company are used and merged with other available spatial data at the municipal level in Switzerland to explain geographically varying market shares of its tied agent channel within the country. The effects of distribution intensity by the tied agents and channel competition, among others are considered in the proposed regression models, which also identify geographic areas of over- and underperformance by spatial residual analysis. However, many studies have highlighted that ordinary least squares (OLS) regression approaches lack the ability to consider spatial dependency and spatial heterogeneity, leading to biased and inefficient estimations. Consequently, spatial simultaneous autoregressive models are applied.

In the second application, the hedonic approach is used to understand and explain geographic variation of asking rents in Canton Zürich for an integrated land use and transport simulation. Here, the aforementioned problem of spatial effects is considered by applying and comparing OLS, spatial simultaneous autoregressive and geographically weighted regression (GWR) techniques. The latter one performs best with regard to model fit, but the issue of spatially correlated coefficients remain a problem. Therefore, a spatial simultaneous autoregressive model is proposed for further use. Beyond, the work reveals that when housing markets are a particular concern in integrated land use and transport simulation, significant efforts are needed for the price data generation and modelling.

The implementation of a land use simulation for the Greater Zürich area is the content of the third application. Such a simulation might be a helpful tool for example for forecasting transport investments impacts on land use and development as well as for investigating the interdependencies and for comparing certain policy scenarios. The location choice decisions of households and businesses/jobs are explicitly modelled. However, in the presented work, focus is given to data issues as they constitute a crucial factor in the implementation process. Moreover, preliminary hedonic rent and household location choice estimations are presented, interpreted and compared in detail.

The fourth example involves an agent-based modelling approach. The work is based on a literature review of retail location planning and a report of a series of in-depth interviews, which were conducted with retailers of various product groups in Germany and Switzerland. The interviews revealed that advanced spatial analysis and spatial decision tools are rarely applied in retail location planning practice. An additional literature overview on agent-based descriptions of retail markets highlights potentially suitable approaches of the modelling of retailers. Finally, a conceptual model for a retailer module within an agent-based transport simulation is proposed, and the main issues arising with the introduction of such types of agents are discussed.

Overall, the various applications show not only how spatial methods can help in analysing and explaining geographic variation of entities, but also how the methods and tools can help in certain location planning and resource allocation tasks.

Zusammenfassung

Der Begriff *Raumanalyse* (engl. Spatial Analysis) bezeichnet mathematische Methoden, die räumliche Informationen nutzen. Dies dient dazu jene Prozesse besser zu verstehen, die die beobachteten Merkmalsausprägungen generieren (Fotheringham and Rogerson, 2009, 1). Entsprechende Techniken werden in vielen Fachgebieten angewendet, u.a. in der Biologie, Epidemiologie, Ethnologie, Geografie, Soziologie und der Statistik. Trotzdem werden aber die Eigenschaften der räumlichen Variation zuweilen nicht gut erfasst und die Muster der räumlichen Abhängigkeit und Heterogenität werden ignoriert. Die Dissertation beschreibt verschiedene Anwendungsfelder der Raumanalyse, um die geografische Variation von verschiedenen Untersuchungseinheiten unter Berücksichtigung solcher räumlicher Effekte zu verstehen. Ausserdem werden räumliche Analyse-, Modellierungs- und Simulationstechniken behandelt, um räumliche Verteilungs- und Lokationsprobleme zu lösen. Namentlich werden in vier Beispielen Regressionstechniken, eine integrierte Flächennutzungs- und Verkehrssimulation sowie agentenbasierte Modellierungstechniken angewendet, wobei alle diese Techniken Methoden der Raumanalyse verwenden.

In der ersten Anwendung werden Geschäftsdaten eines grossen Versicherungsunternehmens kombiniert mit anderen räumlichen Daten auf Gemeindebene in der Schweiz, um die räumlich variierenden Marktanteile des Agenten- bzw. Kundenberatervertriebskanals der Versicherung zu erklären. In den vorgeschlagenen Regressionsmodellen werden unter anderem die Effekte der Vertriebsintensität der gebundenen Agenten und der Wettbewerb zwischen den Vertriebskanälen berücksichtigt. Die Modelle erlauben auch die Identifizierung von Gebieten mit über- oder unterdurchschnittlichen Vertriebserfolg durch die räumliche Analyse der Residuen. Zahlreiche Studien haben darauf hingewiesen, dass in einfachen multiplen Regressionen räumliche Abhängigkeiten und Heterogenität nicht berücksichtigt werden. Dies führt zu verzerrten und ineffizienten Schätzungen. Aus diesem Grund werden räumlich-autoregressive Modelle verwendet.

In der zweiten Anwendung wird ein hedonischer Ansatz eingesetzt, um die räumliche Variation von Angebotsmieten im Kanton Zürich zu verstehen und zu erklären. Dabei wird das zuvor erwähnte Problem der räumlichen Effekte gelöst durch die Anwendung und den Vergleich von einfachen, räumlich-autoregressiven (engl. Spatial Simultaneous Autoregressive Models) und räumlich-gewichteten Regressionen (engl. Geographically Weighted Regressions). Der letztgenannte Ansatz hat in der vorgestellten Anwendung die

höchste Erklärungsgüte, aber es besteht ein Problem der räumlich korrelierten Koeffizienten. Deshalb wird auch hier ein räumlich-autoregressives Modell zur Nutzung vorgeschlagen. Ausserdem zeigt die Anwendung, dass ein grosser Aufwand erforderlich ist für die Generierung und Modellierung von Preisdaten, wenn Immobilienmärkte in einer integrierten Simulation von Flächennutzung und Verkehr von besonderem Interesse sind.

Bei der folgenden Anwendung geht es um die Implementierung eines Flächennutzungsmodells für den Grossraum Zürich. Eine solche Simulation kann ein hilfreiches Werkzeug sein zum Beispiel zur Abschätzung der Wirkung von Verkehrsinvestitionen auf Flächennutzung und Siedlungsentwicklung sowie den Wechselwirkungen, aber auch für den Vergleich von gewissen Politikszenarien. Der Fokus liegt in dieser Arbeit jedoch auf Datenbelange, da sie ein kritischer Faktor sind im Prozess der Implementierung. Des Weiteren wird ein vorläufiges hedonisches Mietpreismodell und ein Standortwahlmodell für Haushalte vorgestellt, interpretiert und im Detail verglichen.

Das vierte Beispiel beinhaltet den Vorschlag für einen agentenbasierten Modellierungsansatz. Die Arbeit basiert auf einer Literaturdurchsicht der Standortplanung im Einzelhandel sowie auf einer Serie von Interviews, die mit Standortentscheidern verschiedener Produktgruppen in Deutschland und der Schweiz durchgeführt wurden. Die Interviews machten deutlich, dass hochentwickelte räumliche Analyse- und Entscheidungswerkzeuge nur selten in der Praxis angewendet werden. Eine zusätzliche Literaturzusammenschau der agentenbasierten Beschreibung von Einzelhandelsmärkten geben zusätzliche Hinweise für die Modellierung von Einzelhändlern. Schliesslich wird ein konzeptionelles Modell für ein Einzelhandelsmodul innerhalb einer agentenbasierten Verkehrssimulation vorgeschlagen und verschiedene Fragen bei der Einführung eines solchen Agententyps diskutiert.

Insgesamt zeigen die verschiedenen Anwendungen nicht nur, wie räumliche Methoden hilfreich sein können bei der Erklärung von geografischer Variation von Daten, sondern auch, wie entsprechende Methoden und Werkzeuge helfen können bei gewissen Aufgaben der Standortplanung und der räumlichen Ressourcenverteilung.

1 Introduction

The importance of accounting for the spatial characteristics of data[1] has been long recognised for example in economics, geography, and regional science. Improved data availability and significant methodological and computational advances have brought increased interest particularly to the field of spatial analysis during the past two decades (Fotheringham *et al.*, 2000, 10; Fotheringham and Rogerson, 2009, 3). Many other scientific fields, besides the aforementioned ones, have contributed to the rise of spatial analysis in modern form, including biology, epidemiology, ethnology, sociology and statistics. This development has also led to some differing terms in spatial analysis[2]. Recently, the growing appreciation across academic disciplines is observable by rapidly increasing related publications as described by Anselin *et al.* (2004, 3), Florax and Van der Vlist (2003, 229f) as well as Janelle and Goodchild (2009, 22). Spatial analysis methods continue to diffuse into many areas of research. Nevertheless, the nature of the spatial variation of interest is sometimes not well understood and the patterns of spatial dependence and heterogeneity are disregarded (Atkinson and Tate, 2000, 607).

The term spatial analysis has been described as a collection of techniques and models that explicitly use the spatial referencing associated with each data value or object that is specified within the system under study (Haining, 2003, 4). Haining additionally remarks that spatial analysis methods need to make assumptions about or draw on data describing the spatial relationships or spatial interactions between cases. Unwin (2008, 392) summarises it as the systematic use of the geographic location(s) of objects of interest as an important variable in description, analysis, and prediction. Often, spatial analysis is applied to understand events, processes and spatial patterns or is used in areas where the identification of any regularities in spatial data is taken to signal something of substantive interest that justifies a closer investigation (Fotheringham and Rogerson, 2009, 1; Haining, 2003, 22). Therefore, spatial analysis is often called upon to address questions relating to outcomes that are in turn consequences of processes that by their nature are spatial (Haining, 2009, 13). Haining (2003, 21f) identifies

[1] It is believed that up to 80 percent of all data stored have a spatial component (Klinkenberg, 2003, 38; Nedović-Budić and Pinto, 1999, 183), although there is considerable concern that this figure is rather a guess than based on scientific research.

2 For example, the term 'spatial analysis' is sometimes used as a synonym for quantitative geography (Fotheringham *et al.*, 2000, 7f), geographic information analysis (O'Sullivan and Unwin, 2003) or geospatial analysis.

four generic groups of spatial processes: diffusion, exchange and transfer, interaction and dispersal processes.

Common software tools for spatial analysis are geographic information systems (GIS). GIS is a means to collect, store, retrieve, analyse, model and map spatial data (Church, 2002, 541; Longley, 2008). Therefore, it supports a variety of needs with regard to spatial analysis, including mapping, spatial queries and data supply to models and statistical tools (Church and Murray, 2009, 13). In part, the increasing popularity of spatial analysis is due to an increasing familiarity with GIS and the increasing availability of user-friendly, GIS-based tools for spatial analysis (Boots and Okabe, 2007; Rey and Anselin, 2006). However, GIS vendors often use the term spatial analysis as the ability of the GIS system to perform manipulations on the basic geometry of the data such as buffering, point-in-polygon, overlaying etc. (Unwin, 2008, 392). Fotheringham *et al.* (2000, 8) emphasise that this forms an extremely minor part of what is typically thought of as spatial analysis.

All problems tackled in this dissertation fall into the realm of spatial analysis. One focus is the explanation of spatial variation of certain measures of interest. Theoretically, there are two reasons for spatial variation, namely compositional and contextual influences (Haining, 2003, 17) which might operate at several spatial scales and levels. Differences of housing prices for example might be due to different structural attributes such as size or age (compositional effects) or due to differences a place makes (contextual effects). The effects might even be compounded (Orford, 2000, 1647). Another focus is the examination of certain spatial analysis, modelling and simulation techniques for solving location problems and informing spatial decisions and deployment of resources. Namely, regression techniques, integrated land use and transport simulation and agent-based modelling approaches are applied, whereas all those techniques are grounded in spatial analysis or are applying spatial analysis methods. Therefore, the dissertation brings together a wide set of spatial analysis techniques to solve real world problems.

1.1 Overview of the thesis and context

The core of the thesis consists of four applications. The applications are extracts from several working streams, but all are concerned with methods and applications in the field of spatial analysis.

Chapter 2 provides background information and forms a frame for the main topics of the dissertation. It is thought as complementary to the work described in more detail in the respective method sections of the following chapters.

The work presented in Chapter 3 is an outcome of collaboration with a major insurance company. The initial work included descriptive statistical analysis in location and market research for the optimisation of facility and agent allocation. However, the full potential of the data is explored by establishing a model which evaluates the market share of the tied-agent distribution channel, considering the distribution success of the broker and direct channel of the company, competitor density and other measures while controlling for spatial effects by a spatial simultaneous autoregressive model. The outcome potentially helps in commercial planning and forecasting, market potential assessment as well as monitoring and workforce planning.

In Chapter 4, the hedonic regression technique is applied to a residential asking rent data set, whereas global and local regression approaches are compared. The estimations were a follow-up to the efforts of setting up a land use model for the Greater Zürich area, which is presented in Chapter 5. For the land use simulation, a basic hedonic model was applied, but the results motivated further work and the models presented in Chapter 4 were extended in terms of variable generation, selection and methods. More specifically, spatial simultaneous autoregressive models were applied and compared to the results of ordinary least squares regression (OLS) and a local modelling approach, called geographically weighted regression (GWR). In the earlier working stream, the land use model UrbanSim was implemented for the Greater Zürich area, involving significant work in data collection, preparation and modelling. The content of Chapter 5 includes a description of the land use model itself as well as efforts and data challenges that were faced during the implementation. Additionally, some model estimates are presented, interpreted and compared in detail. Finally, some notes on validation and transferability are being made.

In Chapter 6, a review of retail location planning techniques is presented and agent-based modelling is suggested as a promising location choice and policy analysis tool. More specifically, a literature review about retail location planning is given and the results of a set of in-depth interviews with retail location decision makers in Germany and Switzerland is summarised. Based on these findings, a concept is formulated for extending the agent-based transport simulation MATSim-T (Balmer *et al.*, 2009) with an additional retailer module. It seeks to model the location planning and choice of retailers based on traffic volumes.

Chapter 7 concludes with a synthesis, including a consolidation and discussion of the experiences from the various spatial analysis applications. Additionally, alternative approaches and potential further work are described.

2 Background

Spatial data may relate to point, line or area 'objects', or may be a sample of some continuous scalar or vector field. The data can be ratio scaled, categorical properties of objects, aggregate values for discrete areas of the earth's surface or even a mixture of several variables that ideally should be examined together (Unwin and Unwin, 1998, 415). Fischer (2006, 18) additionally names spatial interaction data, which consists of measurements, each of which is associated with a link or pair of locations representing points or areas. The common factor linking these disparate types of data is that they have a geographical location that is of value in spatial analysis. Working with spatial data involves various specific properties compared to aspatial data. Besides those 'fundamental' properties mentioned above, they might be due to the chosen space and a consequence of measurement processes (Haining, 2009, 5).

2.1 Spatial analysis and locational problems

A classification of the techniques of spatial analysis is difficult because of several reasons: many fields of research are involved as stated above, the approaches are manifold and the data itself can take many forms. Therefore, Upton and Fingelton (1995, 1) state that it is not easy to arrive at a system of classification that is simultaneously exclusive, exhaustive, imaginative, and satisfying. Nevertheless, several authors have suggested classifications. Haining (2003, 4) identifies three main elements of spatial analysis: spatial data analysis, mathematical modelling and cartographic modelling. According to Fischer (2006, 17), spatial data analysis focuses on detecting patterns and exploring and modelling relationships between such patterns in order to understand processes responsible for observed patterns. In spatial modelling, model outcomes are dependent on the form of spatial interaction between objects in the model, or spatial relationships or the geographical positioning of objects within the model. Finally, cartographic modelling concerns the representation of the data on a map or map-based operations, such as overlay techniques, to generate a new map (Haining, 2003, 4). The purpose of most activities is to add to the understanding of and maximise knowledge on spatial processes (Fotheringham *et al.*, 2000, 4).

Fotheringham and Rogerson (2009, 2) categorise four main spatial analysis techniques:

1. Those spatial analytical techniques aimed at reducing large data sets to smaller amounts of more meaningful information. Summary statistics, various means of visualising data and a wide body of data reduction techniques are often needed to make sense of what can be extremely large, multidimensional data sets.

2. Those techniques collectively known as exploratory data analysis which consists of methods to explore data (and also model outputs) in order to suggest hypotheses or to examine the presence of unusual values in the data set. Often, exploratory data analysis involves the visual display of spatial data generally linked to a map.

3. Those techniques that examine the role of randomness in generating observed spatial patterns of data and testing hypotheses about such patterns. These include the vast majority of statistical models used to infer the process or processes generating the data and also to provide quantitative information on the likelihood that the inferences are incorrect.

4. Those techniques that involve the mathematical modeling and prediction of spatial processes.

The focus in this dissertation is on the latter two categories, which are collectively called the model driven spatial data analysis approach (Fischer, 2006, 23), although elements of the first two categories are employed as tools as well.

The terms location science and location analysis often refer in the literature to problems which were summarised as siting facilities in some given space (ReVelle and Eiselt, 2005, 1) and are predominantly entrenched in logistics and most notably in operations research (see also Brandeau and Chiu, 1989; Hale and Moberg, 2003), which is concerned with algorithmic efficiency and with the mathematical properties of optimal solutions. This type of modelling in the context of location-allocation has been named prescriptive or normative (Church and Murray, 2009, 2). Therefore, it focuses on the determination of the best location for one or more facilities, although the techniques are equally applicable in spatial and non-spatial domains (Church and Murray, 2009, 12). In many industries, there is the need to allocate, for example, plants, distribution centers as well as sales and other facilities.

A second perspective on location analysis concerns the understanding and explanation of the entire landscape of activity locations (Johansson and Forslund, 2008, 39). Therefore, the approach aims to describe arrangements that have emerged and tries to explain why certain de-

cisions were made as for example described in O'Sullivan (2007). Church and Murray (2009, 2) indicate that there is a third way of how location can be studied for which they refer to as techniques of mapping by positioning an object's location in space by its locational coordinates.

All location analysis perspectives involve elements of spatial analysis methods as listed above and they can inform and support a broad variety of business and policy decisions. The retail industry for example forms a prime example for the application of operations research techniques for location planning (Nwogugu, 2006). Nevertheless, many common techniques in retail location analysis practice are not based on operations research methods as described for example in Rogers (2007), Wood and Browne (2007) and in more detail in the literature review in Chapter 6, but on conceptual links between geography and management/marketing (Clarke *et al.*, 1997, 60). Related to the allocation of one or several facilities is the tactical and strategic deployment of resources in space. Such tasks with their manifold peculiarities might be found both in governmental/public and private business realms. As an example, Haining (2003, 36f) describes a framework for a spatial deployment task: identify the areas of need according to specified criteria and a set of objectives, target the intervention, manage and monitor the intervention and evaluate the outcomes in relation to the objectives set. These areas might be large such as regions entitled for example to bid for public funding or small areas such as neighbourhoods with the need to reduce crime by certain police intervention (i.e., Ashby and Longley, 2005).

Within the last decade, spatial analysis and GIS have been used increasingly in business for identifying the location of customers, customer segmentation and targeting (also known as geodemographics; Webber, 2008), trade area analysis, competitor analysis as well as optimising marketing efforts and facility locations (i.e. Harris *et al.*, 2005; Hess *et al.*, 2004; Zhao and Harris, 2006). However, Clarke and Stillwell (2004, 3) state that there are relatively few examples of applied quantitative geography studies in the literature due to the fact that many such projects are confidential, especially for private sector applications.

Chapter 3 includes an application in the private sector by analysing and modelling the determinants of market shares of an insurance company nationwide at the municipality level. Market share models have been widely used in practice and are core models in marketing science (Hanssens *et al.*, 2001, 3). The argument to use spatial analysis and modelling techniques in the analysis of the national market of a major insurance company at the municipal level is motivated by the observation that the main determinants of the market have substantial spatial variation: location of the company's customers as the origin of premiums, readily available

socio-demographic and socio-economic characteristics of inhabitants of the municipality as well as the location of the agents of both the insurance company and its competitors, resulting in agent densities. The developed model can inform resource allocation decisions.

2.1.1 Issues and challenges in spatial analysis

When working with spatial data, the researcher needs to be aware of the problems and issues associated with them. Fotheringham and Rogerson (1993, 4) note that spatial analysis is not simply aspatial analysis performed upon spatial data and in many instances, aspatial forms of analysis might even be simply not suitable for spatial applications. As an example, they compare the differences between an aspatial problem like the choice of a brand with an equivalent spatial problem of residential location choice. Issues might arise because of the definition of a region which might be perceived and defined in different ways by different people and might not be as obvious as the classification of brand types. A second and related issue is the degree to which alternatives within clusters act as substitutes for one another. In an aspatial choice situation, there is often no ordering of alternatives within a cluster so that all alternatives within a cluster are equally likely substitutes for each other. In spatial choice, however, the alternatives are fixed in space and near locations are more likely to be substitutes than ones farther away.

That properties often vary across space is, for example, described by Unwin and Unwin (1998, 415). Haining (2003, 40) notes that some elements of spatial variability may be an artefact of the data arising from errors that have propagated through a data set as a result of carrying out arithmetic or logical operations on erroneous data as described for example in Arbia (1998). Relationships might vary across space as well. There are at least three reasons to suspect that relationships will vary spatially (Fotheringham *et al.*, 2000, 94f). First, there might be spatial variations in observed relationships caused by random sampling. The second reason is that the relationships might be intrinsically different across space. The third reason is that the model used to measure relationships is a gross misspecification of reality, relevant variables are missing or represented by an incorrect functional form. In practice, it may be difficult to distinguish between these reasons (Lloyd, 2007, 4).

Fotheringham and Rogerson (1993, 4) summarise eight specific impediments that arise in spatial analysis: the modifiable areal unit problem (MAUP), boundary problems, spatial interpolation, spatial sampling procedures, spatial autocorrelation, goodness-of-fit in spatial modelling, context-dependent results and non-stationarity as well as aggregate versus disaggregate models (for an almost similar list see Fischer, 2006, 1). Additionally, Fischer (2006, 19) men-

tions that nearly all spatial data are flawed to some degree. Reasons for data imperfection might be vagueness, error, imprecision and uncertainty (Duckham *et al.*, 2001; Olteanu *et al.*, 2006, 697). Errors may arise in measuring both the location and the attribute properties, but may also be associated with computerised processes responsible for storing, retrieving, and manipulating spatial data. Another origin of problems might be the fact that spatial data are often collected at different spatial scales (Gotway and Young, 2002). For example, Haining (2003, 18) notes that 'place' can refer to areal objects of varying sizes – even within the same analysis. Atkinson and Tate (2000, 608) describe that the scales of variation observable in spatial data are inextricably linked to the scales of measurement through which they were obtained. It becomes obvious that methodologies for spatial data analysis have to be tuned to the properties of spatial data (Haining, 2009, 5) and the problems mentioned above. Additionally, particularly for applications in the social sciences, there is always some degree of error indicating that a model has not captured fully the process it is being used to examine (Fotheringham *et al.*, 2002, 9). Although it is not necessarily an exclusive problem of spatial data, it complicates problem localisation. Finally, issues might arise in the dissemination of spatial analysis results for example when results are presented as maps. Monmonier (1996, 3) describes that maps, like numbers, are often arcane images accorded undue respect and credibility, although many errors can occur for example when a well-intentioned map author fails to understand cartographic generalisation and geographic principles (Monmonier, 1996, 184).

2.2 Spatial effects, local models and spatial econometrics

2.2.1 Spatial effects

Tobler (1970, 236) formulates what he called the first law of geography: Everything is related to everything else, but near things are more related than distant things. According to Haining (2003, 33), this holds not only for single entities, but for values representing aggregates with respect to an areal partition. Many examples have been found for example in socio-economic and environmental analysis. This 'law' implies that there will be positive spatial autocorrelation. Therefore, it poses a challenge for statistical methodology because spatially proximal observations are unlikely to be statistically independent (Doh and Hahn, 2008, 666) and it might be incorrect to assume that results obtained from the whole data set represent the situation in all parts of the study area (Fotheringham *et al.*, 2000, 11). Nevertheless, the fact that events close together in geographic space tend to be more alike than those farther apart can be exploited to handle several scientific and technical problems (Haining, 2003, 35). Examples might be the sampling for estimating a parameter without taking an unnecessary large sample

size, spatial interpolation or imputation. Haining (2009, 7) notes that the law is clearly an oversimplification. Nevertheless, he admits that it is a useful aphorism.

The term *spatial effects* describes spatial dependence (or its weaker expression, spatial autocorrelation) and spatial heterogeneity. Both effects are usually not easily discernable in an empirical sense. The latter is structural instability, also called non-stationarity. In a regression context, it either appears in the form of non-constant error variances in a model (heteroscedasticity) or in the form of variable regression coefficients (Anselin, 2001a, 311). The work of McMillan (2003) and Thériault *et al.* (2003, 26) show that spatial heterogeneity will produce spatial autocorrelation among model residuals if it is not appropriately handled in the model specification. As described above, spatial dependence is generally taken to mean the lack of independence which is often present among observations in cross-sectional data sets. In a regression context, spatial dependence is present when the housing rent or price is similar at neighbouring locations for reasons other than those explicitly incorporated into the model. The causes for spatial effects might be manifold: It can be caused by a lack of correspondence between the spatial scope of the phenomenon under study and the delineation of the spatial units of observation. Anselin (2001b, 705) also named the inherent need to "integrate" data from various sources and spatial scales as a cause in this context. Further reasons might be the existence of spillover effects, such as the impact of the price of one housing unit on the price of its neighbours, spatially correlated variables that have been omitted as well as measurement error or misspecification of the functional form (Anselin, 1988, 8 and 12; Wilhelmsson, 2002, 92). Anselin (1988, 8f) summarises that the inherent spatial organisation and spatial structure of phenomena will tend to generate complex patterns of interaction and dependencies. Therefore, there will be often a mixture of both spatial dependence and spatial heterogeneity.

Ignoring spatial effects in ordinary least square regression would result in unbiased but inefficient parameter estimators and biased variance estimators. Therefore, they will produce incorrect confidence intervals for estimated parameters and for predicted values, making inference unreliable (Basu and Thibodeau, 1998, 62; Militino, 2004, 194). There are a wide set of methodological approaches available to assess and treat the problem. Particularly since the mid-1990s, a tremendous growth in publications incorporating spatial effects has been detected (Anselin *et al.*, 2004, 1) although inferential methodology in the presence of spatial autocorrelation has been a topic for some time (Cliff and Ord, 1973).

2.2.2 Local models

Similar, the attempt to understand and specify local relations in multivariate data has some tradition in spatial analysis (Longley and Tobón, 2004, 510). Often, it has been shown that a global model does not represent well the variation at any individual location and interesting insights have been obtained by investigating spatial variations. Indeed, the spatial structure of a process may vary from place to place (Lloyd, 2007, 3). Fotheringham *et al.* (2000, 11) note that simply reporting one 'average' set of results and ignoring any possible spatial variations in those results is equivalent to reporting a mean value of a spatial distribution without seeing a map of the data. Additionally, global results will have limited application to any particular part of that region and may not represent the actual situation in any part of it (Fotheringham *et al.*, 2000, 93). Therefore, the focus of attention has shifted from identifying and understanding global regularities and similarities to differences across space and local exceptions (Fotheringham *et al.*, 2000, 11).

Various techniques for local spatial analysis were developed particularly in the last two decades, acknowledging heterogeneity of spatial data and the tendency to exhibit spatial non-stationarity (Goodchild, 2009, 471; Lloyd, 2007, 1). They have been summarised as local models (Boots and Okabe, 2007; Lloyd, 2007). Global methods make use of all available data, whereas local models are often defined as those that make use of some subset of the data whose characteristics are statistically 'significant' in some way (Boot and Okabe, 2007, 356). One simple way to incorporate geographical information at the local level is to introduce region specific dummy variables or regional interaction terms into the regression equation to control for spatial components. While this method addresses geographical issues to some extent, the discrete nature of this approach might be arbitrary, as Clark (2007, 189) notes. But beyond, more sophisticated local models are available, covering a wide set of methods, such as approaches for exploring local spatial autocorrelation (Anselin, 1995) and for exploring and modelling variation in spatial relations between multiple variables such as geographically weighted regression, which is described and applied in Chapter 4. Due to different research traditions of various disciplines, there is some considerable overlap with approaches developed in the realm of spatial econometrics (see next chapter). Lloyd (2007, 3) notes that the application of local models may be more problematic than the application of global models because of the additional complexity. Factors such as the size of a moving window or the type of transformation applied may have a major impact on the results obtained from an analysis.

2.2.3 Spatial Econometrics

In the 1980s, Anselin (1988, 7) defined spatial econometrics as the collection of techniques that deal with the peculiarities caused by space in the statistical analysis of regional science models. At that time, he described that the modelling perspective distinguishes spatial econometrics from the broader field of spatial statistics/analysis (Anselin, 1988, 10). Since the end of the 1990s, applications in agricultural, environmental and natural resource topics are emerging as well (Anselin, 2002, 247; Florax and Van Der Vlist, 2003, 229f) and nowadays, spatial econometrics is considered a subarea of spatial analysis (Voss, 2008, 410). However, Voss also describes it as the branch of economics that applies statistical and mathematical rigor to the study of economic theories and relationships (Voss, 2008, 409). More specifically, spatial econometrics can be characterised as the set of techniques to deal with methodological concerns that follow from explicit consideration of spatial effects. This includes four broad areas of interest (Anselin, 2001a, 311):

1. the formal specification of spatial effects in econometric models,

2. the estimation of models that incorporate spatial effects,

3. specification tests and diagnostics for the presence of spatial effects and

4. spatial prediction (interpolation).

The field of spatial econometrics is held together by various tests on the spatial autocorrelation that might exist in spatial economic and related data (Getis, 2008, 304). A prerequisite for detecting spatial autocorrelation and heterogeneity is the definition of relationships between locations. Principally, there are two general approaches (Anselin, 2002, 256). One is the geostatistics-inspired approach, in which the spatial correlations are modelled as a function of separation distance. Several specifications for this distance decay function have been employed, but most are some variant of a negative exponential model. The other approach is using an object view and corresponding lattice model. The covariance structure follows indirectly from the specification of the spatial weights matrix that underlies a spatial process model which can involve a simple binary connection matrix or a generalised weighting matrix as provided by Cliff and Ord (1973, 11f). It indicates the presence, absence or intensity of relationships between the locations concerned by a set of operators (Bivand, 2008, 13). The lattice model approach is further described and applied in this dissertation, involving work on polygon (Chapter 3) and point data (Chapter 4). Griffith (1996) shows that a parsimonious

specification of the relationships between observations is to be preferred for example for one making assumptions about distance decay. Moreover, the definition of a spatial weights matrix has to be done with caution since alternative definitions can lead to very different neighbourhoods and hence values for the statistics (Unwin and Unwin, 1998, 417). Some general guidelines are given by Griffith (1996).

Typically, the motivation for applying a spatial econometric model is not driven by formal theoretical concerns, but data-driven due to data "problems" (Anselin, 2002, 253). This might be due to a mismatch of the scale and location of the process under study with the available data or caused by explanatory variables constructed by spatial interpolation in order to make their scale compatible with that of the dependent variable. A similar situation is when data on important explanatory variables are missing and those variables show spatial structure. In any of these cases, the error term in the regression model will tend to be spatially correlated. The question if a spatial pattern observed in model residuals are a reaction to model misspecification or if they signal the presence of substantive interaction between observations remains an unresolved issue (Bivand, 2008, 18; Cliff and Ord, 1981, 141f; Olsson, 1970, 228). Bivand *et al.* (2008, 273) suggest to spent efforts on improving the model itself in case of spatial autocorrelation, i.e. by handling heteroscedasticity, by adding a missing covariate, by revisiting the functional form of included covariates, or by reconsidering the distributional presentation of the response variable instead of trying to model the spatial structure. Cliff and Ord (1981, 142) mentions that the choice of model must involve the scientific judgement of the investigator and careful testing of the assumptions. Nevertheless, the developments in theory and the desirability of having theory rather than data driving the modelling process are gradually steering spatial econometrics away from the traditional techniques according to Florax and Van Der Vlist (2003, 235).

So called spatial simultaneous autoregressive models (so simply spatial regression models) are often applied in spatial econometrics. This family of models, including spatial lag and spatial error models, was originally solely based on maximum likelihood as introduced in Ord (1975) and elaborated in Anselin (1988). Once again, there is a need to define neighbourhood or 'local' respectively whereas the same approaches are applicable and can be used as described above. Spatial simultaneous autoregressive models are further detailed and applied in Chapter 3 and Chapter 4 of this dissertation.

2.3 Hedonic approach in housing markets

The hedonic method was originally applied in the field of agricultural economics in the US in the 1920s (Colwell and Dilmore, 1999, 620). More famous is an application to automobile price indices in the 1930s by Andrew Court (Goodman, 1998, 292). In housing, the approach was not adopted before the articles of Griliches (1971), Lancaster (1966 and 1971) and Rosen (1974). A more comprehensive review of the history might be found in Malpezzi (2003).

According to the seminal article of Rosen (1974, 34), the hedonic price model is based on the hedonic hypothesis that goods are valued for their utility-bearing characteristics and it is determined by a set of choices made by consumers and producers under market clearing conditions. In the real estate and housing context, a dwelling is a special commodity with bundles of attributes that cannot be traded separately (see also Lancaster, 1966). This includes structural attributes referring to the characteristics of the dwelling itself, such as size, age and quality. Moreover, it also includes locational characteristics such as accessibility, attributes of the neighbourhood, proximity to externalities and environmental amenities. Finally, additional attributes such as contract and landlord/previous owner characteristics might be considered as well. The explicit market is for the overall bundles themselves, therefore for the dwellings with their observed total prices. At its simplest, a hedonic equation is a multiple regression of dwelling prices or rents on its various characteristics. Marginal prices of various characteristics of the properties are identified as the partial differentiation of corresponding attributes. This marginal price is usually referred to as the hedonic or implicit price of the attributes (Rosen, 1974, 50). Therefore, the explicit price of a dwelling is the sum of the implicit prices of the various characteristics contained in the bundle. Overall, the hedonic approach provides a methodology for decomposing the overall dwelling prices into prices of the various characteristics and for identifying the structure of prices of the dwelling attributes. The estimated parameters can be interpreted as the willingness to pay for the referring dwelling attributes. A more detailed recent discussion of the theoretical foundations of the hedonic approach might be found in Taylor (2008).

There is a wide set of common applications of the hedonic approach in housing issues, including appraisal for property taxation and mortgage underwriting, demand studies, policy analysis for estimating the impact of environmental amenities or nuisances on dwelling prices and the development of quality-adjusted price indices. The hedonic method has matured to the standard way economists deal with housing heterogeneity (Malpezzi, 2003, 87). This is certainly due to its flexibility and wide applicability. Another benefit of the approach is that it usually uses revealed preference data, such as observed paid transaction prices. Therefore, the

method is based on households' real willingness to pay for the dwelling's characteristics as opposed to the alternative of stated preference surveys. Moreover, it integrates and values environmental and locational amenities and nuisances in a coherent framework, besides considering structural characteristics of the dwelling units (Baranzini *et al.*, 2008b, 4).

Much literature is concerned with the assessment of real estate price variation across metropolitan areas and its determinants (for a comprehensive review of around 125 studies, see Sirmans *et al.*, 2005). However, early work considered only structural variables since the consideration of locational characteristics was technically difficult at that time (Kain and Quigley, 1970, 533). The application of spatial analysis and GIS as tools for calculating various spatial explanatory variables in hedonic studies began to be explored more than 15 years ago and allowed to incorporate neighbourhood and location variables (i.e. Bateman, 1994; Waddell *et al.*, 1993). It was about the same time, when spatial analysis begun to be used more broadly to control for spatial dependence and heterogeneity. One of the first empirical analyses is by Can (1992), who revealed that the incorporation of both spatial dependence and heterogeneity is superior to specification which only considers neighbourhood effects. Similar housing and real estate studies, in which spatial effects were considered, affirmed the substantial improvement in the quality of predictions and statistical inference (i.e. Basu and Thibodeau, 1998; Bowen *et al.*, 2001; Chica-Olmo, 1995; Dubin, 1998; Gelfand, 1998; Pace *et al.*, 1998) and the approaches were further developed and advanced in the meantime. Nevertheless, Ekeland *et al.* (2004, 61) state that the full potential of hedonic models remains to be exploited and Orford (2002, 105) describes that the conceptualisation and measurement of locational factors in hedonic models still remains to be a source of controversy since the results often contradict the theory of locational externalities. Moreover, Bowen *et al.* (2001, 484) note that the need to incorporate an explicitly spatial hedonic specification is not always necessary and needs to be tested.

There is often spatial dependence in residential rent and price data analysis, sometimes also in the hedonic price residuals. This is due to several reasons. Shared environmental and locational amenities and proximity externalities of adjacent houses such as the same distance to recreation and shopping facilities, social services, public transport, noise pollution or general accessibility are one reason. The same applies to favourable neighbourhood variables such as education, income and nationality of inhabitants and the density of the neighbourhood (Griffith, 2003, 5; Militino *et al.*, 2004, 193). Another, related reason is that occupiers with the same socio-economic characteristics, attributes, preferences and life-styles tend to spatially cluster together. Finally, structural characteristics such as housing types, age of construction,

building material etc. might be similar in adjacent houses as they were often developed at the same time (Basu and Thibodeau, 1998, 61).

Spatial heterogeneity might be present in housing markets in which the same set of dwelling characteristics yield different rents or prices in different parts of the study region. Therefore, there are variable marginal prices of characteristics across the region. This has been shown in various studies in the past (Fik *et al.*, 2003; Fotheringham *et al.*, 2002, 27; Thériault *et al.*, 2003). One reason is for example that various types of households have different needs and preferences and are not evenly distributed within the study region. This can locally distort the demand for specific structural attributes and amenities for dwellings, resulting in submarkets with unique contextual composition (Thériault *et al.*, 2003, 26). Nevertheless, Baranzini *et al.* (2008c, 706) describe that estimators of hedonic models are quite reluctant to include person or household characteristics, because the assumed competitive market prices are independent of individual buyers and sellers. Often, study areas are composed of several submarkets, which can be characterised by functional disequilibrium and segmentation and empirical studies differ regarding how submarkets are specified, and hence how spatial heterogeneity is treated (Yu *et al.*, 2007, 1087f). This is reflected in the various approaches to incorporate location characteristics and to consider spatial effects, which are described in Chapter 4.

Besides the problems of spatial effects described above, multicollinearity and heteroscedasticity are the other main sources of problems in hedonic models (Des Rosiers, 2000, 294; Fletcher *et al.*, 2000), which can also caused by neglecting spatial elements (Orford, 1999, 23). Sheppard (1999, 1631f) summarises that some problems associated with estimation of the hedonic price function itself are conventional, even if not easily solved. Often, estimation must confront inadequate data and make use of whatever information sources are available. Mostly, transaction prices are used for hedonic estimations, although asking prices are an acceptable alternative when sales data are unavailable (Brunauer *et al.*, 2009; Cheshire and Sheppard, 1998; Henneberry, 1998). Residential asking rent prices were available for the application of the hedonic application in Chapter 4.

2.4 Spatial simulation approaches

The spatial analysis methods described in Chapter 2.3 fall into the realm of equation-based approaches. Referring regression techniques have been applied in Chapter 3 and 4. Authors have described that the equation-based approach is competing with the simulation approach since both simulate the system by constructing a model and executing it on a computer (Van

Dyke Parunak *et al.*, 1998, 10). Moreover, in both cases, the underlying models are a simplification of some other structure or system (Gilbert and Troitzsch, 2005, 2).

Simulation in spatial analysis and modelling (also called spatial or geo-spatial simulation) has been one of the key approaches of many researchers (Crooks *et al.*, 2008; Xie and Brown, 2007, 229). According to Berry *et al.* (2008, 230), the work of Hägerstrand (1968) has stimulated the adoption of simulation experimentation as part of spatial analysis in the early development.

In general, computer simulation has been recognised as a third way of doing science, besides theoretical analysis/ deduction (i.e. testing of sets of assumptions and their consequences) and empirical analysis/ induction (the development of theories by generalisation of observations). It is often chosen to obtain a better understanding of some features of the real world, to help humans make better decisions about systems that are too complex for any one person to easily understand and to be able to forecast certain scenarios (Gilbert and Troitzsch, 2005, 2). Additionally, simulation techniques have been used for example to express a process in space in the absence of suitable analytical methods (Bivand, 2008, 7). Therefore, one starts with a set of assumptions, but then uses an experimental method to generate data which can be analysed inductively (Axelrod, 1997b, 24; Gilbert and Troitzsch, 2005, 26). The goal is to evaluate action alternatives or to generally enrich our understanding of a situation, sometimes even of fundamental processes. In the latter situation, simplicity of the assumptions is important, and realistic representation of all the details of a particular setting is not (Axelrod, 1997a, 5). However, researchers often work towards complexity to increase the sensitivity or their simulation system. Then, the simulation approaches are tempered by the challenges posed by representing complex systems, such as the need for data and the difficulty of validating complicated models (Manson, 2008, 4). On the other hand, one of the benefits of the simulation approach is that it partially overcomes the empirical problem of data availability, since a simulation often produces its own 'virtual' data (Harrison *et al.*, 2007, 1230).

A number of typologies of simulation models have been proposed (e.g. Burton, 2003; Gilbert and Troitzsch, 2005, 7; Macy and Willer, 2002), in particular for land use and transport simulations (Chang, 2006; Iacono *et al.*, 2008; Timmermans, 2006; Torrens, 2000; Wegener, 2004). The development has been from macrosimulation based on aggregate, equilibrium of spatial interaction models beginning in the 1960s to microsimulation and agent-based modelling more recently, also called multi-agent simulations (Van Leeuwen *et al.*, 2007). Caldwell (1997, as cited in Macy and Miller, 2002, 145) points out that microsimulation is a 'bottom-up' strategy for modelling the interacting behavior of decision makers (such as individuals,

families and firms) within a larger system. This modelling strategy utilises data on representative samples of decision makers, along with equations and algorithms representing behavioral processes, to simulate the evolution through time of each decision maker, and hence of the entire population of decision makers. Moreover, microsimulation models use for example observed population distributions to estimate parameters for models of household characteristics (e.g. number of children, labor force status, income, etc.). Then, the models age the population at the individual (or household) level by updating these characteristics. Microsimulations model changes to each element of the population distribution rather than changes to the distribution at the population level (Macy and Miller, 2002, 145).

2.4.1 Integrated land use and transport simulation

Integrated land use and transport simulation systems are a particular class of spatial simulation models which are used to simulate how land use and transport systems operate. According to Torrens (2000, 5), they combine theory, data, and algorithms to create an abstract representation of the character and functioning of the land use transport system. Therefore, their role in policy analysis is to capture the important relationships in the urban system so that the consequences of alternative policy decisions such as new transport infrastructures can be projected and studied in advance for example with regard to the housing markets or the urban development patterns (Kanaroglou and Scott, 2002, 42).

Early traces of integrated land use and transport simulations appeared in the early 1960s in the US, but only ten years later, many efforts were disrupted and Lee (1973) summarises that these efforts have essentially failed due to hypercomprehensivness, grossness, (data) hungriness, wrongheadedness, complicatedness, mechanicalness and expensiveness of the envisaged models (Lee, 1973, 164ff). Improvements in modelling theory and methodology, data and software availability, computer technology as well as concerns raised by the worsening transport and environmental conditions brought new interest to this kind of simulation particularly in the academic world in the 1990s. But even with the mentioned advancements, not all recent integrated land use and transport simulation projects were successful in this new wave of interest, as Wegener (2009, 2) summarises. According to him, many projects failed to deliver in the time available or had to reduce their too ambitious targets. Additionally, many projects did not reach the state of policy analysis and/or remained in the academic realm. This seems to confirm the doubts described by Timmermans (2006), who alludes that some of the old fundamental problems have still not been solved in the current simulations for example with regard to model formulation and the underlying theory. Moreover, he states that it does not seem realistic to expect that any integrated land use and transport simulation could provide

accurate land use forecasts at the level of individual cells given the still prevailing relative lack of data. He suspects that such simulations could provide some rough qualitative indications for wider areas, rather than a detailed quantitative assessment of tendencies and the likely impacts of land use and transport policy scenarios. Therefore, he presumes that the potential of these simulations might lie in the area of policy scenario development in the sense that they provide a platform for discussion as opposed to being accurate forecast tools (Timmermans, 2006, 239).

Various techniques were developed to simulate the complex land use transport interaction, including econometric and microsimulation models (Iacono *et al.*, 2008, 329 and 333), whereas the latter can be considered as state of the art in integrated land use and transport simulation. Such a simulation was implemented for the Greater Zürich area and the experiences are reported in Chapter 5. Additionally, the chapter includes a short overview about current microsimulation models as well.

2.4.2 Agent-based simulation

Microsimulation models do not fully permit individuals to directly interact. This is a speciality of agent-based simulations, which are considered as one of the most advanced simulation techniques. According to Manson (2008, 4), agent-based simulations are computer models that use agents or small software programs to represent autonomous individual actors that create complex systems and to search for causal mechanisms that may underlie statistical associations. Agent-based simulations are helpful because they can identify how features of complex systems emerge from the simple interactions of their components. The term agent is meant here in the narrow sense of autonomous software objects with cognitive models that guide actions in an environment. Agents are autonomous, acting without other entities having direct control over them. Agent actions are defined in terms of a larger environment, which is everything external to the agent. Each agent individually assesses its situation and makes decisions on the basis of a set of rules. Agent actions typically affect the environment (including exchanging information or resources with other agents) and are, in turn, guided by the environment. Therefore, an agent reacts to environmental changes and it may also be able to learn. Agent-based simulations are increasingly used throughout the social, environmental, and natural sciences for research, policy formulation, decision support, and education (Manson, 2008, 5). They complement other approaches to modelling and answer questions in new ways. In particular, they analyse complex systems emerging from the “bottom up” via local interactions among agents. Although the assumptions may be simple, the consequences may not be at all obvious. The large-scale effects of locally interacting agents are called “emergent prop-

erties" of the system. Emergent properties are often surprising because it can be hard to anticipate the full consequences of even simple forms of interaction (Axelrod, 1997a, 4). Consequently, local interactions can lead to significant changes in overall system behaviour (Manson, 2008, 5).

Manson (2008, 5) describes that many of the characteristics that make agent-based simulations useful also introduce challenges in their use (see also Crooks *et al.*, 2008). He summarises that even when model assumptions are realistic, they are often buried in the model's programming, and results can seem contrived to the point where they reflect underlying programming more than the phenomena modelled. Additionally, agent-based simulations can suffer from high dimensionality, when the number of agents or entities grows to the point where there exist an exponentially large number of possible trajectories (ibid.).

In Chapter 6, an agent-based simulation approach is proposed for analysing and forecasting the location choices of retailers. The proposal is based on a literature review of retail location analysis and planning accompanied by a summary of 11 in-depth interviews with location decisions makers in the retail industry. It covers a functional extension of the agent-based transport simulation model MATSim-T (Balmer *et al.*, 2009).

3 Agents, space and market shares: A spatial analysis of the Swiss insurance market

3.1 Introduction

Since the early work of von Thünen (1842), Weber (1909) and Christaller (1933) researchers of various disciplines have theorised and empirically investigated the role of geographic space in the firm's choice of where to locate and in the performance implications of those choices. According to Doh and Hahn (2008, 661) much of the strategy research involving geographic or spatial constructs focuses on agglomeration economies and positive knowledge spillovers associated with locating near to other firms in a given industry, or proximate to other resources that could be valuable for founding patterns, firm's growth and development (i.e. Greve, 2000; Sorenson and Audia, 2000). An overview of the literature in the field of management is given in Sorenson and Baum (2003). They summarise that strategic interest in location appears recently greater than ever (Sorenson and Baum, 2003, 2), although it has been criticised that spatial constructs, spatial classifications and methodological approaches in empirical strategy research are often quite coarse (Doh and Hahn, 2008, 660). In marketing, there is a long tradition of considering spatial issues as well (Grether, 1983) and since the rapid improvement of data availability, techniques such as geographic segmentation (Ferrell and Hartline, 2005, 148) and the consideration of geodemographics (Harris *et al.*, 2005), locational considerations are common in marketing practice. This is no surprise given the manifold geographical questions in marketing such as in the field of distribution management, customer targeting and spatial resource allocation. Additionally, spatial modelling techniques are evolving in the field, acknowledging its usefulness in describing and analysing brands, consumers, markets, and other units of analysis, though not many references can be found in the literature so far. Jank and Kannan (2005, 623) carefully summarise the current knowledge (Bronnenberg and Mahajan 2001; Bronnenberg and Mela, 2004; Bronnenberg and Sismeiro, 2002; Mittal *et al.*, 2004; Ter Hofstede *et al.*, 2002) by noting that the findings seem to suggest that spatial data capture not only the geographical variations in supply side factors but also variations in demand side factors through variations in physical and psychological landscapes.

In this chapter, spatial analysis techniques are applied to the insurance market in Switzerland. More specifically, the drivers of individual non-life and life premium market share in the tied agent distribution channel of a Swiss major insurance company are spatially analysed and modelled for the whole country at municipal level, i.e. for 2721 municipalities with an aver-

age of 3000 inhabitants per municipality. Cross-sectional business data of the insurance company are merged with other available spatial data at municipal level. The agent locations of all competing insurances with local representation are taken from the official register of insurance agents in Switzerland (Federal Office of Private Insurance, 2008). It is, to our knowledge, the first time that such a nationwide analysis from the financial services industry is reported at this level of spatial detail.

Despite the recent introduction and increasing success of new insurance distribution channels, the distribution via agents remains the main channel for many insurance companies. Usually, an agent of an insurance company is active in a local market and tries to sustain long-lasting customer relationships. Consequently, a close spatial neighbourhood of customers and agents can be assumed, although some restrictions apply. Uncertainties arise due to the fact that a policy holder might have moved from one to a different municipality while a policy is active. The same is not precluded for agent locations, as they might take their portfolio of customers and their associated policies with them to a new location. Moreover, a customer might not necessarily be tied to the next available tied agent as a company might not ensure exclusive sales areas to agents. Therefore, the spatial relationship of a tied agent to its customers might be only moderately strong. This will be further explored in the following.

3.1.1 Study objectives

The study has three objectives. One is to explore to what extent the insurance market is still *spatial* despite the increasing influence of non-spatial direct distribution channels, stipulating a spatial relationship between market share in the tied agent distribution channel, own and competitor agent presence, strength of other distribution channels of the company and characteristics of each municipality. Particular focus is on the role of market cultivation and competition. Does intense market cultivation with a strong field workforce lead to higher market shares and how does high competitor density affect the outcome? Another objective is to estimate the relative weights of the different factors which determine the market share of the company in the tied agent distribution channel in each municipality. The proposed models ought to provide general insights into the determinants of the market position with implications to both strategic management and marketing. The third objective is to determine if various distribution channels are spatially related and to analyse if there is any degree of cannibalization to the tied agent channel through alternative distribution channels, such as the direct channel and brokers.

The results of the research are useful in various ways. Consistent estimates of spatial models of market share and the coefficients of their predictors are of interest for managers and com-

panies for the purpose of commercial planning and forecasting, market potential assessment, monitoring, workforce planning and marketing. For example, the optimal number of agents and their locations is an important strategic question for any insurance company with a tied agent distribution network. Moreover, residual maps of local spatial models can identify areas of over- and under-performance of a company.

The remainder of the chapter is organised as follows: Section 3.2 gives an introduction into the determinants of insurance consumption on the demand side as well as distribution on the supply side. Section 3.3 describes issues in distribution intensity, channel competition and performance measurement. Both sections help to identify explanatory variables for the empirical part of the chapter. Section 3.4 gives insights into the Swiss insurance market and a summary of spatially disaggregated market share studies in the literature. Section 3.5 describes the methodology and data sources while Section 3.6 includes detailed information about the calculation of an agent density measure. Section 3.7 characterises the data used in the following estimations. Section 3.8 includes model estimations and maps the results. Finally, conclusions are given in Section 3.9.

3.2 Insurance consumption and distribution

In the insurance business, the non-life and the life insurance sector are usually distinguished and subject to different regulatory regimes and different tax and accounting rules on the supply side. The main reason for the distinction between the two types is that the life business is long-term in nature. The coverage for life insurance can cover risks over many decades. In contrary, a non-life insurance, such as a liability or property insurance, usually covers a much shorter period. On the demand side, non-life and life insurance have different economic rationales which are reflected in the theoretical models of their consumption as well (see for example Mossin, 1968, for non-life and Villeneuve, 2000, for life insurances). Consequently, the distinction between non-life and life was acknowledged in this work as well.

3.2.1 Empirical insurance demand literature

Two fundamental research approaches can be distinguished and have been applied in the empirical literature to investigate insurance demand (Schlag, 2004, 10): (1) macroeconomic studies and (2) microeconomic studies. Macroeconomic approaches are either aggregated cross-sectional, time-series or panel analysis studies. Aiming to explain both life and non-life insurance demand, these studies (e.g. Beck and Webb, 2003; Enz, 2000; Hussels et al., 2005; Li *et*

al., 2007; Nakata and Sawada, 2007; a synopsis on life insurance demand is given by Schlag, 2004 and Zietz, 2003) are mostly done at the country or cross-country level, where observational units are separated by borders which are to a great extent coincident with the spatial diffusion of the phenomenon of interest (Lenzi and Millo, 2005, 3).

Microeconomic studies try to explain individual decision-making patterns at the household level. Examples are Galabo and Lester (2001), Japelli and Pistaferri (2002) and Dixon *et al.* (2006). Even in light of all this empirical work, some authors are pointing out that the understanding of insurance demand is still limited (Nakata and Sawada, 2007, 1). One reason is certainly that empirical studies in the field of insurance consumption have often the problem of overcoming significant limitations in data availability (Lenzi and Millo, 2005, 1). This is also the reason why there are only few studies on insurance consumption, which have considered the regional or even local spatial dimension of the insurance market in the past. Exceptions are Zhuo (1998) for a selection of regions and cities in China as well as Lenzi and Millo (2005) for Italian regions; but these are still rather large scale units.

Both income and wealth play a major role in insurance demand, and they are positively correlated. According to Nakata and Sawada (2007, 1) the standard model of insurance demand specifies the demand function as a function of the premium and the initial wealth rather than income. However, in the various studies at the country level, wealth is often proxied by income or, when not observable, by the gross domestic product (Schlag, 2004, 12). So is risk exposure, which is in turn related to total wealth and the level of economic activity (Lenzi and Millo, 2005, 6). Moreover, the literature suggests capturing risk aversion by education or the age structure of the population (Lenzi and Millo, 2005, 6). Various studies found out that better educated people purchase more insurance, even when controlling for the higher income levels associated with longer schooling.

Additionally, the spatially disaggregate dimension of insurance marketing and distribution is crucial for the insurance companies as for other businesses. However, very few studies acknowledge those opportunities in the field of insurance demand. One of the rare exceptions is Lenzi and Millo (2005). They studied the demand variation in the Italian insurance market at the regional scale and found spatial effects. Consequently, they highlight the importance of taking the spatial perspective into account. As mentioned before, Clapp *et al.* (1990) analyse the effect of location of agencies and field offices on the profitability of life insurance. The authors concluded that it would appear worthwhile for firms to invest in a systematic analysis of locational and demographic factors when planning distribution systems (Clapp *et al.*, 1990, 447).

3.2.2 Distribution of insurance

Multiple distribution methods are applied in the insurance business (Kim *et al.*, 1996; Regan and Tennyson, 2000, 709). Although there is large heterogeneity in the importance of the various distribution channels among European countries (CEA Statistics, 2008, 48), the predominant distribution channel for the larger insurance companies in the individual market is often still the *tied agent channel* (Dumm and Hoyt, 2003, 28), particularly in German speaking countries (Accenture & Universität St. Gallen, 2005, 35). Tied agents are paid by a particular insurance company to sell only its products. In Switzerland, the number of field staff paid by the insurance companies has increased 30 percent between 2005 and 2009 while the overall premiums income saw only a slight increase (Swiss Insurance Association, 2009). *Non-employee agents*, sometimes also called *exclusive agents*, are independent from the insurer and are typically small businesses and franchises with a well-specified contractual relationship with a single insurer. Agents with non-exclusive sales relationships are independent businesses with contractual agreements to sell the products of more than one insurer. These agents are called *multi-tied agents* or *independent agents*. *Brokers,* too, are independent businesses who may sell the products of more than one insurer. However, they have no formal contractual relationships with insurance firms and ideally represent the insurance purchaser as a client (Regan and Tennyson, 2000, 711). The *direct channel* via internet and phone exhibits rapid growth in new business in recent years but has achieved only a low absolute level to date. Although direct distribution services were first introduced in the 1990s, the total market share has, for example, only reached 3 percent in Germany (Knospe, 2008) and 5 percent in Switzerland (Baumann, 2007; Schletti, 2008) so far, where anecdotal evidence suggests that standardised non-life products are predominant in this distribution channel due to its simplicity. Therefore, the tied agent system remains the back-bone of the distribution strategy for many insurance companies, but they are often following multi-channel strategies.

The ongoing competitive and technological revolution in the financial services industries has resulted in greater segmentation of distribution channels by lines of business, and greater use of multiple distribution methods by companies (Coelho and Easingwood, 2008; Webb and Hogan, 2002, 338), including the establishment of marketing relationships and alliances with non-insurance companies (Regan and Tennyson, 2000, 709). In the last years, non-insurance companies such as retailers, automotive manufacturers, banks as well as integrated financing and credit card companies have entered various national markets in Europe. The newcomers' main aim is to earn distribution commissions rather than bear the risk (PartnerRe, 2001, 7). Banks in particular are getting involved in manufacturing, marketing or distribution of insurance products, known as *bancassurance* (Artikis *et al.*, 2008). Certainly, the importance of al-

ternative distribution channels has increased in the last years and will further expand in the near future (Accenture & Universität St. Gallen, 2005, 35; Graf and Maas, 2008, 19).

3.3 Distribution channel intensity, competition and performance

The effects of distribution intensity on demand and market share are not widely covered in the literature. While there are few studies available on retail goods (for overviews, see Bucklin *et al.*, 2008, 473f or Frazier and Lassar, 1996), none has been found in the financial services industry, considering in particular multichannel distribution strategies where competitor and own market presence are taken into account. Bucklin *et al.* (2008, 474) summarise that in light of the scant evidence, advice to managers has been largely based on theory, logic, and example. However, an older study Clapp *et al.* (1990) used the ratio of agents in a market area divided by aggregate market income to analyse the effects of location on the profitability of life insurance agencies. This measure seems to be attractive, but one has to assume that spatial reach of the agents is uniform across space.

The issues of distribution channel competition and performance have found much more interest in the literature (i.e. Coelho *et al.*, 2003; Coelho and Easingwood, 2004; Frazier, 1999; Gaski, 1984; Neslin *et al.*, 2006; Tsay and Agrawal, 2004; Webb and Hogan, 2002;). As in many other industries, insurers have the challenge to balance channel competition to maximize the premium income and profits. The phenomenon of the coexistence of tied agent and independent agent/broker distribution channel has been several times described in the literature and has been generally explained by either differences in service quality and clientele they attract (*product quality hypothesis*) or by prevailing information asymmetries, a lack of market transparency and other differences between the two distribution channels (*market imperfection hypothesis*) (Trigo-Gamarra, 2008, 390). Additionally, direct channels have been introduced in the more recent past, leading potentially to the cannibalization of other distribution channels used. In practice, there is often resistance of tied agents to any direct channel efforts, as they see them as pure rivals (Eastman *et al.*, 2002; Bannwarth, 2008, 9). However, recent theoretical work on insurance distribution systems (Pfeil *et al.*, 2008, 51) suggest that whether large sales through the direct channel do impact other channels and their profitability depends on the relative magnitude of cannibalization versus market enlargement and the degree to which the increase changes the composition of offline customers' types with respect to service cost. Results from other industries have revealed that the addition of a direct channel through the internet does not have to cannibalize existing channels, at least in the long term (i.e. Avery *et al.*, 2009; Biyalogorsky and Naik, 2003; Deleersnyder *et al.*, 2002), and can

even contribute to both financial and strategic company performance (Wolk and Skiera, 2009). Nevertheless, in a study of 62 U.K. financial services companies, it appeared that multichannel companies enjoyed higher sales levels but lower profits (Coelho *et al.*, 2003, 567). Valos and Vocino (2006, 27) suggest to enhance the integration between market segmentation strategy and channel strategy.

In any case, companies have to direct marketing efforts at the customers who have the highest potential to respond to those efforts in order to use resources effectively. This task includes the identification of the most suitable customers and their place of residence. Research has revealed that acquisition costs and the quality of an acquired customer including its value and retention differ substantially by distribution channel in the insurance business (Verhoef and Donkers, 2005). Therefore, optimisation models for allocation acquisition resources across channels have been suggested to maximize profit for a given budget (Neslin *et al.*, 2006). Nevertheless, their model provides only a global measure and lacks the ability to identify local markets where resource allocation might be particularly promising.

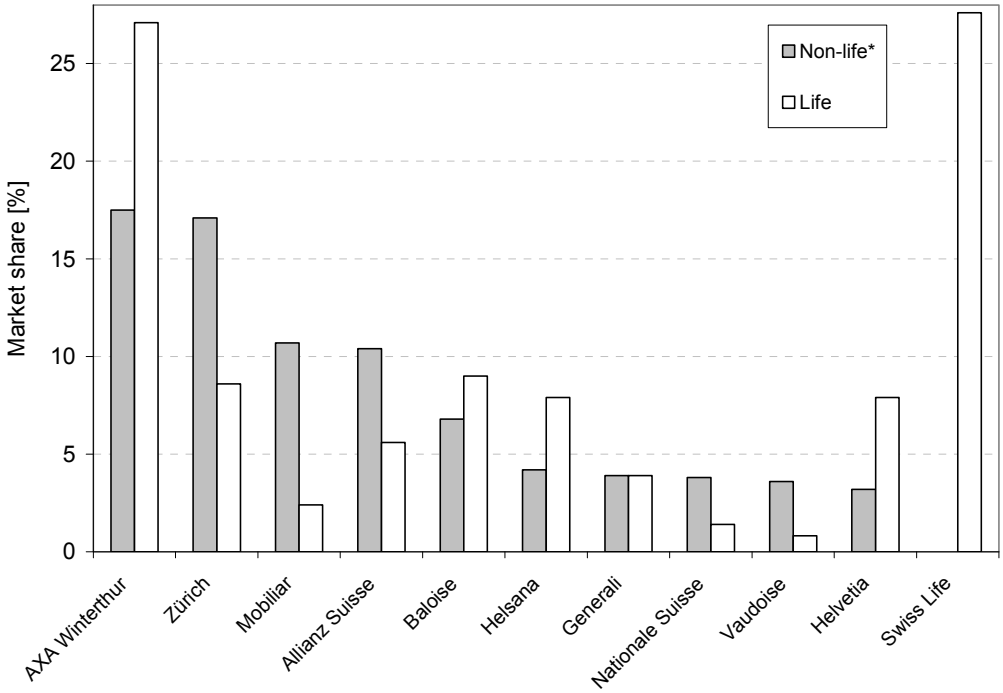
Many channel performance measures have been suggested in the past but particularly financial performance indicators have been criticised of being myopic (Webb and Hogan, 2002, 19). Valos and Vocino (2006, 17) state that current channel performance measurement guidelines are too generic for strategy and marketing managers as well as too reliant on financial measures. Moreover, they describe that research in distribution channels has highlighted a channel performance metric paradox. According to this paradox, it is impossible for a business organisation to maximise all channel performance measures concurrently, because different systems and different channels necessitate particular measurement structure characteristics. Performance measures tailored for single distribution channels have been described from theory and practice (King and Liou, 2004; Löning and Besson, 2002) but approaches for multi-channel strategies are very scarce (Gensler *et al.*, 2007), mostly because comprehensive and cross-channel measures are needed as different channels might benefit from each other, for example when they are coordinated across stages of the customer decision process. In German-speaking countries, it has been found that the internet is often used for information and customer relationship management, but insurance policies are rather sold through face-to-face distribution channels (Hattemer, 2008, 93; Psychonomics, 2007). General evaluation criteria and performance measures such as gaining new customers, customer retention, customer satisfaction, market share, balanced source of revenues, cost of sales, sales growth and profitability have been suggested or empirically detected (Coelho *et al.*, 2003, 561; Löning and Besson, 2002; Vuorinen *et al.*, 1998). Nevertheless, such global measures miss the spatial dimension. For example, Bronnenberg and Albuquerque (2003, 217) suggest to commence by

looking at how distribution and communication channels are structured geographically when investigating the spatial concentration of market share of a company. This study acknowledges these suggestions and applies them to an example from the Swiss insurance market.

3.4 Swiss insurance market and market shares

Switzerland has the highest total insurance expenditure per capita in the world and is a well developed market. Within Europe, Switzerland has by far the highest non-life insurance expenditures per capita with 2450 US-Dollars and the third highest (after United Kingdom and Ireland) life insurance expenditures per capita with 3112 US-Dollars in 2006 (SwissRe, 2007, 38). The average Swiss contract duration is around 10 years (Cap Gemini, 2008). The Swiss market for non-life and life insurances has been steadily deregulated during the 1990s. One of the last set of controls was scrapped in 1996 when the fixed tariff regime for third-party vehicle insurance was abolished (Credit Suisse Economic Research, 2005, 1). The deregulation was followed by the introduction of new services (i.e. assistance insurances, help points, etc.), product variety as well as new distribution channels (direct channels, such as phone and internet distribution as well as brokers). All this made the Swiss insurance market attractive for local as well as international insurance companies and has resulted in intense competition. According to the Federal Office of Private Insurance (FOPI), there were 26 life and 117 non-life insurers active in the market in 2007, which was however dominated by a minority of these companies. Figure 1 shows the ten largest non-life insurers in the Swiss market including both their non-life and life market share with regard to gross written premiums plus Swiss Life as the leading player in the life sector. These national numbers are silent about the market share of the companies by region or municipality within the country.

Figure 1 Market share of major insurer in Switzerland in 2006 (top 10 non-life)



Source: Federal Office of Private Insurance (FOPI), 4th quarter 2006
* excl. non-compulsory individual health insurance

Little research has explored the geographic disaggregation of those nationally aggregated data and their determinants, mostly due to data restrictions. Still, it is very useful for strategy managers and marketers to have knowledge of the processes which generate those aggregate market outcomes. A few studies can be found in the field of consumer packaged goods (Ataman *et al.*, 2007; Bronnenberg *et al.*, 2007; Bronnenberg and Albuquerque, 2003) and banks (Berger and Dick, 2007), focusing on the regional level like metropolitan statistical areas in the US. However, the studies predominantly investigate the advantage of an early market entry and the persistence of geographic differences in market share over time or the geographic variation in response to marketing-mix variables (Dubé and Manchanda, 2005; Lodish, 2007). The available studies do not investigate determinants of geographically differentiated market shares further.

3.5 Methodology and data sources

The main source of data for this study is a major Swiss insurance company. Over the years, it expanded nationally and abroad both by taking over other insurers or by organic growth. In Switzerland, it gained considerable market share in all regions of the country while its posi-

tion remained particularly strong in its initial home market. However, no information is available about the time of entry in the various local markets. In the 1990s the company was the first to introduce direct telephone marketing to the Swiss market, soon added by a full direct distribution via phone and internet with a steadily increasing share of the yearly written new premiums. Despite this multi-channel distribution strategy, involving independent agents and brokers as well, tied agents remain the predominant distribution channel of the company with a total share of about four fifth of all active policies in summer 2007 while the direct distribution channel was responsible for about a tenth.

The company provided the zip code of each customers location in Switzerland, policy type, distribution channel of purchase (including ID of the tied agent if applicable) and yearly premiums of all active policies as of summer 2007. Premiums were assigned to the location of the customer, only considering customers in the individual lines of business and therefore disregarding companies and other institutions. Although the policy information did not include the date of the contract, it can be assumed that many policies have been signed quite some time ago, particularly life insurance policies, as they have a much longer contract period on average. The locations of all tied agents selling individual insurances are known as well. The company is not assigning market areas exclusively to their tied agents, and the raw data distinguish three types of agents in the company data set: commercial agents, individual agents and asset/financial consulting agents. However, for all three types, individual insurance policies constitute the majority of their portfolios. Consequently, all three agent types have been combined for the calculations below. More than 99 percent of the customers and all tied agent locations were successfully geocoded at municipal level.

Data about the total insurance demand in terms of written premiums by individuals from all insurance companies in the national market are available from a data provider[3] at the municipality level, acknowledging that insurance demand differs significantly across Switzerland. The data have been used in this article to calculate the market share of the company for each of its distribution channels by dividing the written premiums found in the company data by the total insurance demand[4]. Additionally, the data have been used as an auxiliary variable to calculate agent density as described below. The data provider has disaggregated the data of the Swiss Household Budget Survey (HBS) from the regional level (Grossregionen) to the municipality level, based on the distribution of household age, household income and house-

[3] provided by bwv its GmbH, St. Gallen

[4] Due to confidentiality issues, data maps of market shares of the company cannot be displayed here.

hold size. Therefore, those variables cannot be included as explanatory variables in the estimates below due to methodological restrictions. Moreover, the data does not allow the study of the effect of distribution intensity on overall market demand, at least not at the municipality level.

3.6 Calculation of an agent density measure

Insurance agents usually travel to their customer's home by car. Since it is certain that distribution and acquisition efforts of agents go well beyond of their home municipality of the partly very small Swiss municipalities, some kind of distribution intensity measure of the agent channel across municipal boundaries is needed. Therefore, the observed proportional origins of premiums by travel time of the company's agents to the location of their customers have been investigated and assumed to be the proportions of time agents are willing to allocate their distribution efforts. For those calculations, a car travel time matrix of all Swiss municipalities based on the national transport model (Vrtic and Fröhlich, 2008) was used. Intra-municipal (-zonal) travel time was calculated based on distance as suggested by Rietveld and Bruinsma (1998, 118):

$$d_m = \sqrt{\frac{\left(\frac{a_m}{\pi}\right) \times (\pi - 1)}{\pi}} \qquad (1)$$

where d is the average distance between two arbitrary points in municipality m and a is the area of municipality m. The distance has been divided by an average assumed speed of 25 km/h. With the resulting values, it is possible to plot the share of premiums by travel time, measured from the municipality of the agents to the origin municipalities of the premiums, based on the following equation:

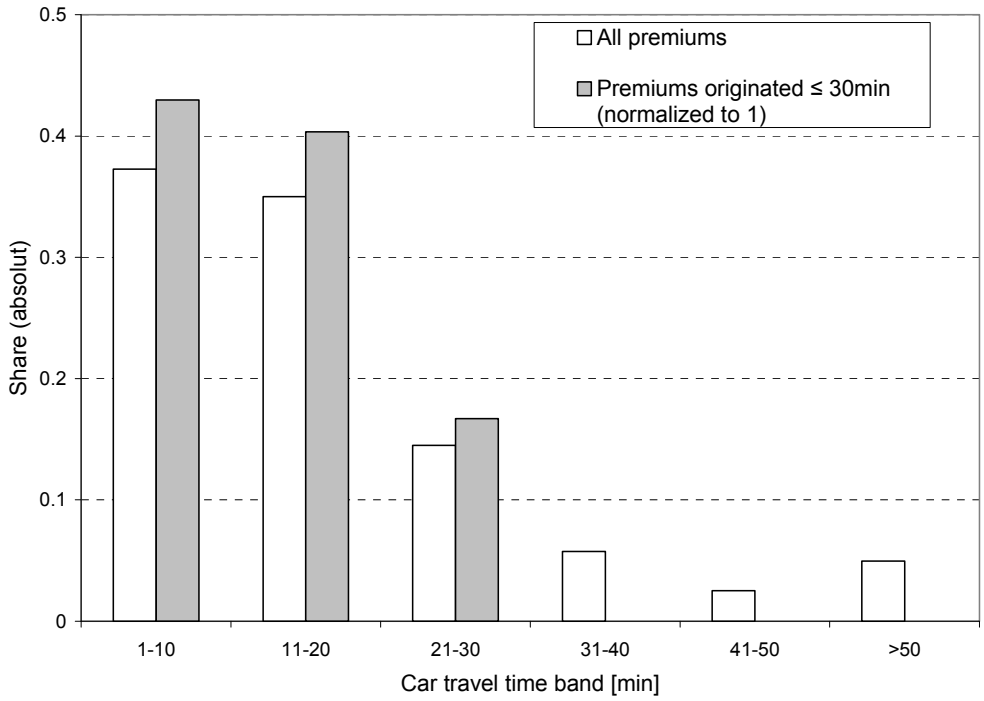
$$Sh_T = \frac{p_T}{p} \qquad \text{with } T = [t_{1\text{-}10}, t_{11\text{-}20}, t_{21\text{-}30}, t_{31\text{-}40}, t_{41\text{-}50}, t_{>50}] \qquad (2)$$

where Sh_T is the share of the company's tied agent premiums p_T in car travel time band T measured in minutes from the agent home municipality to the premium originated municipality which is the same as the customer location. p is the sum of all premiums in the tied agent distribution channel of the company. The resulting shares are shown by the white bars in Figure 2. The analysis reveals that the majority of the premiums originate within a distance of

30 minutes car travel time or less from the agent location, namely more than 87%. However, around 13% of locations, which are originating premiums, are beyond 30 minutes car travel time and 5% beyond even 50 minutes. Earlier micro-analysis with the company's data revealed that customers of an agent are spread over the region and even scattered across the country (Hauri *et al.*, 2008, 28). This reflects the fact that there is often a long-term relationship of the agent to her/his customers which can last for example beyond a move of the customer. Exploratory data analysis also showed that the travel time distributions differ locally, due to variations in market sizes, the spatial distributions, natural barriers, accessibility etc., but they do not show any systematic pattern, therefore a uniform measure of agent density is pursued here.

For calculating agent distribution density, only the observed premiums within 30 minutes car travel time from the agent location in the company data were considered. Travel times beyond 30 minutes are omitted since it is assumed that these are special cases (i.e. the customer and/or the agent has moved since the moment of insurance purchase, the insurance has been purchased at the place of a second home or the home of the customer is exceptionally far away from the next tied agent). Again, Equation 2 has been employed but only considering travel time bands $t_{1\text{-}10}$, $t_{11\text{-}20}$ and $t_{21\text{-}30}$. The results have been normalised to one as shown by the grey bars in Figure 2. Based on the calculations, it is assumed that an agent dedicates 42% of her/his distribution efforts to customers within 10 minutes travel time.

Figure 2 Origin of premiums by car travel time band from the company agent's municipality to the home municipality of the customers



However, this measure would uniformly allocate the same weight to any municipality within a travel time band, disregarding total market demand size and the market selection of the agents. To adjust for this bias, it is assumed that an agent allocates its distribution efforts proportional to the market demand per municipality within each of the three travel time bands. The resulting agent presence of each tied agent was summed up for all tied agents of the company per municipality as follows:

$$AgP_m = \sum_{a=1}^{A} Sh_{Ta} \frac{d_{mT}}{d_T} \tag{3}$$

where AgP_m is the agent presence measure in municipality m, a is an agent located within 30min travel time of the municipality, A is the sum of all agents within 30min travel time, Sh_{Ta} is the weight of agent a in travel band T, d_{mT} is the total demand of all households for insurances at municipality m in travel time band T and d_T is the total demand in all municipalities in travel time band T. The total demand for insurances was taken from the external database mentioned.

This agent presence measure per municipality was divided by the market demand in the municipality in terms of total annual individual non-life or life premiums, again coming from the external database:

$$AgD_{sm} = \frac{AgP_m}{d_{sm}} \tag{4}$$

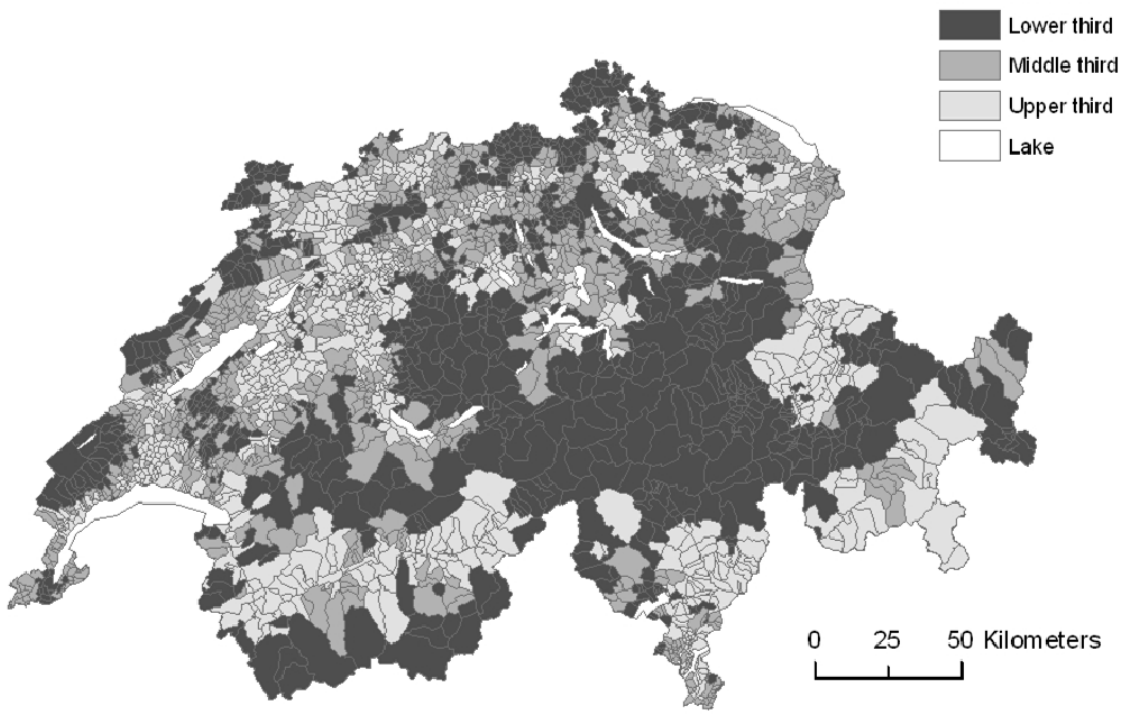
where AgD_{sm} is the agent density in insurance sector s (individual non-life or life) in municipality m, AgP_m the agent presence measure in the municipality as calculated in Equation 3, and d_{sm} the total demand in insurance sector s in municipality m. All company agents of the company are selling both non-life and life insurance policies. Therefore, there is no distinction of insurance sectors in the agent density measure for company's agents. This agent distribution density measure extracts the theoretical efforts all active tied agents are devoting to any local market while adjusting for total market demand. However, this measure does weight all individual agents in the same way, not considering differences in experience, weekly hours of work or functional specifications. One agent might have additional administrative duties or focus on wealth management of customers. Nevertheless, explorative data analysis revealed that a negligible small portion of the company's tied agents do not work full time and that the amount of written premiums is not correlated with the functional job classification.

Clearly, the market share of a company is also a function of the intensity of competition in any local market. Agent location data of all competitors (as of November 2007) were taken from the Swiss Federal Office of Private Insurance (FOPI), which maintains a database of registered agents since 2007. It is mandatory for independent agents and brokers to register and optional for tied agents[5]. The data included zip code as well as the city of the registered agent's work place and was manually enriched by the primary affiliation to an insurer if applicable and the assignment to a municipality through geocoding. If primary affiliation information was present, it was assumed that this relationship is exclusive. There are several insurance companies active with tied agents in the Swiss market, which are offering either non-life or life insurances only. This is why, different than for the uniform agent presence calculation based on the company data, two separate competitors' agent density variables have been generated: one for non-life, leaving out agents exclusively or predominantly working for a company only offering life insurance, and the other for life respectively. For both variables the same spatial diffusion has been hypothesised as for the company's tied agents and the same

[5] Comparing company agent data with the register entries, around 60 percent of the tied agents are registered.

methodology for generating the agent density measure has been applied. In general, it was assumed that the tied agents of the various competitors missing in the database are missing at random, i.e. have the same spatial distribution as those registered. Only agents associated with an insurer have been considered while disregarding independent agents and brokers, as they can potentially sell insurances both from the company and its competitors. The agent density result for competitor non-life agents is mapped in Figure 3. This measure describes the density of non-life agents from all competitors of the company. It is displayed in terciles so that the reader can easily identify areas of comparable low and high competitor non-life tied agent density. Therefore, there is low competition in lower third municipalities and high competition in upper third municipalities for the company from the tied agent channel of the competitors.

Figure 3 Map of competitors non-life tied agent density in Switzerland at municipality level



3.7 Additional variables and descriptive statistics

Some other variables at the municipal level, which proved to have an impact on insurance demand in aggregated empirical insurance demand studies (as described in Section 3.2.1) were considered, namely education as measured in percentage of inhabitants with an university degree as well as vehicle ownership, as its correlation with income is weak. Additional factors, which are usually taken into account in cross-country studies, such as life expectancy,

monetary stability, interest rate, social security, insurance regulation etc., are not useful here as there is no variation within the country. Consequently, they are omitted from the specifications. Finally, a set of dummy variables for the seven statistical regions (Grossregionen) of the country, as defined by the Swiss Federal Statistical Office, have been considered.

Due to different data sets referring to varying years and frequent municipal boundary changes over time due to municipal mergers in Switzerland, all data were adjusted to the municipal boundary system as of January 2007 with a total of 2721 municipalities. The complete set of variables used for the final estimations is summarised in Table 1 and their correlations are given in Table 2. More variables were tested, but were either insignificant or did not improve the models presented in Section 3.8 (see Appendix for a list of tested variables). Both the non-life and life market share in the tied agent distribution channel of the company show significant spatial dependence, indicated by a significant Moran's I (Moran, 1950) of 0.306 for non-life and 0.111 for life.

Table 1 Descriptive statistics of variables at municipality level ($N = 2721$)

Variable	Abbrevation	Type[1]	Min	Max	Mean	S.D.
Company non-life market share in tied agent distrib. channel	MShNAg	C	0.000	0.340	0.068	0.033
Company non-life tied agent density	AgDN	C	0.000	0.036	0.001	0.001
Competitors non-life tied agent density	CtAgDN	C	0.000	0.035	0.003	0.002
Company non-life market share in broker distrib. channel.	MShNBr	C	0.000	0.091	0.005	0.007
Company non-life market share in direct distribution channel	MShNDi	C	0.000	0.051	0.008	0.006
Company life market share in tied agent distrib. channel	MShLAg	C	0.000	0.265	0.042	0.029
Company life tied agent density	AgDL	C	0.000	0.055	0.001	0.001
Competitors life tied agent density	CtAgDL	C	0.000	0.033	0.004	0.003
Company life market share in broker distribution channel	MShLBr	C	0.000	0.357	0.005	0.011
Company life market share in direct distribution channel	MShLDi	C	0.000	0.162	0.006	0.009
Vehicles per inhabitant	VepC	C	0.000	1.631	0.559	0.103
Percentage of inhabitants with university degree	UniDeg	C	0.000	0.455	0.096	0.064
Grossregion Lemanique	Regio1	D	0.000	1.000		
Grossregion Espace Mitteland	Regio2	D	0.000	1.000		
Grossregion Northwest Switzerland	Regio3	D	0.000	1.000		
Grossregion Zürich[2]	Regio4	D	0.000	1.000		
Grossregion East Switzerland	Regio5	D	0.000	1.000		
Grossregion Central Switzerland	Regio6	D	0.000	1.000		
Grossregion Ticino	Regio7	D	0.000	1.000		

[1] C = continues; D = dummy
[2] Reference case

Table 2 Correlation matrix of variables

Variable	MshNAg	AgDN	CtAgDN	MShNBr	MshNDi	MShLAg	AgDL	CtAgDL	MShLBr	MshLDi	VepC	UniDeg	Regio1	Regio2	Regio3	Regio4	Regio5	Regio6	Regio7
MshNAg	1.00	**0.11**	-0.02	**0.08**	**0.18**	**0.43**	**0.11**	-0.02	0.02	**0.11**	**0.19**	**-0.18**	**-0.17**	0.01	-0.04	**0.10**	**0.10**	**-0.08**	**0.13**
AgDN		1.00	**0.38**	**0.05**	**0.06**	**0.08**	**0.99**	**0.39**	**0.05**	**0.06**	**0.07**	**0.05**	-0.02	**-0.04**	-0.01	0.03	-0.02	0.03	**0.10**
CtAgDN			1.00	0.02	0.00	0.03	**0.37**	**0.94**	0.00	0.01	**0.07**	**0.09**	**0.09**	**0.04**	**-0.08**	**-0.06**	**-0.07**	-0.04	**0.07**
MShNBr				1.00	**0.09**	**0.06**	**0.04**	0.01	**0.21**	**0.09**	**0.11**	**0.16**	-0.03	**-0.15**	0.03	**0.09**	0.02	-0.03	**0.18**
MshNDi					1.00	**0.10**	**0.05**	0.00	**0.08**	**0.36**	**0.12**	**0.17**	**-0.09**	**-0.20**	**0.19**	**0.43**	0.00	**-0.08**	**-0.07**
MShLAg						1.00	**0.08**	**0.04**	0.03	**0.12**	**0.15**	0.01	**-0.04**	0.00	-0.01	-0.02	0.00	**-0.04**	**0.14**
AgDL							1.00	**0.39**	**0.05**	**0.05**	**0.06**	0.04	-0.02	**-0.04**	-0.02	0.02	-0.02	0.03	**0.10**
CtAgDL								1.00	**-0.02**	**0.01**	**0.09**	**0.11**	**0.12**	**0.05**	**-0.11**	**-0.06**	**-0.10**	**-0.06**	**0.12**
MShLBr									1.00	0.02	0.03	**0.05**	-0.02	**-0.07**	0.01	**0.09**	0.01	0.01	0.03
MshLDi										1.00	**0.05**	**0.13**	-0.01	**-0.09**	0.00	**0.23**	-0.01	-0.01	-0.02
VepC											1.00	0.16	**0.12**	**0.00**	0.03	0.01	**-0.20**	**-0.11**	**0.16**
UniDeg												1.00	**0.34**	**-0.18**	-0.01	**0.12**	**-0.19**	**-0.11**	**0.06**
Regio1													1.00	**-0.35**	**-0.19**	**-0.13**	**-0.23**	**-0.14**	**-0.14**
Regio2														1.00	**-0.24**	**-0.17**	**-0.3.**	**-0.17**	**-0.18**
Regio3															1.00	**-0.09**	**-0.16**	**-0.10**	**-0.10**
Regio4																1.00	**-0.12**	**-0.07**	**-0.07**
Regio5																	1.00	**-0.12**	**-0.12**
Regio6																		1.00	**-0.07**
Regio7																			1.00

Bold: significant at 0.05 level

3.8 Estimates

It is hypothesised that the market share of the company in terms of individual premiums of the tied agent distribution channel in a municipality is systematically related to both the density of the company's and competitors' tied agents assuming that they are seeking to serve the same set of customers. Moreover, it is assumed that this tied agent market share is related to the market share of other distribution channels of the company and that every municipality constitutes a local insurance market. Nevertheless, one has to keep in mind that municipal data are aggregated according to rather arbitrary but historical administrative delimitations of territory which are, however, usually homogeneous with regard to language, legal regulations, tax system and so forth. Consequently, spatial effects might be present and have to be tested before relying on cross-section estimations. But in a first instance, a cross-section ordinary least square (OLS) regression is estimated where market share of the company's tied agent distribution channel has been selected as the dependent variable.

A logit transformation has been applied to the market share variables as suggested by Mosteller and Tukey (1977, 109). This transformation stretches the tails of the dependent variable of market share (originally 0 to1):

$$z = \ln\left(\frac{p}{1-p}\right) \tag{6}$$

where p is the market share and z the transformed variable.

Logarithmic transformations were applied to the tied agent density measures. A quadratic term (for example, Reibstein and Farris, 1995, suggest a generalised convex cross-sectional relationship between retail distribution and market share) did not improve the models significantly. Since cannibalization effects of the broker and the direct distribution channel on the tied agent channel performance of the company might be present, their market shares are included in the estimates as independent variables. The data of the company's broker channel market share comprises both broker and independent agents.

With regard to the seven statistical regions (Grossregionen) of the country, the Zürich region, the biggest market, is the reference case in the estimates. The dummy variables are testing if there is any systematic over- or underperformance in the regions while controlling for all other independent variables in the models.

Separate OLS models were estimated for individual non-life (Table 3) and life market share (Table 4) in the tied agent channel of the company. In both models, the variance inflation factors (VIF) are below two for all independent variables, except for the regional dummies, indicating no problems with multicollinearity. Apparently, the tied agent non-life market share is dependent on the chosen factors as the significance of most variables at the 0.05 level and a moderate adjusted R-square of 0.174 indicate. The life model reveals a lower adjusted R-square of 0.125. Therefore, the non-life sector model achieves a better fit than the life model. It might be surprising that the life market seems to be less 'spatial' than the non-life market since in general life insurances tend to be sold more through face-to-face distribution channels than non-life insurances. The reason is probably that life policies usually have a longer duration, meaning that the spatial relationship of agent density and customer location is relaxed due to relocations of the customer and/or the agent over time. Additional factors might be omitted variables and the generally higher market penetration in the non-life market of the company, indicated by higher and more evenly distributed market shares of the non-life tied agent channel of the company compared to the life tied agent channel. As can be observed in Table 5, the mean market share of the non-life tied agent distribution channel of the company is 2.6 percent points higher than the market share in the life tied agent channel, while both variables have almost the same standard deviation.

But the Moran's I test of the OLS residuals indicated spatial autocorrelation for both models, based on a queen type contiguity matrix. The Moran's I for the non-life OLS models had a value of 0.172 and the life OLS a value of 0.051, both highly significant. Therefore, the Lagrange Multiplier test (Breusch and Pagan, 1980) was applied. The results indicated a spatial lag model (Anselin, 1988, 34), which can be written as follows.

$$y = \rho Wy + X\beta + \varepsilon \tag{7}$$

where y is a N by 1 vector of market shares of the company's tied agent distribution channel (N is the number of municipalities), ρ is a spatial autoregressive parameter, Wy is the corresponding spatially lagged dependent variable for weights matrix W, X is a N by K matrix with observations on municipality characteristics K, β is a K by 1 vector of regression coefficients and ε is a N by 1 vector of error terms.

Comparing the OLS results with the spatial lag models, the models for both non-life and life have improved measures of fit. An increase in the log-likelihood value as well as the decrease of the Akaike Information Criterion (AIC) and the Schwartz Criterion (BIC) all suggest an improvement of fit for the spatial lag specification for both non-life and life. Moreover, most

coefficients are becoming lower or less negative with the spatial lag specification except for the non-life broker channel market share and some of the regional dummy variables.

In all models, the coefficient of the tied agent density of the company is positive and the competitor's density negative. Therefore, the coefficients show the expected result and a higher company tied agent density increases whereas a higher competitors tied agent density decreases the tied agent market share of the company. The company might be less vulnerable in the life sector to additional competitor tied agents indicated by the low values of the competitor tied agent density (CtAgDL) compared to the CtAgDN parameters in the non-life sector, but the coefficients are insignificant both in the OLS and the spatial lag model.

Remarkably, both the company's market share in the broker and direct distribution channel are significant and have positive signs for non-life and life, suggesting that a high market share in these distribution channels leads to higher market share in the tied agent distribution channel. Although the relationship could be vice versa or actually interact in both directions, the results cast strong doubts on the notion that the direct and broker channels impact the tied agent channel negatively. Apparently, the company manages to coordinate the channel strategies by targeting different customer groups or the like. The distribution channels might also benefit from each other for example through word-of-mouth reputation and visibility. While the coefficient values are roughly the same for the direct distribution channel in non-life and life, the values are lower in the broker distribution channel for non-life insurances compared to life.

The company tends to have higher tied agent market shares in the non-life insurance sector in municipalities with high vehicle ownership, possibly because of its particular strong market position in car insurance, and marginally lower market shares in municipalities with a high percentage of well-educated people with college or university degree. The same positive relationship between tied agent market share and vehicle ownership can be observed for the life sector, but there is also a positive relationship with regard to percentage of well-educated people with college or university degree. Therefore, the agents seem to achieve a higher market share in municipalities with more well-educated people, who usually have an above average income.

Further insights into the spatial distribution of the market shares are given by the coefficients of the regional dummy variables. While all significant coefficients are negative in the non-life spatial lag model, all coefficients are positive in the life sector. Therefore, the market share of the tied agents of the company is weaker in several regions beyond the reference region of

(Grossregion) Zürich in the non-life sector but stronger in the life sector. Apparently, even the regional market share distribution in the tied agent channel differs considerably among the non-life and life sector.

Table 3 Estimates for individual non-life market share in the tied agent distribution channel of the company ($N = 2721$)

y = MShNAg (logit)	OLS				Spatial Lag Model		
Independent variable	b	beta	sign.	VIF	b	beta	sign.
Constant	-2.192		***		-1.454		***
Ln(AgDN)	0.130	0.148	***	1.747	0.090	0.102	***
Ln(CtAgDN)	-0.057	-0.086	***	1.763	-0.041	-0.062	***
MShNBr (logit)	0.023	0.069	***	1.253	0.024	0.072	***
MShNDi (logit)	0.090	0.205	***	1.271	0.081	0.185	***
VepC	0.939	0.178	***	1.120	0.907	0.172	***
UniDeg (logit)	-0.122	-0.177	***	1.386	-0.102	-0.148	***
Regio 1	-0.289	-0.217	***	3.794	-0.168	-0.126	***
Regio2	-0.125	-0.107	***	4.610	-0.053	-0.045	
Regio3	-0.228	-0.135	***	2.568	-0.137	-0.081	***
Regio5	-0.008	-0.006		3.305	0.023	0.016	
Regio6	-0.275	-0.124	***	2.018	-0.151	-0.068	***
Regio7	0.054	0.025		2.105	0.049	0.023	
rho					0.359		***
F-Test	48.720		***				
Adj. R-square	0.174						
Log likelihood	-1932.860				-1835.150		
AIC	3893.719				3700.290		
BIC	3976.442				3788.921		
Moran's I	0.172		***		-0.013		

Probability of rejecting H_0 = *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 4 Estimates for individual life market share in the tied agent distribution channel of the company ($N = 2721$)

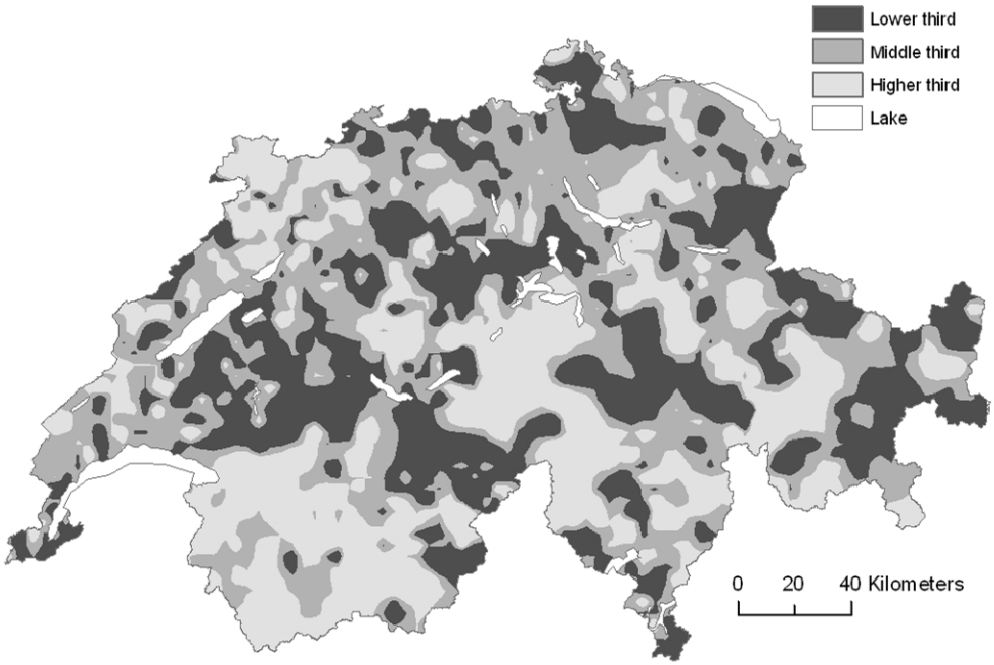
y = MShLAg (logit)	OLS				Spatial Lag Model		
Independent variable	b	beta	sign.	VIF	b	beta	sign.
Constant	-1.813		***		-1.541		***
Ln(AgDL)	0.198	0.115	***	1.689	0.172	0.100	***
Ln(CtAgDL)	-0.019	-0.016		1.754	-0.017	-0.014	
MShLBr (logit)	0.055	0.103	***	1.253	0.050	0.093	***
MShLDi (logit)	0.103	0.196	***	1.249	0.098	0.188	***
VepC	1.269	0.124	***	1.106	1.235	0.121	***
UniDeg (logit)	0.096	0.072	***	1.359	0.092	0.069	***
Regio 1	0.051	0.020		3.874	0.064	0.025	*
Regio2	0.276	0.121	***	4.615	0.271	0.119	
Regio3	0.204	0.062	**	2.598	0.199	0.061	*
Regio5	0.184	0.066	**	3.330	0.197	0.070	*
Regio6	0.112	0.026		1.987	0.125	0.029	**
Regio7	0.296	0.072	***	2.071	0.267	0.065	
rho					0.143		
F-Test	33.370		***				
Adj. R-square	0.125						
Log likelihood	-3802.866				-3790.815		
AIC	7633.733				7611.631		
BIC	7716.455				7700.262		
Moran's I	0.051		***		-0.007		

Probability of rejecting H_0 = *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

By mapping the residuals of these models, it is possible to identify local markets and regions where the tied agent distribution channel of the company is either over- or underperforming. As an example, the residuals of the non-life spatial lag model are mapped in Figure 4. Due to confidentiality restrictions, the data presented have been interpolated by an inverse distance weight measure (Lloyd, 2007, 98) based on the residual values at the centroid of each municipality. Actually, the models provide metric values for every single municipality. In the figure, the data have been grouped into terciles. Areas in the lower third indicate that the actual tied agent distribution market share is lower than predicted by the model, while areas in the upper third show higher market shares than predicted. Therefore, company's tied-agent distribution

channel performance is rather weak for areas in the lower third and strong in the upper third, considering own to competitor presence as well as the other variables in the model. Local and regional variation across the country can be observed, and it is well below the spatial scale of the regions considered by the dummy variables in the models.

Figure 4 Smoothed residuals from the individual non-life spatial lag model



3.9 Conclusions

Behind aggregated national market share figures of nationally operating companies, local market shares can vary considerably within a country. Here, an example from the insurance industry was used the municipality level in Switzerland, applying business data from a major insurance company. In particular the market share of the tied agent distribution channel was analysed at municipal level for all of Switzerland. It was revealed that the market share is influenced by certain characteristics of the local markets. Therefore, differences in market shares can be explained by considering some crucial local factors. It can be stated that the non-life and life insurance market are spatially organised, meaning that the market share in the tied agent distribution channel follows spatial principles, explained by the significant independent variables in the models at the municipality level.

A significant positive relationship between the company's tied agent density and market share in the tied agent distribution channel was verified by developing a density measure based on

revealed company data which takes into account the fact that the distribution area of an agent goes well beyond its home municipality. As expected, a strong tied agent workforce helps to increase market share in the tied agent distribution channel whereas strong competition due to a high density of competitors tied agents certainly makes it more difficult to gain market share. With regard to channel competition, no indication of cannibalization has been found. The different distribution channels of the company even seem to be complementary as the correlation matrix (Table 2) and the model estimates (Table 3 and Table 4) suggest that higher market shares in the broker and direct channels lead to higher market shares in the tied agent channel, although the cause-and-effect chain could be vice versa.

However, it has to be noted that several limitations apply to the analysis presented. First of all, the model fit particularly of the life model is relatively low. Apparently, there are factors other than the ones considered and unavailable for this study, which determine the performance. On the supply side, regional performance differences in the tied agent channel might be present due to varying accessibilities leading potentially to longer travel times in some areas of Switzerland where the agents have to spend more time to reach their customers compared to other areas. Additionally, varying effectiveness among the tied agents might be present, which is certainly affected by experience but could also be due to variation of what the literature describes as sales person service behaviour, including factors such as diligence, information communication, inducement, empathy and sportsmanship (Ahearne *et al.*, 2007). Another limitation is the lack of information about spatially varying pricing policies, promotions and advertisement campaigns both of the company and its competitors. However, the company claims that there are no locally varying price regimes after considering varying risk levels.

On the demand side, there might be additional influences present as well, which were not captured in the models. For the retail market, Bronnenberg and Mahajan (2001, 286) describe that differences in demand can be caused by sheer inertia of initial market conditions, local order-of-entry effects, or can simply reflect regional consumer tastes. However, the time of local market entries are unknown and unavailable for both the company and the competitors, which forms a major limitation of the analysis.

By introducing marketing-related variables to the models, such as marketing expenditure in local markets, including billboards, event sponsoring, regional press advertisement or TV spots etc., it would potentially become possible to evaluate and forecast marketing actions of the company. The analysis focuses on the investigation of an individual company. Certainly, it would be helpful to include disaggregated customer data of the competitors of the company. But it is rather unlikely that data of this spatial detail will ever become available for multiple

market actors. Finally, only a static analysis based on cross-sectional data was performed, consequently ignoring industry trends. Further insights might be revealed by repeating the analysis at a later point in time or by a longitudinal analysis. Overall, the present data analysis would not be suitable as a single performance measure due to the mentioned data restrictions although to some degree the models proved to be a helpful instrument to identify determinants as well as local areas of strength and weaknesses for the tied agent channel of the company.

Agent presence/density of the company is the only variable in the model which is directly controllable by the company's managers, while they might have only indirect control over the market share from the broker and direct distribution channels. However, maps of the resulting residuals of the models give suggestions of market success of the tied agent channel. There might be specific local reasons for underperformance, but it is up to decision makers to use those maps and data in strategic management and marketing. Particularly the insight that the different distribution channels seem to be complementary but also the measurement of the tied agent densities effect allow development of geographically differentiated multi-channel strategies and optimisation of resource allocation of marketing efforts and across distribution channels.

Finally, this example from the financial services industry revealed that spatial analysis and methods can help in various areas of strategic management and marketing, encouraging the use and exploitation of the spatial reference information of business data to improve the investigation of the spatial dimension of the unit of analysis.

4 Modelling hedonic residential rents for land use and transport simulation while considering spatial effects

4.1 Introduction

One of the key components of UrbanSim is the use of land or real estate price data. These are applied in the model system as an indicator of the relative market valuations for attributes of housing, non-residential space, and location (Waddell *et al.*, 2003, 63). However, finding suitable data sources of real estate transaction prices and rents might be a challenge while setting up an UrbanSim application when transaction and rent price data from data suppliers are unavailable to researchers. As a minimum prerequisite for modelling purposes, sufficient information about a decent amount of properties is needed, including price (transaction cost or rent) and some explanatory locational variables the model system should be sensitive to in the application. These typically include at least some kind of regional accessibility, proximities and neighbourhood characteristics. The hedonic approach is a suitable method to model price values for every cell or parcel in the UrbanSim application based on such a sample. Moreover, the analysis exposes the implicit prices of housing and location characteristics. In the case of the UrbanSim application for Zürich (Löchl *et al.*, 2007), tax assessor data and data from commercial sources were unavailable. Therefore, an alternative method of data acquisition was needed. One pillar of those efforts included a household survey whose purpose was manifold. For the hedonic modelling of residential rents, its sampling strategy turned out to be too clustered, resulting in insufficient variance of the locational explanatory variables. Another pillar was the use of a publicly available web-based portal of residential asking rents. The data were collected for the area of Canton Zürich from the end of 2004 to fall of 2005. It was used to generate a basic hedonic model for the first application of UrbanSim in the Greater Zürich area. However, further analysis of the data revealed the need to consider spatial effects and the introduction of additional explanatory variables.

The hedonic approach was not adopted in the field of housing and real estate before the work of Lancaster (1966 and 1971), Griliches (1971) and Rosen (1974). Today, it is regularly used in the field of real estate for property taxation and mortgage underwriting, but it has also been used for property price generation in land use and transport models (i.e. Waddell and Ulfarsson, 2003).

Location is essential for determining housing prices. Bitter *et al.* (2007, 7) note that controlling for location and the spatial structure of markets is essential to explaining price differentials and deriving accurate coefficient estimates in hedonic residential price models. One common way to incorporate information about location in hedonic models is to introduce distance to the central business district (CBD) or sub-market indicators including regional, local or neighbourhood specific binary-coded dummy variables or interaction terms into the regression equation. However, previous studies have revealed that inclusion does not necessarily take all of the spatial effects into account (Clark, 2007, 189; Willhelmsson, 2002, 100). Two types of spatial effects have been identified: spatial dependence and spatial heterogeneity (Anselin, 1988, 8). These have been major challenges in spatial data analysis (i.e. Du and Mulley, 2006, 201). Anselin (1988, 12) defines spatial dependence, also called spatial autocorrelation, as the existence of a functional relationship between what happens at one point in space and what happens elsewhere. Spatial heterogeneity (or spatial non-stationarity) may be present when there is a lack of uniformity from the effects of space or the spatial units of observation are not homogeneous. For example, price contributions of housing attributes may not be constant throughout a study area and may vary over space. Therefore, there may be spatial heteroscedasticity or spatially varying parameters present. Páez *et al.* (2008, 1566) summarise dependency as a locational/adjacent effect and heterogeneity as market segmentation. In general, housing markets often involve both spatial dependence and spatial heterogeneity due to localised supply and demand imbalances (Bitter *et al.*, 2007, 8). De Graaff *et al.* (2001, 259) list three reasons as to why spatial dependence and heterogeneity should be considered jointly. First, there may be no differences between heterogeneity and dependence in an observational sense. Second, spatial dependency induces a particular form of heteroscedasticity (see also Kelejian and Robinson, 2000). Finally, it may be empirically difficult to separate the two effects. Overall, failure to incorporate spatial effects will result in biased or misleading coefficients and a loss of explanatory power.

For more than two decades, researchers and practitioners have considered spatial effects in hedonic regression models. Several advanced methods have been proposed to incorporate spatial structural instability, spatial drift and spatial lag into models (Leung *et al.*, 2000, 10). These models commonly propose to make use of the spatial characteristics of variables to improve the results (Gao *et al.*, 2006, 1040). Some of the most popular approaches are the spatially adaptive filtering methods (Trigg and Leach, 1967; Widrow and Hoff, 1960;), expansion methods (Fotheringham and Pitts, 1995), multi-level approaches (Jones, 1991; Goldstein, 1987), spatial simultaneous autoregressive approaches such as spatial lag and spatial error models (Anselin, 1988) as well as geographically weighted regression (GWR) models (Fotheringham *et al.*, 2002). Beyond that, there are, for example, the local kriging and co-kriging

methodology of Haas (1995, 1996) and the Bayesian spatially varying coefficient process models of Gelfand *et al.* (2003). Each of these approaches has its benefits and drawbacks, but all emphasise that parameters identified in global models may not resemble parameters estimated in local models. Therefore, they are often non-stationary.

It is the aim of this article to comparatively analyse outcomes from spatial simultaneous autoregressive approaches with results from GWR models that use the same data set, and furthermore, to suggest one method for the housing price modelling for the next application of UrbanSim in the Greater Zürich area. At the same time, the chapter seeks to add to the discussion about appropriate methods for the consideration of spatial effects and non-stationarity in hedonic rent models. Moreover, this chapter describes an innovative way to acquire publicly available data and highlights possible further development potentials of UrbanSim applications. The focus is on residential rents in this article due to the fact that Switzerland has (as compared with the US and other European countries) a very low ownership rate of 37 percent in 2006 (Wüest & Partner, 2007, 76), which is even lower in Swiss urban areas. The reasons for the low ownership rate are many, including low land supply, rigorous mortgage down payment constraints, and tax regulation and landlord-tenant laws. But foremost, prices are high relative to rents and relative to household incomes and wealth (Bourassa and Hoesli, 2010).

While there are many international studies on hedonic house and apartment transaction prices (for a comprehensive review of approximately 125 studies, see Sirmans *et al.*, 2005), only few studies have focused on residential rents. Those rent studies either use the net annual or monthly rent (Baranzini *et al.*, 2006; Baranzini and Ramirez, 2005; Baranzini and Schaerer, 2007; Brunauer *et al.*, 2009) or the monthly gross rent (Banfi *et al.*, 2006). For other hedonic rent studies, it is not specified if gross or net rents have been used (i.e. Djurdjevic *et al.*, 2008; Sirmans *et al.*, 1989; Valente *et al.*, 2005). Of all those studies, only a minority has focused on rents per square meter (Brunauer *et al.*, 2009), while the vast majority have been using the absolute rent or its log-transformation.

The remainder of this chapter is organised as follows: Section 4.2 introduces the methods of spatial simultaneous autoregressive models and geographically weighted regression. Section 4.3 gives a short overview of earlier studies, which compared the results of various hedonic modelling techniques. The data are introduced and described in Section 4.4, followed by the estimation results in Section 4.5. In Section 4.6, the findings are summarised.

4.2 Spatial simultaneous autoregressive models and geographically weighted regression (GWR)

Spatial simultaneous autoregressive modelling is a popular approach to consider spatial effects. It assumes that the response variable at each location is a function not only of the explanatory variable at that location, but of the response at neighbouring locations as well. The models are based on maximum-likelihood estimations and commonly applied in the fields of regional science, sociology, political science and the various fields of economics (Anselin, 2001, 310). A comprehensive introduction might be found in Anselin (1988). Various studies are also available in the field of real estate appraisal, such as Kim *et al.* (2003), Shin *et al.* (2007) and Willhelmsson (2002).

Three different spatial simultaneous autoregressive models are usually distinguished, depending on where the autoregressive process is believed to occur (Kissling and Carl, 2008, 61). The spatial simultaneous autoregressive lag model (SARlag) assumes that the autoregressive process occurs in the response variable. A spatial lag hedonic rent model can be written as follows:

$$P = \rho WP + \beta X + \varepsilon \qquad (8)$$

where P is a vector of rents, ρ is a spatial autocorrelation parameter, W is a $N \times N$ spatial weight matrix (where N is the number of observations), β is a vector of regression coefficients**,** X is a matrix with observations on structural and spatial explanatory characteristics and ε is assumed to be a vector of independent and identically distributed (iid) error terms. Typically, the definition of neighbours used in the weights matrix is based on a notion of distance decay or contiguity.

When spatial dependence is present in the error term, a spatial autoregressive specification for this dependence is usually assumed. This is called the spatial simultaneous autoregressive error model (SARerr) and can be formulated as follows:

$$P = \beta X + u \qquad (9)$$

i.e., a linear regression with error vector u, and

$$u = \lambda Wu + \varepsilon \qquad (10)$$

where λ is the spatial autoregressive coefficient, W is the spatial weight matrix, and u is assumed to be a vector of independent and identically distributed errors. This model is a special case of a regression specification with a non-spherical error variance-covariance matrix. Therefore, W now pertains to shocks in the unobserved variables (the errors u) but not to the explanatory variables of the model (X). Consequently, the price at any location is a function of the local characteristics but also of the omitted variables at neighbouring locations. This is the most popular spatial simultaneous autoregressive model and it is widely used in the literature (Taylor, 2008, 25; Willhelmsson 2002, 94).

Finally, spatial autocorrelation can affect both response and explanatory variables, having both 'inherent spatial autocorrelation' and 'induced spatial dependence'. In this case, an additional term ($WX\gamma$) must appear in the model, which describes the autoregression coefficient (γ) of the spatially lagged explanatory variables (WX) as Kissling and Carl (2008, 61) describe. The so-called spatial Durbin model (SARmix) takes the form

$$P = \rho WP + \beta X + WX\gamma + \varepsilon \qquad (11)$$

According to Valente *et al.* (2005, 110), one of the advantages of these models is a (nearest) neighbour-based smoothing of the means and convenient computation besides the often shown improvement over OLS models.

The technique of GWR has been developed by Brunsdon *et al.* (1998). A more comprehensive overview might be found in Fotheringham *et al.* (2002). The method attempts to incorporate geographical information into a regression model using a series of distance-related weights. Essentially, it consists of a series of locally linear regressions that utilize distance-weighted overlapping samples of the data (Farber and Yeates, 2006, 412). The method explicitly allows parameter estimates to vary over space which leads to independent spatial error terms. Rather than specifying a single global model to characterize the entire housing market, GWR estimates a separate model for each data point and weights observations by their distance to this point, thus allowing unique marginal price estimates at each location (Bitter *et al.*, 2007, 10).

The typical output from a GWR model is a set of parameters that can be mapped in the geographic space to represent non-stationarity or parameter drift. Similarly, local measures of standard errors and goodness-of-fit statistics are obtained (Fotheringham *et al.*, 2000, 113). Therefore, the additional benefit of the GWR approach is that it offers the potential of increased understanding of the nature of varying relationships between variables across space.

Du and Mulley (2006, 201) describe GWR as an alternative to spatial lsimultaneous autoregressive models which is perhaps more intuitive. The GWR approach has recently found use in various applications. There are examples in the fields of climatology (Brunsdon *et al.*, 2001), ecology (Zhang and Shi, 2004; Kimsey *et al.*, 2008), education (Fotheringham *et al.*, 2001), marketing research (Mittal *et al.*, 2004), regional science (Huang and Leung, 2002), political science (Calvo and Escolar, 2003) and transport research (Chow *et al.*, 2006; Clark, 2007; Hadayeghi *et al.*, 2003; Lloyd and Shuttleworth, 2005; Nakaya, 2001). In the housing field, there are studies for example by Bitter *et al.* (2007), Farber and Yeates (2006), Fotheringham *et al.* (2002), Kestens *et al.* (2006), Páez *et al.* (2008) and Yu *et al.* (2007). To our knowledge, there have been no studies on residential rents employing GWR so far as all of the above mentioned authors focus on apartments or single-family house transaction prices. Overall, GWR is considered a standard tool in exploratory spatial data analysis due to its effectiveness and wide applications (Wang *et al.*, 2008, 987). Nevertheless, there are some limitations to the method and authors have pointed out that GWR results should be interpreted with caution (Shearmur *et al.*, 2007, 701). In particular, multicollinearity and correlation among local regression coefficients is a problem in GWR, even in the presence of uncorrelated exogenous variables in the data generating process. The issue was raised by Wheeler and Tiefelsdorf (2005), who emphasised that the effects of multicollinearity are substantially stronger in the GWR model than in global regression models. They suggested that this potentially invalidates any interpretation of individual GWR parameter estimates and reduces confidence in the method for more than exploratory purposes – for example, there might be inflated variances and at times counterintuitive and contradictory in sign to the global regression estimates. Young *et al.* (2008, 4013) notes that local regression models are designed as exploratory smoothing methods and not as inferential statistical tools.

Therefore, several authors have compared the GWR models with alternatives in this regard, in particular with methodological improvements of the GWR approach (Griffith, 2008; Wheeler, 2007 and 2009) and Bayesian spatially varying coefficient process models (Waller *et al.*, 2007; Wheeler and Waller, 2009). Waller *et al.* (2007, 585) note that more methodological research is required to determine the most flexible and robust structures. Additionally, the computational burden of alternative methods might be high (Wheeler and Waller, 2009, 21). Overall, the approaches are still in an experimental phase and the debate about most the most appropriate methodology is ongoing.

A weighting function is applied in GWR in order to give greater influence to close data points. A spatial kernel is usually used. Choosing the shape of the kernel and its bandwidth are both important issues and various options are available (Fotheringham *et al.*, 2002, 56ff).

The bandwidth choice is not a parameter relating to the model itself, but is essentially part of the calibration strategy for a given sample (Fotheringham *et al.*, 2002, 63).

In the case of rent estimations, a GWR model can be written as follows:

$$P_i = \beta_{i0} + \sum_k \beta_{ik} X_{ik} + \varepsilon_i \qquad (12)$$

where P_i is the ith observation of the rent, β_{i0} is the constant estimated for local regression i, β_{ik} is the regression coefficient of structural or spatial explanatory variable k estimated for local regression i, and ε_i is the ith value of a normally distributed error vector with mean equal to zero. This differs from OLS by using distinct constants and regression parameters for each data point. The estimation algorithm iterates through N OLS, each one modified by a unique distance-decay weight matrix. The estimation takes the form:

$$\beta_i = (X^T W_i X)^{-1} X^T W_i P \qquad (13)$$

where β_i is the vector of estimated coefficients for observation i, X is the N x K matrix of explanatory variables, W_i is a diagonal distance-decay weight matrix customised for i's location relative to the surrounding observations and P is the vector of observed rents.

4.3 Comparative studies in the literature

Recently, hedonic literature has shifted from simple applications of local models to comparative studies, where several local model approaches are tested and evaluated. Because of the increasing popularity of the approaches, the need for assessing the relative merits of the different modelling techniques is obvious (Páez *et al.*, 2008, 1566). Consequently, comparisons of the performance of different measures, which take into account spatial effects, have begun to emerge in the literature. Clark (2007) applies a spatial error model and GWR for estimating local car ownership in the United Kingdom, but did not focus on comparing the results in particular. Wall (2004) compares conditional autoregressive models (CAR) and spatial simultaneous autoregressive models, but finds counter-intuitive or impractical results.

Bitter *et al.* (2007) compare the spatial expansion method and the GWR approach for a dataset from Tuscon, Arizona, applying seven housing characteristics, including two from a principal components analysis. While they found spatial variation in both models, they could not separate the observed spatial heterogeneity by localised supply and demand characteristics and the impact of omitted variables. However, GWR outperformed the spatial expansion

specification in terms of explanatory power and predictive accuracy in their study, although the differences were narrowed to some degree with the addition of the spatial lag term in the expansion specification. The authors conclude that when explanatory power and predictive accuracy are the primary objectives, GWR is the superior approach (Bitter *et al.*, 2007, 23).

Farber and Yeates (2006) compare OLS with a spatial lag model, GWR, and a moving window regression (MWR) approach. They found that the GWR approach is superior to all others; therefore it may be regarded as the one which accounts best for the spatial variation. They conclude that by allowing parameters to vary spatially, estimation accuracy of the response variable improves dramatically while spatial biases diminish to nominal amounts (Farber and Yeates, 2006, 417). However, they observed some extreme coefficients in their GWR model, which were clustered in specific neighbourhoods of their study area. They note that in the light of the unavailability of a robust statistical framework for GWR and MWR, irrational coefficients pose a major threat to the adoption of GWR by assessment authorities (Farber and Yeates, 2006, 418). The finding of extremely high coefficient variability is not unusual and often found in practice. This has raised some concern that GWR results may be misleading because of the suspicion that the variability observed is somehow built into the model by its own calibration and estimation mechanisms (Páez, 2005, 163). By comparing GWR and the expansion method in a simulation exercise, Páez concludes that, on average, spatial variability in GWR is not an artifact of the calibration procedure, and that GWR is sufficiently flexible to reproduce the type of map patterns used in the simulation experiment. Moreover, he finds that both approaches are able to provide reasonable representation of the spatial patterns inherent in the simulated data. Kestens *et al.* (2006) compare GWR and the expansion method as well, highlighting the individual merits of both methods. While GWR provides additional insight by measuring local regression statistics, the expansion term makes it possible to analyse and explain the cause of parameter heterogeneity, whether its structure is spatial or not. They conclude that both methods are complementary rather than substitutes for each other (Kestens *et al.*, 2006, 94).

Páez *et al.* (2008) compare moving windows regression (MWR), GWR and moving windows kriging. They find that the MWR approach leads to superior results relative to single market and global approaches and also to modelling spatial dependencies. Therefore, they conclude that market segmentation may be more important than spatial dependencies.

The authors of all these articles use various measures to compare and evaluate the model approaches. The most popular OLS goodness-of-fit measure is certainly the adjusted R-square. This measure is useful and applied in GWR as well, but it is no longer applicable in spatial simultaneous autoregressive models (Yu *et al.,* 2007, 1092). Therefore, other measures such

as residual sum of squares (RSS), sum of squared errors (SSE), maximum likelihood value, Log likelihood (Loglik) and Akaike's Information Criterion (AIC) are applied. The null hypothesis that the contribution of a relationship is zero is often investigated based on t-test or F-test (Gao *et al.*, 2006, 1041). Additionally, model prediction power verification is evaluated through out-of-sample testing (i.e. Bitter *et al.*, 2007).

4.4 Data and variables

Butler (1980, 97) states that, in principle, all characteristics relevant to the determination of market price should be included in a hedonic function. In practice, this cannot be done because the number of such characteristics is unmanageably large, and data on many of these are either unavailable or of poor quality. In addition, even without data constraints, a relatively small number of explanatory variables leads to considerable multicollinearity. Therefore, the aim is to find a broad set of statistically significant variables with expected signs and moderate impact of multicollinearity while the estimates should have a sufficient model fit. Butler (1980, 97) notes that for the aforementioned reasons, any estimate of the hedonic relationship must be misspecified because some of the relevant explanatory variables must be omitted. He concludes that consequently all estimates are to some extent "incorrect" and differences among them must be attributed at least in part to differences in adaptation to the specification problems common to all. However, ample and accurate variable specification is essential to infer properly and to generate statistically significant results. Finding suitable data sources in the field of real estate prices when transaction and rent data are not open to researchers presents and additional challenge.

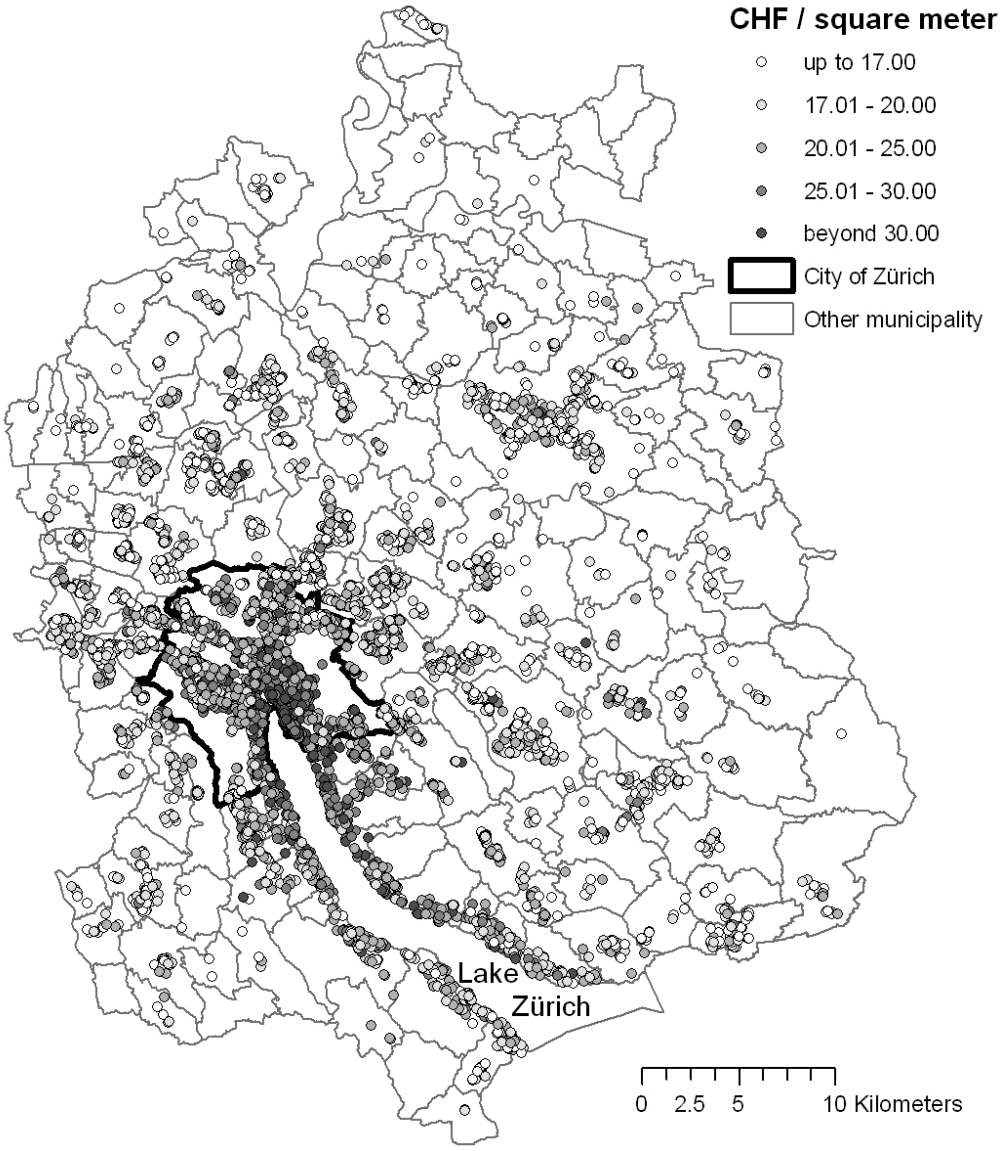
The underlying data for this study were taken from a publicly available Web site[6] between December 2004 and October 2005. The database included rent offers from various Swiss real estate online platforms. Duplicate entries, furnished and shared apartment offers as well as apartments with unusual sizes (<20 and >500 m^2) were dropped. The addresses for all dwelling units in the dataset were geocoded at building level and matched with a wide set of spatial variables. The generation of some variables included significant further work (see Löchl, 2007, for details on the calculation of solar radiation and visibility variables based on a digital elevation model), others were simply matched with available data or layers by a geographic information system (GIS). Generally, the availability of additional non-spatial and spatial data is rather good in Switzerland – for example, both the population census and the business cen-

[6] www.comparis.ch

sus are available at the hectare level. Overall, the dataset comprehended rents and additional information of 8592 dwelling units in Canton Zürich, which consists of 171 municipalities.

Because of the source and characteristics of the input price data, several restrictions apply. As the dataset includes asking rents, it does not necessarily reflect paid market prices. Moreover, the sample might be slightly biased because certain vacant dwelling units do not make their way to the Web site. However, an earlier data comparison showed that the differences in the structural variables to the complete inventory in the Federal Building and Apartment Register of the Federal Statistical Office are minor (Löchl, 2006, 6). Additionally, the structural characteristics of the properties might include faulty information, as they are self-reported by the person placing the ad. Finally, only a limited set of structural dwelling unit characteristics were available for the estimations. It included the price information and the dwelling unit size in square meters, the number of rooms (a value which counts bedrooms, living room, and kitchen), and some information about available facilities, such as balcony, fireplace, terrace/garden and the availability of a lift in the building. In cases where the age of the building was unavailable, the records were matched with the Federal Building and Apartment Register to generate the information. Some variables which might have been important contributors were unavailable and had to be left out from the equations, such as information about the last major renovation and garage availability. However, parking costs are usually not included in the net rent. Another known relevant factor is landlord identity since non-profit organisations such as cooperative building associations offer rents well below average market rents. They can be found in the whole of Canton Zürich and have a market share of almost 20 percent (Statistical Office of Canton Zürich, 2004, 15). However, the landlord identification was not possible with the data, which leads potentially to some bias. In Figure 5, the observations are displayed as the monthly asking rent per square meter.

Figure 5 Net asking rent per square meter in CHF per month in Canton Zürich



Administrative boundaries from Vector25 © 2009 swisstopo (JD082776)

The selection of explanatory variables was the result of exploratory data analysis through OLS stepwise regression in SPSS Version 16[7]. The final dataset included two response and 31 explanatory variables. The descriptive statistics are found in Table 5. Basically, all available structural variables in the source data were used except for the number of rooms as there is high multicollinearity with the dwelling unit floor area. Additionally, a broad set of spatial

[7] The GWR software does not provide functions for variable selection such as forward, backward and stepwise methods.

variables were found to be significant without considerable correlation among each other and with the structural variables. One was public transport accessibility (PTACC[8]) besides car accessibility (CARACC[9]), as the study area has a well developed public transport system. The last microcensus in 2005 (Statistical Office of Canton Zürich, 2008) revealed that public transport is used by the inhabitants for almost 30 percent of all kilometres travelled. A remarkable feature of the sample is that the mean distance to the next railway station is 910 meters on average. The whole region of Zürich is connected by a dense network of suburban railways, and all stations are served at least once per hour, with most served twice or even four or more times per hour, assuring a very high quality of service. The seven explanatory time dummy variables not only reflect rent changes over time, but also include the known process that landlords tended to reassign specific costs from the net rent to the additional costs in that period.

[8] Accessibility is measured with the log-sum term of a choice model.

[9] See footnote before.

Table 5 Descriptive statistics of the final variable selection

Variable	Description	Type[1]	Min	Max	Mean	S.D.
Dependent variables						
RENT	Monthly net asking rent in CHF	C	476.00	15000.00	1845.26	908.60
SQMRENT	Monthly net asking rent per square meter in CHF	C	10.60	53.57	20.62	5.35
Structural explanatory variables						
SQM	Floor area in square meter	C	20.00	400.00	91.58	36.92
LIFT	Building has a lift	D	0.00	1.00		
FIREPLACE	Dwelling unit has a fireplace	D	0.00	1.00		
BALCONY	Dwelling unit has one or more balconies	D	0.00	1.00		
GTERRACE	Dwelling unit has a garden terrace	D	0.00	1.00		
ISHOUSE	Single family house	D	0.00	1.00		
BUILBEF20	House built before 1921	D	0.00	1.00		
BUIL21TO30	House built between 1921 and 1930	D	0.00	1.00		
BUIL31TO80	House built between 1931 and 1980[2]	D	0.00	1.00		
BUIL81TO90	House built between 1981 and 1990	D	0.00	1.00		
BUIL91TO05	House built between 1991 and 2005	D	0.00	1.00		
Spatial explanatory variables						
CARTT_CBD	Average travel time to Zürich CBD by car in min	C	8.00	58.4	29.92	9.22
CARACC	Regional car accessibility to employment	C	7.18	10.30	9.22	0.57
PTACC	Regional public transport accessibility to employment	C	-19.46	12.39	10.69	1.60
RAILSTATION	Euclidean distance to next rail station in km	C	0.01	5.73	0.91	0.66
AUTOBAHN	Autobahn within 100 m	D	0.00	1.00		
AIRNOISE	Daily average aircraft noise above 52dB	D	0.00	1.00		
HOTREST_JOBS	Number of jobs in hotel and restaurant industry within 1 km	C	1.27	7187.78	316.33	782.43
POP_DENS	Number of inhabitants in hectare	C	0.00	2004.00	92.95	71.43
FOREIGNERS	Proportion of foreigners[3] in hectare	C	0.00	0.50	0.05	0.04
TAXLEVEL	Local income tax level for individuals	C	69.00	122.00	110.36	14.12
SLOPE	Slope by 25 m raster	C	0.00	0.26	0.36	0.03
VIEW_LAKE	Visibility[4] of lake surface (>1 km^2) in hectare	C	0.00	8887.81	439.68	1050.07
VIEW_ALL	Total visibility[4] of terrain surface in hectare	C	90.75	78421.00	8840.15	5638.71

SOLAR_EVE	Evening solar exposure index	C	0.25	607.11	238.22	80.85
Time explanatory variables						
Q1_04	Ad from 1st quarter 2004	D	0.00	1.00		
Q2_04	Ad from 2nd quarter 2004	D	0.00	1.00		
Q3_04	Ad from 3rd quarter 2004	D	0.00	1.00		
Q4_04	Ad from 4th quarter 2004	D	0.00	1.00		
Q1_05	Ad from 1st quarter 2005	D	0.00	1.00		
Q2_05	Ad from 2nd quarter 2005	D	0.00	1.00		
Q3_05	Ad from 3rd quarter 2005	D	0.00	1.00		
Q4_05	Ad from 4th quarter 2005[2]	D	0.00	1.00		

[1] C = continues; D = dummy
[2] Reference case
[3] Foreigners are defined as inhabitants with nationalities outside of North-Western Europe, North America, and Australia
[4] Visibility is calculated based on a 25m DEM, not considering buildings, trees etc.

4.5 Estimation

In general, hedonic theory does not strictly specify a functional form (Cropper *et al.*, 1988; Halvorsen and Pollakowski, 1981, 37). Sirmans *et al.* (2005, 6) explain that studies have wrestled with the problem of the correct functional form and no consensus has been found of which is most appropriate[10]. The semi-log specification has some advantages, such as the coefficients can be easily interpreted as the percentage change in the price given a one-unit change in the characteristic. Moreover, it helps to minimize the problem of heteroscedasticity and it mitigates the impact of nonlinear relationships between market price and the explanatory variables (Malpezzi, 2003). Therefore, hedonic pricing equations are typically estimated using either linear or semi-logarithmic models (Sirmans *et al.*, 2005, 4). A second challenge is the adequacy of parametric specification. Some authors (e.g. Anglin and Gençay, 1996, 633; Martins-Filho and Bin, 2005, 93f) indicate that this problem arises from the inability of economic theory to provide guidance on how characteristics of similar products relate functionally to their market price. Consequently, there have been attempts to use semi- or nonparametric methods (Clapp, 2003 and 2004; Fahrländer, 2006; Pace, 1993 and 1998; Parmeter *et al.*, 2007), which allow for the possibility of nonlinearity in the hedonic price functions and flexible modelling of the influence of continuous covariates on the dependent variable.

[10] Functional forms for the hedonic price function that have been proposed and used in the literature include the linear, the quadratic, the log-log, the semi-log, the inverse semi-log, the exponential, and the Box-Cox transformation (Freeman, 1993, 379).

In this study, after variable selection, further data modelling and diagnostics were performed in the open-source statistical software R Version 2.8[11] (R Development Core Team, 2008) in this study. Fox (2002) provides a good introduction into regression analysis with R. For a technical introduction into the R package *spdep* and available functions for spatial data analysis, see Bivand *et al.* (2008). Kriström (2008) provides an example of the estimation of a hedonic model in the housing market with the statistical software R. Both the dependent and the explanatory variables were transformed as suggested by Mosteller and Tukey (1977, 109), whereas a logarithmic transformation was taken for an amount or count and a logit transformation was used for fractions[12]. Due to memory capacity issues of the GWR methods in R (the computational overhead of this method is considerable, especially for a large dataset), the GWR 3.0 software by Martin Charlton and Stewart Fotheringham was applied for the GWR approach. A documentation of the software is available in Fotheringham *et al.* (2002).[13]

Two OLS models (Models 1 and 2) were selected for the analysis including both structural and spatial explanatory variables, taking the logarithmic transformation of RENT as the dependent variable (see Table 6). Similar models, considering rent per square meter as the dependent variable, did not show any significant difference of the coefficients, besides the one referring to the floor area variable, but had a much lower model fit and are therefore not presented here. The OLS results of two additional models (Models 3 and 4), which are only considering spatial explanatory variables, are found in the appendix. While models with solely spatial explanatory variables can be used in UrbanSim currently, it is hoped in the future the simulation system becomes sensitive to structural variables of the building stock, since this presents the opportunity to better reflect the local situation and to make the simulation more realistic.

The explanatory variable selection is essentially the same in Models 1 and 2 with one exception: Model 1 uses car travel time to Zürich Central Business District (CARTT_CBD) while Model 2 considers regional car accessibility to employment (CARACC). Both explanatory

[11] www.r-project.org

[12] Tests with Box-Cox transformations did not succeed since there are some independent dummy variables where more than two thirds of the sample is coded zero, making it impossible to automatically compute a sufficiently large constant in R (Fox, 2002, 107). As it is, such specifications are not readily implemented in the presence of spatial dependence (Kim *et al.*, 2003, 31).

[13] The fact that some variables do not show local variability (e.g. AIRNOISE) actually requires a mixed GWR approach, in which the referring coefficients are gobal and the others are local as the basic GWR model (Fotheringham *et al.*, 2002, 65ff). However, this procedure is not implemented in the GWR 3.0 software; it is planned to be included in the upcoming 4.0 release.

variables could not be used in one model due to multicollinearity issues. Model 1 using CARTT_CBD reveals the best overall model fit whereas Model 2 including CARACC has more policy relevance, as it is sensitive to any improvement in the street network beyond those affecting the travel time to Zürich CBD.

Outliers were identified as suggested by Fotheringham *et al.* (2002, 78) and Chatfield (1995, 265) where a data point with the absolute value of the studentised residual exceeding 3 is a potential outlier. This threshold has been applied for both models. Efforts of modelling interaction of the selected explanatory variables, particularly those with slope and the total view variable, did not improve the models. SLOPE might indeed have an intrinsic value, as it usually comes with a certain potential view (at least of immediate surroundings), while VIEW_ALL and VIEW_LAKE include the total visible area in hectares, which is rather coarse and considers, in particular, the long-range view.

The adjusted R-square values indicate good model fits, while Model 1 performs slightly better than Model 2. Researchers often apply out-of-sample testing for the purpose of model accuracy and model prediction power verification, where a proportion of the available data records are left out from the estimations and used to be compared to estimated outcomes. However, in this study, no such analysis was performed, as data fitting was the priority, and there was less focus on the predictive power of the models.

Table 6 Estimated OLS parameters (N = 8592)

Y = Ln(RENT)	Model 1				Model 2			
Explanatory variable	Estimate	SE	sig.	VIF	Estimate	SE	sig.	VIF
(Constant)	5.214				3.638			
Ln(SQM)	0.777	0.005	***	1.392	0.776	0.005	***	1.394
LIFT	0.026	0.004	***	1.160	0.024	0.005	***	1.159
FIREPLACE	0.108	0.010	***	1.167	0.118	0.011	***	1.167
BALCONY	0.033	0.004	***	1.186	0.033	0.004	***	1.186
GTERRACE	0.105	0.015	***	1.013	0.107	0.016	***	1.013
ISHOUSE	0.133	0.013	***	1.153	0.125	0.013	***	1.153
BUILTBEF21	0.077	0.006	***	1.230	0.109	0.006	***	1.209
BUILT21TO30	0.086	0.011	***	1.062	0.094	0.011	***	1.061
BUILT81TO90	0.018	0.005	***	1.199	0.018	0.005	***	1.199
BUILT91TO05	0.067	0.005	***	1.305	0.067	0.005	***	1.306
Ln(CARTT_CBD)	-0.294	0.007	***	2.089				
CARACC					0.119	0.005	***	2.471
PTACC	0.006	0.001	***	1.472	0.011	0.001	***	1.493
Ln(RAILSTATION)	-0.012	0.002	***	1.184	-0.012	0.003	***	1.232
AUTOBAHN	-0.060	0.013	***	1.033	-0.067	0.013	***	1.033
AIRNOISE	-0.039	0.006	***	1.281	-0.096	0.006	***	1.276
Ln(HOTREST_JOBS)	0.020	0.002	***	2.339	0.032	0.002	***	2.447
Ln(POP_DENS)	-0.028	0.002	***	1.242	-0.026	0.002	***	1.243
FOREIGNERS (logit)	-0.018	0.002	***	1.340	-0.023	0.002	***	1.338
Ln(TAXLEVEL)	-0.130	0.014	***	1.326	-0.223	0.015	***	1.292
SLOPE (logit)	0.016	0.002	***	1.285	0.026	0.002	***	1.255
Ln(VIEW_LAKE)	0.008	0.001	***	1.759	0.012	0.001	***	1.944
Ln(VIEW_ALL)	0.005	0.002	**	1.450	-0.001	0.002		1.528
Ln(SOLAR_EVE)	0.037	0.003	***	1.107	0.018	0.003	***	1.089
1Q_04	0.080	0.013	***	1.195	0.079	0.014	***	1.195
2Q_04	0.038	0.012	***	1.228	0.032	0.013	**	1.228
3Q_04	0.059	0.009	***	1.618	0.057	0.009	***	1.618
4Q_05	0.032	0.007	***	2.612	0.033	0.007	***	2.613
1Q_05	0.045	0.007	***	2.616	0.044	0.007	***	2.616
2Q_05	0.026	0.007	***	2.035	0.024	0.008	***	2.035
3Q_05	0.037	0.007	***	2.526	0.034	0.007	***	2.526
Adjusted R-square	0.854				0.836			
F-test	1677.000		***		1460.000		***	

Probability of rejecting H_0 = *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

When detecting and incorporating spatial effects, it is necessary to produce a weight matrix based on some kind of contiguity. Several approaches are available (e.g. Anselin, 2002, 258; Bivand *et al.*, 2008, 251ff). Because of a relatively high heterogeneity of spatial distribution of the data points (see Figure 5), a k-nearest neighbours approach by Euclidean distance was chosen, whereas k=9 produced the best results in terms of model fit measured by log likelihood (LogLik) and Akaike Information Criterion (AIC) for the spatial simultaneous autoregressive models. Therefore, each observation has the nine next observations in terms of linear distance defined as contiguity in the weight matrix.

A set of diagnostic tests for spatial autocorrelation were performed based on the contiguity weight matrix for both models, which clearly indicated the need to consider spatial autocorrelation in the models. This is unsurprising, given that most variables in the estimation actually have a spatial relation. The global Moran's I measure clearly indicated spatial autocorrelation in both cases, while the Lagrange Multiplier tests (Anselin *et al.*, 1996) pointed to both spatial lags and spatial errors in the models, as can be seen by the high significance level even in their robust versions, although the test values are much higher for the robust Lagrange Multiplier test of the spatial error (Robust LMerr).

Table 7 OLS diagnostics ($N = 8592$)

	Model 1		Model 2	
Measure	Value	sig.	Value	sig.
LogLik	3685.653		3183.261	
AIC	-7307.306		-6302.522	
Global Moran's I	0.213	***	0.287	***
LMerr	2164.867	***	3918.393	***
LMlag	454.444	***	977.235	***
Robust LMerr	1722.903	***	2981.065	***
Robust LMlag	12.480	***	39.907	***

Probability of rejecting H_0 = *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Consequently, the SARerr and the SARmix approach were considered besides the GWR approach in the following analysis, all applying the variable selection from Model 2 due to its higher policy relevance as described above. Visually inspecting of the resulting residual scatter plots of the model indicated a problem with heteroscedasticity, as very low and very high observed asking rents tended to be overestimated. The problem may have been caused by

missing and unavailable crucial structural variables, such as information about building conditions or about the landlord. However, the standard error of the OLS model was compared to the results based on the heteroscedasticity consistent covariance matrix and did not show any significant difference (for the procedure in R, see Bivand *et al.*, 2008, 290). Therefore, heteroscedasticity was not considered to be a serious problem in this case.

Since the data points were not randomly located in particular and sometimes spaced widely about (see Figure 5), an adaptive kernel was used for the GWR approach. As there was no prior justification for supplying a particular bandwidth, the two available options for finding the most appropriate bandwidth were tested: 1) minimisation of the Cross-validation (CV) score and 2) minimisation of the AIC measure (Fotheringham *et al.*, 2002, 212). Both methods suggested the same bandwidth. In Table 8, the results of the OLS, SARerr, SARmix and GWR estimations, based on the variable selection in Model 2, are shown. For the GWR model, the mean and the standard deviations of the parameter values are presented. The parameter estimates do not vary dramatically among the models (exceptions are CARACC, AIRNOISE, and VIEW_ALL, for which the GWR model indicates a large spatial variation of those variables indicated by multiple standard deviation values compared to the mean of the estimates). For model evaluation purposes, both the sum of squares of the errors (SSE) and the AIC have been calculated. In both measures, smaller values indicate better model fit. In this case, both measures identify the best model fit for the GWR estimates; second best the SARmix model. However, four explanatory variables (PTACC, AIRNOISE, TAXLEVEL and SOLAR_EVE) are not significant in the SARmix (as indicated by the bold values in Table 8), whereas AIRNOISE even has a counterintuitive sign. Particularly with regard to the envisaged application of the hedonic model in UrbanSim, the insignificance of the public transport accessibility coefficient creates a major limitation of the model.

Table 8 Comparison of coefficients from various modelling approaches

Y = ln(RENT)	OLS	SARerr	SARmix	GWR	
Explanatory variable	Estimate	Estimate	Estimate	Estimate mean	Std.dev.[1]
(Constant)	3.638	3.611	1.329	3.954	
Ln(SQM)	0.776	0.773	0.773	0.772	0.055
LIFT	0.024	0.026	0.025	0.022	0.017
FIREPLACE	0.118	0.085	0.088	0.096	0.038
BALCONY	0.033	0.030	0.030	0.031	0.011
GTERRACE	0.107	0.094	0.097	0.098	0.073
ISHOUSE	0.125	0.137	0.141	0.138	0.047
BUILTBEF21	0.109	0.082	0.075	0.091	0.026
BUILT21TO30	0.094	0.075	0.070	0.084	0.042
BUILT81TO90	0.018	0.019	0.020	0.026	0.024
BUILT91TO05	0.067	0.080	0.085	0.078	0.023
CARACC	0.119	0.113	0.068	0.011	0.098
PTACC	0.011	0.006	**0.002**	0.039	0.034
Ln(RAILSTATION)	-0.012	-0.013	-0.016	-0.009	0.016
AUTOBAHN	-0.067	-0.076	-0.079	-0.071	0.066
AIRNOISE	-0.096	-0.083	**0.032**	-0.014	0.032
Ln(HOTREST_JOBS)	0.032	0.029	0.014	0.025	0.017
Ln(POP_DENS)	-0.026	-0.028	-0.028	-0.029	0.014
FOREIGNERS (logit)	-0.023	-0.014	-0.013	-0.018	0.007
Ln(TAXLEVEL)	-0.223	-0.201	**-0.002**	-0.160	0.268
SLOPE (logit)	0.026	0.016	0.007	0.010	0.009
Ln(VIEW_LAKE)	0.012	0.010	0.003	0.008	0.007
Ln(VIEW_ALL)	**-0.001**	**0.008**	0.017	0.006	0.024
Ln(SOLAR_EVE)	0.018	0.012	**0.006**	0.024	0.027
1Q_04	0.079	0.078	0.079	0.072	0.034
2Q_04	0.032	0.046	0.044	0.040	0.036
3Q_04	0.057	0.067	0.066	0.060	0.021
4Q_05	0.033	0.038	0.036	0.032	0.020
1Q_05	0.044	0.050	0.049	0.048	0.020
2Q_05	0.024	0.027	0.026	0.024	0.018
3Q_05	0.034	0.040	0.039	0.037	0.022
Lambda		0.637			
Rho			0.599		
SSE	239.781	182.655	179.700	174.202	
AIC	-6302.522	-8169.98	-8319.190	-8392.987	

Bold: not significant at 0.05 level

[1] Standard deviation of coefficient values across study area

In order to further assess and compare the predictive accuracy of spatial simultaneous autoregressive approaches and the GWR model, the predicted rents were compared with the observed values. The results of this analysis (presented in Table 9), show that the OLS approach has the lowest accuracy. Less than 66 percent of the predictions were within the range of +/- 2 percent of the actual rent price and it showed the lowest accuracy within +/- 5 percent and +/- 10 percent as well. The other three models perform considerably better, whereas around 73 percent are in the +/- 2 percent range and slightly above 98 percent are in the +/- 5 percent range. The GWR approach has the best predictive accuracy in the two tighter ranges but its performance was slightly inferior in the +/- 8 percent range, although the differences were only marginal among the models. The SARerr model is more accurate in the +/- 2 percent range than the SARmix model, but slightly lower in the +/- 5 percent and +/- 8 percent range. Finally, the residuals of all four models were tested for autocorrelation, measured by the Moran's I test. Besides the OLS, the GWR approach seems not to solve the autocorrelation, while there is essentially no autocorrelation observable anymore in the residuals of the SARerr and the SARmix model.

Table 9 Predictive accuracy of models: Percentage of predicted rents within specified range of actual asking rents (Ln(RENT)) and Moran's I of the residuals

	2% range	5% range	8% range	Moran's I
OLS	65.77 %	96.81 %	99.91 %	0.287
SARerr	72.65 %	98.02 %	99.93 %	-0.018
SARmix	72.59 %	98.14 %	99.94 %	-0.016
GWR	73.79 %	98.46 %	99.92 %	0.119

The fact that the GWR approach does not solve the autocorrelation in the residuals is alarming. Wheeler and Tiefelsdorf (2005, 186) emphasise that spatially auto-correlated residuals can produce severely correlated local regression coefficients. This is proven by a further investigation of the correlation matrix of the regression coefficients. In the correlation matrix, significant values beyond +/- 0.5 can be found in almost 8 percent of the cases and values beyond +/- 0.6 in 4 percent. Therefore, the issue of multicollinearity among the coefficients clearly exists in the GWR results. This raised strong doubts, if the GWR method is the most suitable and reliable approach here despite its good model fit. The SARmix model has some limitations as well, namely the fact that four explanatory variables, including public transport accessibility, are insignificant. Given that the SARerr model accuracy is almost as good as the SARmix model, it is preferred here and suggested for the next update of the UrbanSim application for Zürich.

4.6 Conclusions

Setting up an UrbanSim application for a metropolitan area is a major task and includes, in particular, a significant data collection effort. This chapter focused on hedonic residential rent modelling in order to establish a housing price surface for the UrbanSim application in Zürich. The task was particularly difficult since there was no tax assessor data or data from commercial sources available for the study area at the required spatial resolution. Eventually, publicly available residential asking rents from a web-based portal were used. The variable selection was found by considering significant explanatory variables while strictly controlling for multicollinearity. Spatial autocorrelation later proved to be a problem and several spatial simultaneous autoregressive models and the GWR approach were tested. Although, all goodness-of-fit measures indicated slightly better performance for the GWR model compared to the spatial simultaneous autoregressive models, its residuals were still auto-correlated. Additionally, the resulting GWR coefficients were correlated, which severely reduced confidence in the appropriateness of the method in this study. The SARerr model has been preferred against the SARmix model, because the latter had several insignificant variables, including public transport accessibility, which is a crucial measure in land use and transport modelling. Additionally, the SARerr model showed good accuracy of the predicted values compared to observed values.

Overall, the analysis highlighted the complex spatial structure of housing markets. The need to explicitly address spatial effects is obvious since a failure to do so may result in loss of explanatory power and erroneous estimates. The GWR method may not always be the best choice, although an additional benefit of GWR is to provide a means to visualize the spatial structure of housing markets which was also emphasised in earlier studies. In this study, sufficient locational explanatory variables were available but other studies have highlighted that the GWR method is also helpful in situations where locational information is difficult to obtain or when knowledge of local submarkets is unavailable (Bitter et al., 2007, 24). However, locally correlated GWR coefficient estimates are a remaining problem which was present in this study as well. Methodological improvements of the GWR approach have been suggested in the literature, but are subject to an ongoing debate among econometricians. Spatial simultaneous autoregressive approaches proved to be a reasonable alternative in the analysis, which can be implemented in UrbanSim more easily because of its structure of a single set of resulting parameters.

Technically, UrbanSim is somewhat flexible concerning variable selection for the various needed models, as long as the variables are available and constantly updated. Therefore, the

results of this article open discussion about the further development of the model system, in particular with regard to real estate price data used in the hedonic modelling. It is of particular concern since real estate prices are also a major determinant of location choice. The hedonic modelling efforts for the Zürich application of UrbanSim are based on residential rents due to the predominant role of renting in Switzerland. Moreover, UrbanSim may not become sensitive for some explanatory variables which have been used in the models for this article. But it can be stated that besides using spatial variables, more information about the building stock such as size, type and age (presuming data availability) would greatly improve the hedonic estimations in addition to advancing the household location choice models, perhaps even the developer model. Moreover, local rent and buying prices are certainly not perfectly correlated. Therefore, combining rent models with transaction price models would additionally incorporate the rent/purchasing decision of households. With regard to model methodology, it again became obvious that hedonic housing models considering spatial effects are more reliable and are therefore suggested for further exploration in future applications of UrbanSim.

5 Implementing the integrated land use model UrbanSim for the Greater Zürich area – A field report

5.1 Introduction

The transformation of spatial structures in Switzerland has an increasing pace in the last decades. The suburbanisation and urban sprawl processes have not come to an end yet (Schulz and Dosch, 2005). New and additional transport infrastructures have certainly made their contribution to this development but their long-term spatial impacts are only partly understood so far (Bundesamt für Raumentwicklung, 2003; Tschopp and Axhausen, 2007).

Consequently, the complex interactions among spatial development and transport are insufficiently considered in planning decisions. One reason is the lack of applicable instruments, with which interdependencies could be appropriately represented and planning consequences forecasted or alternatives compared.

Integrated land use transport models might close this gap. The addition "integrated" reflects the interdependency among land use and transport in the model system. Short- and long-term feedbacks among the arrangement and the environment of the transport network, i.e. the accessibility, and land development through the kind of location choices made by households and businesses, are incorporated into such a model.

While the use of pure transport models is common in many planning authorities, the land use part is currently neglected. The increasing urbanisation world-wide, but also the rising interest in sustainability of spatial development as well as new theoretical findings and technical improvements of modelling techniques has encouraged academics and consultants to develop a new generation of urban simulation models. Without doubt, interest for integrated land use and transport models has also increased at local and regional planning authorities. Nevertheless, there are very few evaluation reports in the European realm about the implementation efforts needed and the integration of this new generation of models into planning practice. This article describes the experiences, made during an implementation of UrbanSim for the Greater Zürich area. The system is running, but has yet to be fully calibrated.

5.2 State of the art

Some model systems have evolved during the last 15 years, which were successfully applied in practice more than once. They consider the most important processes, which affect spatial development, such as land development, location decisions of households and businesses as well as transport. In the following discussion, those land use models are mentioned, which have found most application in practice. Beyond, there are many more model systems, which are either approaches in the academic realm or have found less application (e.g., ALBA-TROSS/PUMA, Arentze und Timmermans, 2004; Ettema und Timmermans, 2006; ILU-MASS, Moeckel *et al.*, 2006; ILUTE, Salvini und Miller, 2005; MUSSA, Martínez, 2000). For a more exhaustive review, additional information about and differences among the simulation models currently in use, see Hunt *et al.*, 2005 and Wegener 2004.

ITLUP (also known as DRAM/EMPAL – Putman, 1994) is currently the most widely used model in the US, while MEPLAN (Hunt, 1994) and TRANUS (Modelistica, 2006) are primarily applied in Europe and South-America. DELTA (Simmonds and Feldman, 2005) was used in several regions of Great Britain. All four are proprietary software packages, which have been used for years. In contrast, UrbanSim (Waddell *et al.*, 2003) is an open-source software, whose explicit modelling of land and real estate markets makes it exceptional. Moreover, UrbanSim does not include the presupposition of equilibrium demand and supply with regard to building space, but does allow dynamic disequilibrium.

There is an ongoing discussion in the international literature about the theoretic foundations of the models and their current and future capabilities in planning practice (Beckmann, 2006; Miller, 2006; Timmermans, 2006). One focus is the question of the required and sufficient complexity of a simulation system – that is, the probability of faulty prediction increases in case the model is too coarse and simple and does not consider crucial factors sufficiently. Important feedbacks are not considered and critical secondary and tertiary effects of measures and processes are ignored (Miller, 2006). Conversely, an excess of complexity can require undeliverable data requirements as well as time and cost efforts, so that the practical application is almost impossible. This was the experience with the first generation of land use models (Lee, 1973). Overall, the issue of necessary complexity cannot be answered universally without considering the needed depth and functionality coverage of the envisaged model.

5.3 Project "Infrastructure, Accessibility and Spatial Development"

One part of the so-called poly project "Future of Urbanised Landscapes"[14] was the project "Infrastructure, Accessibility and Spatial Development". The purpose was to implement an integrated land use model for the Greater Zürich area. The simulation software UrbanSim was chosen for various reasons. Functionality and disaggregation of the simulation software as well as its implementation can be considered advanced. Moreover, there are various applications worldwide and an extended documentation, which is available online (CUSPA, 2006). As mentioned, UrbanSim is not a proprietary software package and therefore available for free.

UrbanSim was adjusted to the circumstances of the Greater Zürich area and coupled with the transport model of Canton Zürich. Within the project, all necessary data were collected, generated and collated for application in the Greater Zürich area, including seamless data of inhabitants, jobs and building stock at the hectare level as well as parameter estimates for the discrete choice models. Based on this work, various analysis and impact assessments of planning alternatives can be conducted, i.e. the effects of new, enlarged or reinstated streets on land use development. Test scenarios of how the enforced or restricted declaration of developed land or the effect of increased density on the location choices of households and businesses are possible.

5.4 UrbanSim

UrbanSim is a software package for the integrated simulation of land use and transport. It was developed at the University of Washington in Seattle/US (i.e. Waddell, 2002). The program is published under the GNU General Public Licence (Free Software Foundation, 1991). Therefore, the software code is available for free and can be adjusted by any user. Since summer 2006, UrbanSim is available in a new version with improved program design, simplified user and programming interfaces and implemented in the script language Python. Additionally, UrbanSim was put into the larger context of the newly created framework OPUS (Open Platform for Urban Simulation) (CUSPA, 2006; Waddell *et al.*, 2005). OPUS follows a modular

[14] The interdisciplinary research project was funded by a competitive ETH research grants programme from 2004 to 2007. It investigated the characteristics, forms, functions and development of urban landscapes in five specialised areas: construction planning and operation, the regional balance of raw materials, urban development, landscape architecture and transport and spatial development.

approach with the aim to encourage the testing and linking of various model approaches for land use simulations. The development of additional or improved single modules is supported by a set of basic functions. Standardised interfaces can simplify the collaboration of different development teams in the field of land use simulations across the globe dramatically. Python (Python Software Foundation, 2006) and the database management system MySQL (MySQL AB, 2006), both required software to run UrbanSim, is available for free. Preferably easy access ought to encourage the use of UrbanSim. At the same time, improvements and extensions through third parties is fostered. The aim is to establish a worldwide community of users and developers, who are advancing the level of maturity of the software as well as its usability.

UrbanSim is currently applied by a range of user groups. Applications are found predominantly in the US, for example in Amherst, El Paso, Eugene/Springfield, Honolulu, Houston, Salt Lake City, San Antonio and Seattle. The reason is not only proximity to the developers. Another reason is a requirement in the US to include a systematic consideration of environmental impacts of transport planning. UrbanSim understandably assumes the US data acquisition and availability conditions. For US data, which are not available in the necessary format, UrbanSim offers specialised tools. However, only users in the US can benefit optimally from it and the tools are not as useful outside the country. Nevertheless, additional applications of UrbanSim are emerging worldwide, besides the one reported in this chapter – for example, for Lyon, Melbourne, Paris, Randstad, Rome and Tel Aviv. Mostly, these applications are implemented in research-oriented institutional frameworks, but there are examples where a local planning authority has taken the lead, for example in Houston and Minneapolis/St. Paul.

Land use and transport are modelled as a result of a set of market controlled decisions of various interacting stakeholders, i.e. households. Additionally, zoning allowances can be considered (Waddell, 1998). Since the behaviour of single actors is the subject of the simulation, UrbanSim can be considered as a microsimulation, although single persons are aggregated to households in the standard version of the model (Waddell *et al.*, 2003). The application of UrbanSim in the Puget Sound Region of the State of Washington is already using partly person-based information.

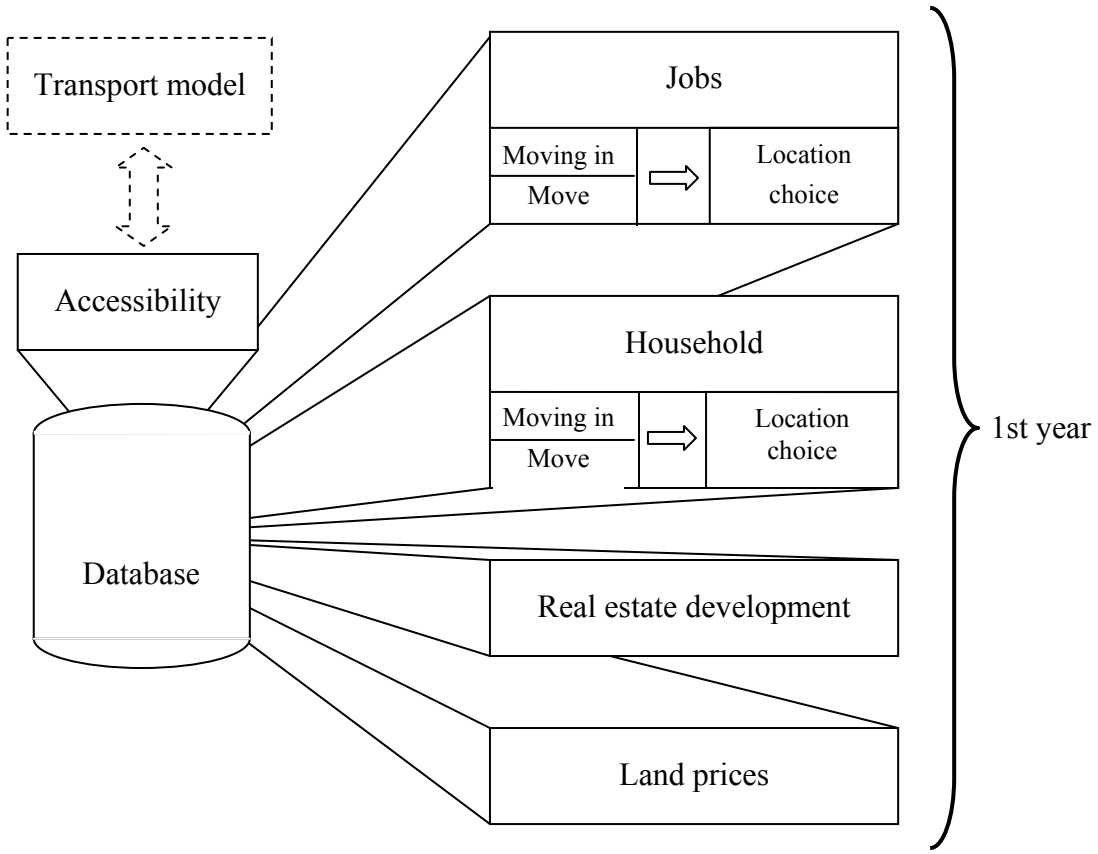
An initial database of land use information has to be provided before the computer-based simulation of the future development of land use and transport can begin. Another prerequisite is the availability of an external transport model, since UrbanSim only provides models of land use processes. The transport model provides information of transport supply among the transport zones (foremost average travel times). UrbanSim uses such information as input data

to calculate accessibility measures. Conversely, the calculated distribution of land uses by UrbanSim can be used as input for a transport model.

5.4.1 Basic design

The basic design of the simulation framework is simple, as Figure 6 shows. Land use and transport supply for the base year, socio-demographic key data and development scenarios for future years as well as model parameters are centrally stored in the database. All models necessary for the simulation sequentially access this database for reading the required variables and for calculations. The model updates the database with its results and the updated database is the base for the calculation of the next models. This procedure is repeated with one run of all models per simulated year.

Figure 6 Basic design of the simulation framework UrbanSim



The allocation of various objects (building developments, households and jobs) in the study area plays an important role. At first, additional objects are generated in the study area for the simulated year. The amount of new buildings depends on the calculated demand of households and jobs as well as the exogenously provided vacancy rates and the building space requirement per job from the previously simulated year. The capacity of new development projects is oriented at real projects. The total number of households and jobs per year in the simulation area must be provided externally. The number of additional jobs is calculated by considering the difference of the externally provided number and the current data in the database for the study area. In a second step, moving households and jobs within the simulation area is selected. They are choosen randomly based on moving probabilities.

All selected objects from both steps are located a new position in the study area. For this task, discrete choice models (DCM) are applied. Additional information about the residential location choice of households might be found below in this chapter. All attributes of all data tables of the current and previous simulation periods are available as model variables, including all calculations based on these. For example, the updated residential rents from the hedonic rent model are an input to the residential location choice model. The number of households within a certain radius of a location is used as a predictive variable for the job location choice.

The sequential use of isolated models is certainly a simplification of reality, in which various agent decisions and processes are executed simultaneously and are interacting. This reduction of complexity allows a systematic analysis and calibration of the subsystems of spatial development. Moreover, it contributes considerably to the clarity and manageability of the simulation system.

The accessibility model holds an exceptional position: It updates the accessibility values by traffic zone in the study area by applying an external transport simulation. This external model obtains the current land use data of the database of UrbanSim as an input for adjusting all transport modelling steps up to the assignment. This update is typically done once per three or five simulation years.

5.5 Data

The database occupies a central position in the implementation of UrbanSim. To implement a model of land use development and transport requires collecting data about the recent past. Rules for the future can be inferred *ceteris paribus*. The use of data about past development it

twofold: some aspects are the basis for model estimations as explanatory variables and data from the past, at several points in time, is also used for validating the estimated models.

The database of land uses is ideally assembled from available data sources. The data requirements arise predominantly from the selection of sub-systems, which are to be represented by the simulation system, e.g. households and jobs. Additionally, minimal requirements of the simulation need to be considered.

5.5.1 Data acquisition

Since the land use model for the Greater Zürich area should deliver results at the hectare level, the hectare is the selected spatial unit of the database. The federal population and business census data of the Federal Statistical Office have been stored consistently at this spatial resolution since the 1990s. Although the amount of collected data about land use has multiplied, the collation of the data needed for the simulation of the Greater Zürich area proved to be difficult for various reasons.

5.5.2 Data availability

On the one hand, if a required data set exists, the data are not automatically available to the modeller: person-based data are subject data privacy protection, other information, e.g. real estate and rent prices respectively, are the basis of business for banks and real estate consultants and are consequently not available to third parties offhand. Therefore, the issue of data availability becomes a question of costs and alliances.

On the other hand, the existence of all wished for data sets cannot be assumed. Often, one needs to manage with estimates or data from other geographical areas. For example, detailed information about households was imputed based on optimisation techniques (Bürgle *et al.*, 2005). The moving probability of businesses of different sectors was taken from data by the cantons of St. Gallen as well as Appenzell Ausser Rhoden and Inner Rhoden (Bodenmann, 2006) for Canton of Zürich. Another approach for completing missing data is to conduct a survey. A household survey was conducted within the project concerning the residential living conditions in the Greater Zürich area (Waldner *et al.*, 2005). However, conducting a survey is probably not feasible while setting up a land use model in practice because of the time required.

5.5.3 Data consistency

Disregarding the nationwide available standard data sets such as the population and employment censuses in Switzerland, efforts and costs depend on the alignment of the study area with existing jurisdictions because the availability and responsibility differ in various cantons considerably. This issue can interfere with the delineation of functional relationships and should consequently be accounted for.

The experience shows that the available data sets can have inconsistencies and gaps, which have to be corrected. The consolidation of different data sets, which have been collected by different institutions with diverging methods and purposes, can lead to additional efforts. This became for example obvious during the reconciliation of the employment census of the Federal Statistical Office with data concerning building capacity and square footage received from the fire insurance of the Canton of Zürich. Finally, the available data sets often refer to different points in time. This needs to be adjusted by interpolation.

5.6 Models

The database is the raw material of the model system, while the various models shape its behavior. Here, the term “model” stands for a mathematical formula describing interrelations in the real world. Methods of model estimation deliver mathematically exact descriptions for processes concerned. But the explanatory power of a model depends heavily on the completeness and quality of the input data as well as on the formulation and parameterisation of the the model itself. The knowledge of the model developer about functional relationships and the characteristics of the available data can affect the final estimates considerably. Moreover, only quantifiable relations can be included in the estimations. Qualitative data needs to be quantified first before it can be used in the model estimations.

As examples for two different model families, a hedonic residential rent regression model and a household location choice model are presented, which have been specified for the UrbanSim application in the Greater Zürich area.

5.6.1 Database

The data base of land use information was expanded by a household survey of 9000 households in 25 municipalities of the Greater Zürich area and districts of the City of Zürich

(Waldner *et al.*, 2005). The questionnaire included questions concerning the residential situation and socio-economic attributes of the surveyed households. A data set with around 20,000 dwelling units, which were offered at an online real estate database between 2004 and 2005, was a large but less detailed addition. The data from the survey and the online data base were geocoded and enriched with spatial data. This included information about the environment such as slope and exposition, distance to lakes and transport infrastructure or regional accessibility measures to inhabitants and employment. With the location information, neighbourhood variables like employment and population density in the surrounding area as well as variables at the municipality level such as share of the population with a university degree can be established and considered in the estimations.

5.6.2 Modelling of rents

Using the data, comprehensive hedonic models of residential rents with multiple regression analysis have been estimated (Löchl, 2006)[15]. The hedonic method is based on the assumption that the price of a good can be calculated by the sum of the value of each attribute. Its theoretic foundation was developed by Lancaster (1966) and Rosen (1974), who assumed that a flat, for example, is composed of a bundle of characteristics and there is a market for each of them. Possible value-creating attributes of flats are structural characteristics such as size, age, equipment as well as locational characteristics. The value of each attribute is implicitly included in the price of a property and can be extracted by a regression model. If the price contributions of every attribute have been estimated, it can be applied to any other property as long as the attribute values are known. Since there were only the locational characteristics available for the land use simulation for the Greater Zürich area, only those could be used for the residential rent modelling. Therefore, characteristics such as size, quality and equipment had to be disregarded and are unavailable for impact studies.

The first two data columns in Table 10 show the hedonic residential rent model with the average monthly net rent per square meter in Swiss Francs as the dependent variable. The parameters are presented standardised and non-standardised. The relationship between specification of the variable and the rent in Swiss Francs can be seen from the parameter values of the non-standardised model: A value of 1.0 means that a one unit change of the variable increases the rent by one Swiss Franc. The last two columns refer to the residential location choice model.

[15] The model is different to the model presented in Chapter 4.5, as it was estimated before the other model and some independent variables were not calculated or available yet.

In all estimated residential rent models, the car travel time to Zürich CBD (Bürkliplatz) proved to be the most important variable with regard to rent. This reveals the importance of the City of Zürich for the whole region. Other accessibility measures, such as those for public transport, could not be considered because of interdependency (multicollinearity). While the proximity to autobahn exits, autobahns and rail lines negatively affect rent, this is in a contrast to the proximity to the next rail station. A high density of jobs in the restaurant and hotel industry as an indicator for urbanity has a positive impact. Additional, proximity to a large lake, good solar exposition as well as locations with a decent slope are appreciated. Finally, there are some variables relevant at the municipality level, such as the tendency for higher rents in wealthy municipalities and in those with a comparably old building fabric.

Table 10 Rent price and residential location choice model for Canton Zürich

	Hedonic rent price model[16]		Location choice model	
Variable	Standardised	Non-standardised	General	UrbanSim
Constant		27.327***		
Accessibility/ transport				
Ln (Car travel time to Zürich CBD)	-0.349***	-5.580***	0.018***	-3.335***
Ln (Accessibility of population by public transport) times dummy „no car“			0.570***	0.600***
Distance to work place (km)			-5.459***	
Exponent of distance to work place			0.167***	
Ln (Distance to next autobahn exit in km)	0.080***	0.581***		0.119**
Ln (Distance to next rail station in km)	-0.033***	-0.242***		-0.115***
Regularly used rail tracks within 50m (dummy)	-0.027***	-0.878***		-0.933***
Autobahn within 100m (dummy)	-0.017**	-0.702**		-0.400*
Noise nuisance (dummy)			-0.236***	
Socio-economic variables				
Population density times dummy „young household“				0.006***
Household of same size within 1km radius			0.0004***	0.0001**
Number of jobs in hotel and restaurants industry within 1km radius (divided by 1000)	0.193***	1.289***		

[16] See footnote before.

	Hedonic rent price model[17]		Location choice model	
Variable	Standardised	Non-standardised	General	UrbanSim
Environmental variables				
Ln (Distance to next large lake in km)	-0.101***	-0.447***		
Sun exposure index	0.090***	0.081***		
Slope (%)	0.064***	0.111***		
Municipality related variables				
Federal tax income per capita in municipality (divided by 1000) (CHF)	0.169***	0.977***	-0.026***	1.037***
Percentage of buildings built before 1971 in municipality (%)	0.146***	0.049***		0.041***
Rent vacancy rate in municipality (%)			-0.224***	-0.110***
Percentage of inhabitant with college degree in municipality (%)				-3.073***
Household related variables				
Ratio of rent to household income			-0.546**	
Rent per sqm (CHF)				-0.600***
Floor area (sqm)/ square root of number of household members			-0.289***	
*** = significant at 1% level ** = significant at 5% level * = significant at 10% level No values: Variable is not included in model	n = 9199; adjusted R^2 = 0.454; F = 695.883		n = 877 rho^2=0.26	n = 1356 rho^2=0.08

[17] See footnote before.

5.6.3 Modelling of residential location choice

While the residential rent model estimates a monetary value of a location while considering various factors, the task is different for the location choice of households. Out of a finite number of different locations, one should be chosen which the house hunting household will most probably choose. Such modelling problems can be solved by discrete choice modelling. Mostly, those models represent decisions among few possible alternatives, for example the choice between public transport and private transport.

However, as shown by McFadden (1978), the problem formulation as a multinomial logit function is possible and reasonable also in choices with a large set of options, such as residential location choice. A random sample of 20 to 50 alternatives from the real estate database as non-chosen alternatives was assembled. One decision maker per surveyed household was selected and the currently occupied dwelling unit by the household taken as the chosen alternative. Only those households moving into their current dwelling unit less than six years ago were considered in order to ensure comparability of the chosen and non-chosen dwelling units. Utility functions were estimated, including structural characteristics of the dwelling unit (e.g., monthly rent), locational characteristics (e.g., linear distance to next autobahn) and characteristics of the decision maker (e.g., work location). Mostly, linear formulations were chosen but some non-linear combinations were tested as well.

Equation 14 shows these formulation options by a simple formula for calculating the utility of a residential location. The expressions with β and upper case letters are parameters of the model, which are calculated with specialised software (Biogeme – Bierlaire, 2006). The expressions in lower case letters are variables referring to characteristics of the location and the household which are used to calculate the utility of a location alternative for a decision maker. The non-linear formulation of the explanatory variable distance work is reflecting that the same absolute distance difference in short-distance has a stronger impact than in long-distance.

$$Utility_{ik} = \beta RENT * rent_k + \beta NOISE * noise_k + \beta DISTANCE\text{-}WORK * (distance\text{-}work_{ik})^{\beta\_EXPONENT} \quad (14)$$

i: Index of the deciding household
k: Index of the location of alternatives

Two location choice models were developed for the Greater Zürich area. All available variables from the household survey and GIS-derived variables were employed for a universal model. The results are presented in the third column of Table 10. The fourth column includes

a model adjusted for UrbanSim, considering only endogenously available variables in the simulation (see Bürgle, 2006b). The sign of a parameter indicates if a higher value of a variable increases the utility of location of a household (positive sign) or decreases (negative sign). For instance, a negative sign can be expected for the distance of the residential location to the work location as shorter distances are preferred. Most of the observed signs are not surprising. A relatively high rent compared to the household income impacts the subjective utility of a location alternative negatively for example.

The relatively high share of variables at municipality and traffic zone level particularly in the UrbanSim adjusted location choice model is interesting because these variables can be obtained with much less effort and costs than variables at a lower spatial level. A conclusion could be that a minimum standard of model fit can already be archived with low data requirements. For further model improvement, however, knowledge about socio-economic attributes of the decision maker is necessary. As documented in Bürgle (2006a), the model estimates differentiated by household type reveal that such characteristics clearly impact the location decision heavily. As already described for the hedonic residential rent model, the household location choice model had to be adjusted to the restricted data availability for the simulation area and period. For example, the work locations of the surveyed household members are available, but they are unknown for all households in the simulation area. Moreover, there is no data structure included in UrbanSim as of 2006, which could hold this information and update it over the simulation period. Therefore, the distance to the work location had to be dropped as an explanatory variable. As expected, the omission of this variable reduced the model fit considerably. The same is the case for dwelling unit size, which is not tracked in UrbanSim so far.

It is striking that with the omission of socio-economic information, some locational variables used in the hedonic rent model become significant as well in the UrbanSim adjusted household location choice model, which was not the case in the universal household location choice model. Moreover, the algebraic signs of the parameters of “travel time to Zürich CBD” and “federal tax income per capita” became similar to the values in the hedonic residential rent model.

The transport-related variables “accessibility” and “travel time to Bürkliplatz/Zürich CBD” deserve a closer look. Contrary to the hedonic residential rent model, only public transport accessibility is significant when households without cars are considered. Accessibility by car is not significant at all. In general, travel time to Zürich CBD has a positive sign, a conclusion, which might point to urban sprawl tendencies. In the UrbanSim-related model, the sign is

negative, similar to the hedonic residential rent model. The surprising reversion at the first glance is due to the fact that the travel time variable takes on the explanation for the affinity for central locations from the omitted variable "distance to work location" since jobs are concentrated in such central locations.

Table 11 shows the correlation of the variables related to accessibility from the universal model. The relative high correlation value for "public transport accessibility" and "travel time to Bürliplatz/Zürich CBD" is probably due to the selective consideration of household without cars.

Table 11 Correlation of accessibilities in the universal model

	Car travel time to Bürkliplatz/Zürich CBD	Distance to work location
Public transport asseccibility (households without car)	0,21	0,01
Distance to work location	0,02	

Since both the hedonic residential rent model and the household location choice model are concerned with the attractiveness of residential locations – although with two different problem formulations – a systematic comparison of the results would be interesting for future work. Where reasonable and possible, the same variables were used for the estimations.

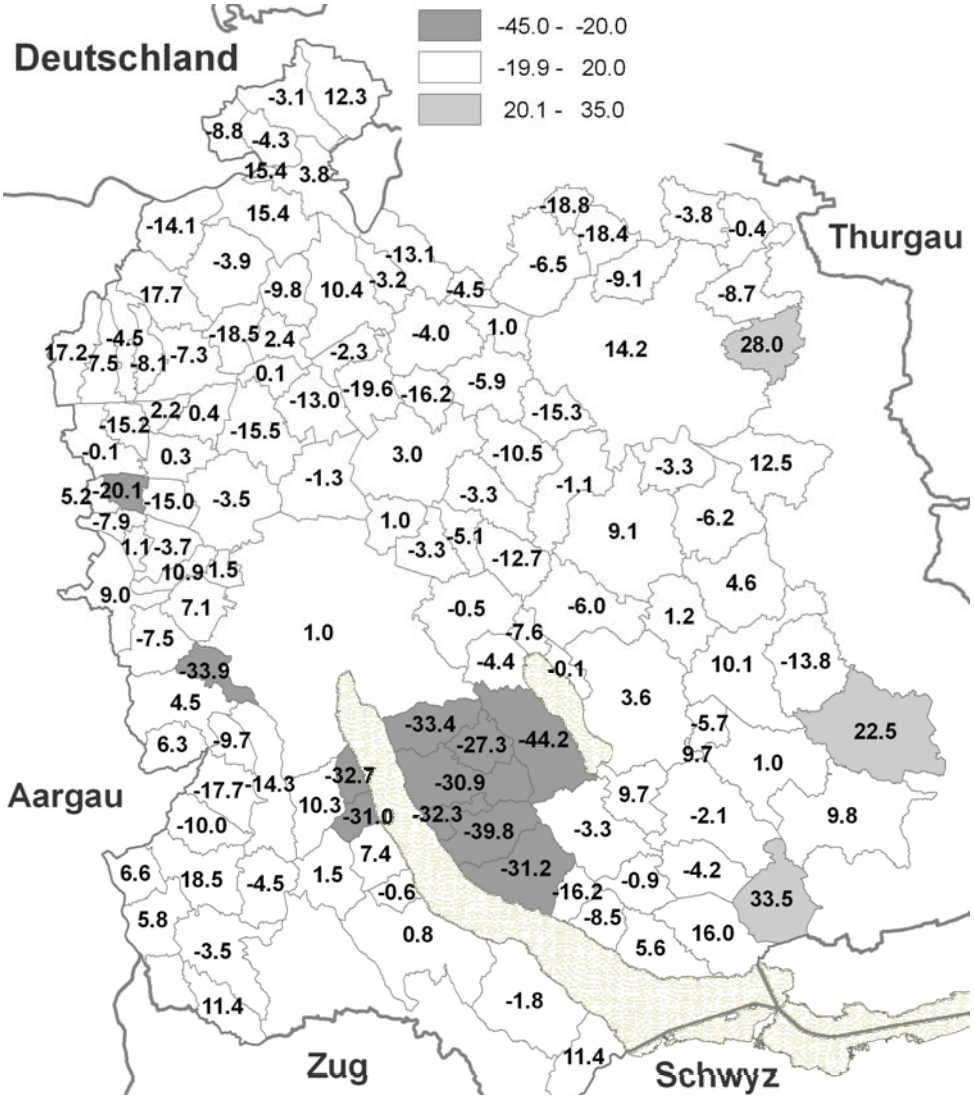
5.7 Validation

Results concerning the quality of the simulation results for the Greater Zürich area were obtained by comparing simulation runs for the years 1996 to 2001 with census data. But, this is just the beginning of the necessary validation work. Start and end time of the validation period are due to data availability. Particularly, the census 2000 and the employment census 2001 are important references.

As an example, Figure 7 reveals the discrepancy between the simulated household distribution per municipality in the year 2000 to the census data by substracting the census data from the simulation results. While the simulation results show approximately the actual development in a large part of the study area, some municipalities differ considerably from the actual value in the year 2000. In the simulation results, there are too few households in those municipalities

with above average residential rents. One reason might be an overestimation of the negative utility value of rents in the household location choice model. Municipalities with simulated above average household balances are situated at the fringe of the simulation area. It remains to be clarified if this is due to boundary effects.

Figure 7 Percentage deviation of households per municipality: simulation 1996 to 2000 compared to census 2000 data



Additionally, a comparison was conducted for the validation period composed of the simulated household distribution and a simple, calculated distribution. For all municipalities, the same percentage increase was assumed. The simulated distribution shows a lower deviation from the census on average with -3.7% compared to -10.8% for the calculated distribution.

Overall, it becomes obvious that validation and calibration work needs to continue. This can be done with various approaches: The improvement of the model fit of the used model specifications could result in a convergence to the real development, without changing the model structure of the simulation software. Additionally, the causes for the outliers should be systematically identified. For detecting sources of errors in the database, it is often necessary to retrace the operations of the simulation model by model, step by step. Finally, improvement by additional models is possible with some programming efforts in order to better exploit the properties of the available data.

5.8 Conclusions

Integrated land use and transport simulations have achieved a level of technical maturation, which enables in principle their use in practice. But, the best simulation system is dependent on the underlying data used for the model estimations and the base year data of the initial state. The efforts for data collection and preparation as well as the amount of uncertain information in the base year data, are two possible sources of problems, which have to be taken into account for the application of a simulation model as a planning support tool. However, a systematic investigation of the relationship between the necessary efforts for the implementation and the explanatory power of such a simulation remains to be done. Nevertheless, this should precede the application.

For the acceptance of such complex simulation systems in planning practice, not only a moderate implementation effort is necessary but also a certain degree of trust of the potential users. The base for that trust is built by well-founded analysis of the reliability of land use and transport simulation results that go beyond the model fit values of single model estimations. Useful approaches might be found for example in Ševčíková *et al.* (2007).

Besides methodological improvements of simulation systems, the early participation of potential users in the development process will play an increasingly important role for the practical implementation at a later stage. In a dialogue, users can communicate their needs and requirements and conversely, they are sensibilised for opportunities and limits of the validity of complex simulation systems (e.g. Borning *et al.*, 2008; Förster and Kytzia, 2004).

Within the project, the land use model UrbanSim were adapted and implemented for the Greater Zürich area. However, validation work revealed that there is more need for calibra-

tion. This remains to be done in future projects, in which UrbanSim can be applied in case studies within the Greater Zürich area for the comparison of development scenarios.

The question of transferability to other regions of Switzerland depends on data availability. As described, a basic requirement for the application of UrbanSim is a transport model in operation, which can calculate and export transport supply data such as travel time between traffic zones at an average traffic load. Beyond, availability of certain data besides the available population and employment census data are essential for the implementation. This comprises information about building capacity, distinguished by use. If this information is unavailable, rough assumptions have to be made which reduces the reliability of the overall simulation system.

Comprehensive modelling approaches as presented in this chapter cannot substitute specific sector planning models due to their complexity. If one decides on such an implementation, it opens new fields of analysis and appraisal opportunities for planners.

6 Location choice of retailers – An agent-based approach

6.1 Introduction

Microscopic models of travel and land use tend to have rich descriptions of the travellers, but the other actors in the urban system are normally abstracted into market clearing mechanisms. This chapter aims to make a central actor, the different retailers, explicit by developing an agent, which can address the location choice problem of a retailer, well knowing that this is only one choice, even if probably the most important, of a retailing firm. The new agent type will be integrated into the existing agent-based simulation MATSIM-T, developed at ETH Zürich and TU Berlin. MATSim-T is a simulation toolkit which is able to deal with large scale scenarios (Balmer *et al.*, 2006). It uses the concept of the Evolutionary Algorithm (EA) in order to generate consistent daily activity schedules for each individual (agent) of a population and travel times on the network. The resulting modelling system will be a first step to a much richer description of the urban system in this simulation toolkit in the future. It will be also a first effort in order to integrate the supply side choices, here for retailing, into such a system. In fact, also if the initial modelling of retailers is simple, the agent-based approach makes the enhancing of the model to a more sophisticated one relatively easy. In the short run, the main goal is to introduce this new agent for modelling the location behaviour of retailers. The shopping behaviour of individuals, in particular the location choice for the shopping, has a big impact on retailers' behaviour and it creates a cycle indeed. In the long run, we imagine thus that both retailers and individuals as consumers will have a much richer description. The work is also aimed to contribute to the current policy debate about road pricing. The evaluation of road pricing as a policy tool hinges to a substantial extent on an assessment of how the local retailers will respond to such charging. Even if road pricing is currently implemented in some cities (London, Stockholm, Singapore) the effect on spatial behaviour of retailers is not yet clear and definitely an open issue. A micro-simulation tool as presented here can be the right instrument to investigate this issue, especially when trying to analyze and compare effects in policy scenarios.

The remainder of this chapter is organised in four sections. In Section 6.2, an overview on retailers location choices is presented. A literature review on location strategies of retailers (Section 6.2.1) and on methodologies employed for the location choice of stores (Section 6.2.2) is enriched through the report of a series of in-depth interviews. In Section 6.4, the

available literature on effects of road pricing on the retail sector is also discussed and in Section 6.5, hypotheses are stated for later work. A literature review on agent-based approaches modelling retail markets is the topic of Section 6.6. Section 6.7 discusses the introduction of the new retailer agent in MATSim-T. The main issues arising with the introduction of the retailer agent are discussed, and a possible approach is proposed. Section 6.8 is dedicated to conclusions and to an outlook on the future work.

6.2 Retailers location choices

This section is intended to provide a general overview on most important contributions on location choices of retailers, covering both the strategic and the methodological levels. The section is enriched with insights of eleven explorative interviews, which were conducted with German and Swiss retailers from January until March 2008 (Löchl, 2008). This allows us to get an understanding on the actual practice of location decisions of retail firms with different sizes and in various product categories. A subsection describes particularities of the Swiss retail market. Finally, a short literature review on reactions of retailers to road pricing is given and hypotheses to be tested are given.

The literature argues strongly that location is the most important factor for the success of a retail store, and sometimes is also acknowledged as the only one. “Good locations are key elements for attracting customers to the outlets and sometimes can even compensate for a mediocre retail strategy mix. A good location therefore can lead to strong competitive advantage, because location is considered one of the elements of the retail marketing mix that is “unique” and thus cannot be imitated by competitors” (Zentes *et al.*, 2007, 143). Moreover, the importance of the location has increased during the last years because of more intense competition in most of the retailing markets. This growing interest on location can also be seen in the academic literature in different disciplines such as economics, marketing, geography, land use, city planning and operations research.

The location behaviour of a retailer can be analysed distinguishing two levels, the strategic and the methodological. Most retailers have a spatial strategy; its realisation is pursued with a specific methodology. This may be chosen from a wide spectrum of existing methodologies, ranging from extremely simple and not very scientific, sometimes simple intuitive, to complex, computer-assisted approaches.

In general, location preferences and spatial strategy of retailers are determined by the retail format. Krafft and Mantrala (2006, 193) define a retail format as comprised of stores that offer the same, or very nearly the same, variety of product categories. Moreover, a retail format represents a specific configuration of the retail marketing mix, which is maintained consistently over time. It includes the nature of merchandise, assortment and service offered, similar promotion, pricing policy, approach to store design and visual merchandising as well as typical location (Zentes *et al.*, 2007, 10).

Depending on the retail format, there are various location factors and preferences. Each one weighs location factors differently and involves various trade-offs. Therefore, the appropriateness of a specific site is based upon the retailer's format and strategy and is influenced by a substantial number of factors that need to be investigated (see Table 12).

Table 12 Selected location factors

Customers	Accessibility	Competition	Costs
numbers by demographics (e.g. population size, age profile, household size)	pedestrian flows	existing retail activity (direct/indirect competitors, anchor stores, cumulative attraction, compatibility)	purchase price
	pedestrian entry routes		building costs
income level	road network (conditions, speeds)		rent costs
disposable income per capita	parking (capacity, convenience, cost, potential)	existing retail specification (selling area, turnover estimates, trade areas, age of outlet, design, parking)	leasing terms
employment by occupation, industry, trends	car ownership level		site preparation
housing density	public transport (types, cost, ease, potential)		building restrictions
housing age/type	barriers such as railway tracks, rivers	competitive potential (outlet expansion, refurbishment, vacant sites, interception, repositioning, competitor policy)	development concessions
neighbourhood classification	visibility		rate payable
home-ownership levels	type of location zone		refurbishment needs
building/demolition plans	access for staff		maintenance costs
main employers	access for delivery	saturation index	security needs
spending patterns			staff availability
shopping patterns		proximity of key competitors, traders, brand leaders	labour rates
population growth, density and trends			delivery costs
lifestyle measures			insurance costs
cultural/ethnic grouping			promotional media/costs

Sources: adapted from Zentes *et al.* (2007), 148; McGoldrick (2002), 240; Gilbert (2003), 293

6.2.1 Location choice strategies of retailers

The specific literature about location strategies of retailers is by far less abundant than the one on methodologies. As Brown (1992) observed academic conceptualizations of retailing location have changed little in the last thirty years. Indeed, this conceptual stasis stands in marked contrast to the dramatic methodological developments that have recently taken place. In fact, this observation can be considered valid also nowadays (Cliquet and Josselin, 2002; Rogers, 2007). Traditionally, location choices were more based more on intuition and on a responsive basis to competitors but also to other stakeholders such as government, legislators, etc…- and therefore often made with a short-term horizon. More recently, many major changes in the re-

tailing environment such as the need to justify decisions to shareholders, increased competition for retail sites, fast changing demand, the disappearance of 'obvious' sites, the increasing cost of location decisions and so forth brought them to a more organic approach (Clarke *et al.*, 1997, Hernandez *et al.*, 1998, Hernandez and Biasotto, 2001). Sometimes it seems that in many cases retailers attempt a sort of a posteriori validation of strategies more than a conscious path following one planned strategy of development. Naturally that does not fit, in most of the cases, larger retailers, whose expansion strategy is usually following an accurately planned and well recognizable pattern.

One systematisation of retailers' spatial strategies is for example used in Laulajainen (1987). In his study, the growth of some retail chains in Sweden and in the US is analysed, distinguishing between contagion and hierarchical expansion. The *contagion* strategy consists of an expansion from areas already occupied to new areas following a vicinity principle. In that way, the growth takes place from a unique central point. Wal-Mart followed such an expansion strategy in the US, where it allowed the company to exploit economies of density in distribution, training and advertising (Holmes, 2008). In contrast, when using the *hierarchical* expansion, a retailer decides to colonize first bigger centres, disregarding the geographic vicinity from already existing own stores, and then tries to saturate the local market expanding to increasingly smaller centres following a hierarchical path.

No one of the interviewees in both Germany and Switzerland reported a contagion strategy. This might be due to the fact that both countries are relatively small and each region can be delivered from a central distribution centre within a few hours. Even discount retail chains in the early market entry and expansion phase prefer to pick the most promising or available sites for new stores without considering their current stores locations and distribution centres.

Anyway, the aforementioned classification does not take into account the effect of the competition on spatial decisions of retailers. In fact, they face two opposing incentives in the location decision.

A retailer might have an incentive to locate close to competitors in an attempt to capture more consumers, but spatial and even product differentiation is reduced and leads to greater competition in the price dimension. Therefore, a firm has an incentive to locate farther from its rivals in order to reduce price competition. This potential ambiguity is reflected in Brown (1992), who proposes a classification, which is uniquely based on the behaviour of retailers versus their competitors. Three different spatial strategies of retailers are recognised: avoidance, confrontation, and predation. *Avoidance* means that the retailer seeks a site as far as possible from

its competitors trying to meet some other requirement (i.e. nearby population, quantity and quality of vehicular traffic) at the same time. In contrast, *confrontation* means that the retailer wants a site as near as possible to the competitors. Hotelling (1929) already observed that when the demand is elastic and the good sold is not much differentiated there is an interest to place a shop in the heart of the market, instead of trying to colonise a virgin area. This approach is quite typical for the furniture market. The last, and maybe most interesting, strategy is the one of *predation*. In this case the retailer tries to fill up spaces left free by the competitors, and once installed near other shops of the same type tries to capture other customers with a price strategy. This strategy is often used in the clothing market.

An interesting analysis is also provided by Karande (2003). Even if it is reported only for location strategies of broad-line specialist retailers, there is an attempt to explain why a certain strategy is used in a given situation, referring in particular to vicinity and distance strategies. Vicinity strategy is justified with the attractiveness for potential customers of doing "comparative shopping". But if "darwinistic" effects prevail, the competition to get customers is the dominating factor, therefore distance strategy is the one preferred. Moreover, it is observed that areas where the two different strategies are implemented are also differing. A proximity strategy tends to be adopted by retailers in areas with relatively high income, high retail expenditures, high population density, younger population and high home ownership.

The actual strategies differ by retail format and product group as the interviews revealed. Generally, clothing retailers look for the agglomeration in city or shopping centres, as neither of them could attract enough customers alone. Therefore, they avoid solitary locations, but they observe any actions by the competitors closely. Others such as sport and outdoor retailers do not generally avoid proximities to competitors, but they acknowledge oversaturated markets in order to prevent ruinous price competition. Moreover, they are sometimes willing to locate away from the agglomeration and do not avoid solitary locations. Grocery retailers follow a different path. They are interested in a dense network and a good market penetration. Therefore, they disregard the competitors and focus on the best locations in regard to market potential for them. Sometimes, they look for joint locations (i.e., supermarket and discounter) when there is a truly minor product overlap.

Based on the interviews the preferred location types of various retailers can be roughly categorised into the following:

- top locations with high frequencies in city centres or shopping centres,
- semi-top locations at the inner city fringe or in subcentres as well as in less frequented shopping centres,

- typical specialist retail warehouse locations,
- locations close to residential areas.

The process of location selection often follows a three step approach with increasing spatial focus (Brown, 1992, 16, Zentes *et al.*, 2007, 147):

1. *Market selection*: Interesting agglomerations, regions or municipalities are filtered based on inhabitants or purchasing power.
2. *Area analysis*: Within the chosen region, a potentially optimal area for the store is selected. However, this step is often neglected in practice, as anything after the market selection is driven by site supply.
3. *Site evaluation*: The best available sites are examined in terms of all features that are relevant to potential store performance.

Certainly, location choices are not limited to simple opening and closing of shops. In fact most of the location decisions include other activities (Davidson *et al.*, 1988; Ghosh, 1990; Ghosh and McLafferty, 1987). More recently, this is also reprised in Hernandez *et al.* (1998). In principle, location strategy is concerned with the size, type and number of outlets coupled with an assessment of intended timing and geographical spread of the chain. Basically, it means to maximize aggregate returns by planning and adjusting a shop portfolio in such a way that each unit is matched closely to the market within its immediate catchment area, or is deleted if this is not possible.

In this optic, six main types of decisions can be identified, which, in descending order of risk, investment commitment, and time taken to implement are as follows:

1. *Roll-out/extension*: Increase in floor space either by the opening of new stores (by organic growth) or through acquisition or extending the floor space in existing stores.
2. *Relocation*: Relocation of existing stores.
3. *Rationalisation*: Closure of stores on an individual basis or selling entire divisions.
4. *Re-fascia*: Altering the image of outlets by changing the name and appearance.
5. *Refurbishment*: Changing the retail appeal with some modification to the store.
6. *Remerchandising*: Altering the product range and merchandising of a retail location, and tailoring the offer to the local consumer.

6.2.2 Retailers' location choice methodologies

The rising competition in most retailing markets has not only increased the interest on location issues, but has also caused major change in retailers' approach to location choice. This is particularly true when it comes to available methodologies, much more than it is the case for strategies. The literature reports that many retailers are passing from really simple location choice techniques to more modern and sophisticated approaches. Even if simple "rules of thumbs" techniques keep being diffused and also simple intuition and experience of the retailer are still actual tools when deciding the location of new shops, the number of retailers using newer, often computer-based, methodologies is increasing internationally. Apart from competition, another explanation of that is the explosion that has occurred in the quantity of available retail data. Moreover, most of it is of spatial nature, about 90%, pushing retailers to the adoption of GIS (Geographic Information System) systems in order to treat such geographical data. Hernandez *et al.* (1998) observed:

> "Although formal techniques of locational analysis have been available for over 50 years, most retailers traditionally made no use of them, relying instead on intuition guided by experience and "common sense". However, the simultaneous advent in the last 15 years of low cost computing and the increasing availability of retail related data of all types has given retailers the opportunity to take a much more rational approach to decision making."

Among the abundant literature produced in the last ten years, some good examples are Hernandez *et al.* (1998), Hernandez and Bennison (2000), Birkin *et al.* (2002) Pioch and Byrom (2004), Wood and Browne (2006) as well as Rogers (2007). Some classic literature on the subject is Hotelling (1929), Applebaum (1966), Craig *et al.* (1981), Ghosh and McLafferty (1987) and Brown (1992).

A classification of location methods which separates them according to the type of analysis involved is the one proposed by Hernandez *et al.* (1998). Reporting on a survey of UK retailers, they include ten methods grouped in three categories:

1. *Comparative:* Rules of thumb, checklist, ratio, analogues.
2. *Predictive:* Multiple regression, discriminant analysis, cluster analysis, gravity models.
3. *Knowledge based:* Expert systems, neural networks.

Most of these methods are described below. Descriptions are adapted from Birkin *et al.* (2002) and Rogers (2007).

- *Rules of thumb* simply mean that experience is the main instrument. Essentially subjective and intuitive guidelines or developed from knowledge of the company and sector, and tempered by "common sense", are used.
- *Checklists* consist of a list of variables considered to have an influence on store performance and perhaps given some variable points rating.
- *Ratio and market support* methods are simple procedures which may be used when data on the existing facility is very limited. In both of them, the trade area of the store is estimated along with the population in it and the relative spending potential. An inventory of potential competitors of the store in that area is also compiled.
- *Analogue methods* are still commonly used despite their simplicity, indeed they are among most popular methods since the 1930s, and still widely used especially among retailers with limited risk related to one store's opening. The idea is to forecast and compare the sales of a store with sales of another store of the same retailer. The latter store should have a similar location in terms of population, geography and trade, compared to the one for which the forecast is desired. The data and computational requirements are small and the costs are low.
- *Catchment area models* are a more sophisticated variant of analogues methods. The main advantage of this method is that the unique profile of a retailer is accounted for. Moreover, it is cheap and simple and might be applied even if the retailer owns only a few prototype stores. The analogies are still based on qualitative evaluations but relies more on sociodemographic characteristics of inhabitants of a certain area. The difficulties to account for interrelated impacts of competition, demographics and distance at a new location, traditionally the main shortcomings of analogue methods, have been successfully overcome by integrating cluster analysis, decision-tree analytical frameworks and statistical modelling in this technique.
- *Rating models* are a good option when markets are very complex. In particular they are suitable when sales are highly unpredictable and based on an irregular clientele, as in the case of use of ATM's. The method consists on giving a rate for each of a predefined group of characteristics of the site. According to such rates the site receives a final score which can be put in relation with an expected level of sales. The method is particularly suitable for a computer based use. However, it has not proved particularly successful so far.

- *Statistical models* are another popular group of methods. Their application is typical in highly segmented markets, such as clothing, books and so forth. Linear multiple regression, the most common of these methods, may be considered an analytical version of analogues methods. In fact it is assumed that sales are dependent on a group of variables which can be measured, and forecasted solving an equation of the type:

$$Y_i = z + a_1X_1 + a_2X_2 + \ldots + a_nX_n \qquad (15)$$

  Coefficients are calculated using already existing stores attributes and sales level. These methods have the advantage to reduce the degree of subjectivity involved in the process and the increased availability of computation power made the application of these methods easier. However, they are often misapplied mainly because of objective difficulties of taking into account all relevant factors. Moreover, a model produced with this method, also a good one, is a sort of a snapshot of the performance of a store of a given retailer at a given time. Thus, only a short term forecast of the sales may be imputed.

- *Cluster and factor analysis* are aimed at grouping data cases and variables together for example, segmenting a portfolio of stores into similar groups (clusters) or grouping together a range of variables which can be used to predict profitability (factors). These techniques are particularly suited to new store format development and network segmentation. However, as with the previous multivariate techniques, they require a combination of statistical expertise and business acumen, along with relatively large quantities of "good quality" data.

- *Gravity models* assume that spatial issues (like distance from a customer's home) play a key role in the attractiveness of a store. Sales are forecasted taking into account distance relationships between competing facilities and population distribution and density. They are frequently employed for evaluating aggregations of stores, such as shopping malls, or extended store networks, like supermarket chains, or fuel station chains. A crucial aspect is that a correct use of gravity models requires the availability of several different data sets. The main problem of these methods is that segments of population are not accounted for, and therefore they might be not suitable for every segmented market.

- An *expert system* is a computer program which contains the knowledge of one or more experts. That means that the system will be able to autonomously suggest solutions to the location problem dependent on a set of input data. In fact, especially in its simpler implementation, it is conceptually similar to rules of thumbs, but with the difference that decisions are more consistent and coherent from case to case. *Neural networks* work similarly. An artificial neural network is a computer program which has an initial

knowledge and the capacity to "learn" if properly "trained". Thus, again, the system is able to autonomously propose solutions. Both methods require a relatively complicated algorithm to be used (at least if compared with simple traditional methods), and are not widely diffused in practice.

After this brief overview on location methods it should be noted that, in practice, techniques that are used most widely are the simpler comparative ones, requiring less expertise and incurring less cost than the more complex ones. The interviews revealed that a lot of location decisions, particularly those at the micro scale, are actually based on intuition and involve less sophisticated methods. These are also the ones with the greatest degree of subjectivity in their formulation, calibration and interpretation, and imply the continuing existence of decision making cultures, which are opinionated and politicised. Moreover, many retailers are habitually using different methods in parallel, and not trusting just one method. Therefore, even when the most sophisticated methods are applied experience and intuition are still essential for successful location choices.

6.3 Retail situation in Switzerland

The Swiss market is relatively small with 7.6 Mio. inhabitants in 2007 and only five cities with more than 100,000 inhabitants (Basle, Berne, Geneva, Lausanne, Zurich). Particularly for retailers which are looking for top locations in inner cities or shopping centres, there are only very few local markets with few potential spots. Therefore, the demand focuses on a small area where rents are extremely high and expansion is restricted by the short supply of available sites. This situation explains the relatively low level of methodological sophistication reported by retailers of products with longer life spans.

The grocery retail chains are looking for locations close to residential areas with good accessibility. Moreover, they pursue a seamless and dense market presence. This is particularly true for the two predominant grocery retailers in Switzerland, Coop and Migros, which together have a market share of more than 46% in the food sector. The two German discounters Aldi and Lidl have recently started to expand in Switzerland. They require more parking and therefore they are found more often at the periphery of settlements. They also want to have a seamless market presence, although their minimum catchment areas per shop are considerably larger than those of the two leading retailers. While Coop and Migros require at least around 8000 inhabitants for a 700sqm-shop and 2,000 inhabitants for smaller (convenience) shops, the discounters require a minimum of around 20,000 inhabitants for a shop. Nevertheless,

those catchment area sizes are considerably lower than those in most other retail product groups.

6.4 Location of retailers and road pricing in the literature

Road pricing is often mentioned as an effective measure for sustainable transport in metropolitan areas. A broad body of literature about this topic is available, focusing mainly on short-term transport, economic and acceptability aspects of pricing policies. Papers focused on spatial aspects of road pricing are relatively rare, and only recently a stream of work analysing them is starting to emerge. Probably the reasons are the few implemented examples around the world and their short duration in service, too short to observe long-term effects on location choices of retailers. In this sense, it is not surprising that a fair number of reports are ex-ante studies attempting to answer the question of how the introduction of a road pricing measure would have influenced the land use. None of those studies is aimed to investigate effects on retail in particular.

However, there are a few ex-post analysis results available. For the London case, what can be measured by now is the level of sales, and eventually its modification, but relocation effects have not been observed yet. Impacts of the congestion charge are investigated in detail by applying econometric models in the work of Quddus *et al.* (2006). One of the datasets used includes weekly sales data for six stores of a retail chain in London, whereas one of the stores is located in the charged zone. The authors found a significant negative impact on sales at this store over a period of about eleven months following the introduction of the charge compared to the other five stores of the chain. However, the competition among shops and the spatial redistribution of sales could not be considered in the analysis. A second dataset covering total retail sales in central London did not show any significant downturn because of the charge. This supports the suspicion of spatial redistribution. Therefore, no congestion charging impact could be found as a whole. Nevertheless, it is possible that there has been some redistribution of sales from certain areas to other stores within central London (Quddus *et al.*, 2006, 20). Similar conclusions can be found in the reports of Transport of London (TfL, 2005, 2007). In the latest report, TfL summarises that no general evidence of any clear differential impact of the central London congestion charging scheme on business activity had been found (TfL, 2007, 95). Earlier, it has been stated that the congestion charging had no significant effects on retail sales (Tfl, 2006, 88) and that some sectors within the charging zone have shown better performance than outside the zone. Other sectors have performed worse inside the zone than outside. These differences are all relatively small, and are not consistent between different

datasets. It is not possible to be certain what part of these differences (positive or negative) result from the congestion charge (TfL, 2005, 5). Therefore, the effects of road pricing often disappear among other factors, which might have a higher influence. Overall, the impacts seem to be rather neutral so far. Nevertheless, one has to take into account that land use impacts are rather long term and the duration since the introduction is still too short to draw final conclusions, particularly since the TfL analysis is limited due to long lags in the availability of published economic and business data (TfL, 2007, 95).

No effects on retail revenues for shopping malls and shops located within the toll area were reported from Stockholm, where a road pricing scheme was introduced in January 2006 (Daunfeldt *et al.*, 2007). The authors suspect that the customers are avoiding the time-depending road pricing by changing the time of day when the shopping is being done. Moreover, they note that shopping behavior is primarily determined by habits that change slowly and that therefore long-term effects were not observed in the data, which covers the time until December 2006.

Similar conclusions can be found in cities with longer road pricing experiences. In Singapore for example, a road pricing scheme was first introduced in the inner city in 1975. Reporting on this experience Armstrong (1986) observed that any impact that the scheme may have had on land values, land use and the environment in Singapore, has been largely eclipsed by other factors in the economy. In the case of Trondheim, there was some research conducted on shopping behaviour after the introduction of road pricing in 1991. A study (Avant Management A/S, 1992) found that 10% of the customers had changed their shopping behaviour by moving their shopping to other destinations or times after the introduction of the cordon pricing (cited in Tretvik, 2003, 88). Moreover, while business people located in the city centre had predicted major negative swings in trade prior to the cordon pricing, the Chamber of Commerce of Trondheim concluded from its own ex-post survey that there was hardly any effect on trade at all. Anyway, there was a long lasting general trend of growth in areas outside and decline in areas of the cordon. Tretvik (1999) even concludes a general trend of modest but steady growth in retail sales in real terms inside the cordon since the introduction (Tretvik, 2003, 89).

In conclusion, it seems that effects on retail are minor particularly at the aggregated level, but for some cities the introduction is too recent in order to allow a definitive word. In any case, it is always hard to separate the individual effects of road pricing from other factors and impacts at the spatial micro-level are still rather uncertain. Moreover, it is difficult to know to what extent experiences and conclusions drawn from one city can be transferred and generalised to

other cities, since the effects depend to a large extent on the road pricing scheme as well as on particular characteristics of a city such as its spatial structure, street network form and capacity, quality of public transport network and so on (Löchl, 2006). Therefore, a micro-simulation, considering all these factors seems to be the right instrument to assess the possible effects of a road pricing measure on a city by comparing policy alternatives.

6.5 Location of retailers and road pricing hypotheses

It is eventually planned to test road pricing scenarios with the implemented retailer module in MATSim-T for the Greater Zurich area. The analysis will be led by the hypothesis that depending on the spatial range and fee schedule of a road pricing measure, it can have minor or major impacts on the retail sector. Therefore, focus will be not on the analysis of a general increase of transport costs due to road pricing but on spatially varying pricing schemes, such as cordon charging. High charges might encourage both a transport mode shift to public transport and a destination shift to retailers outside the charging area. Changes in route, mode and destination choice by consumers can be expected. As a consequence, they might benefit from the reduction of density losses due to less congestion to and within the cordon. But in case retailers relocate outside the charged area, the advantage of having an agglomeration in the inner city might diminish when retailers are spreading beyond.

More specifically, the following road pricing scenarios are going to be tested:

1. *Scenario A*: Road pricing at an inner city cordon in Zürich for private cars only, reduced charges for inhabitants inside the cordon.
2. *Scenario B*: Time depending fee schedules for the inner city of Zürich.
3. *Scenario C*: Cordon pricing for the whole City of Zürich; reduced charges for inhabitants inside the cordon.

The following hypotheses are stated with regard to the aforementioned road pricing scenarios:

Scenario A: The road pricing will impact retailers inside the cordon slightly, nevertheless spatial redistribution occurs. As the pricing is set higher, alternatives beyond the charging zone become attractive for customers living outside the cordon. In this situation retailers might increasingly prefer to establish their business outside the cordon in the long run.

Scenario B: With smart scheduling of fee levels over the day and over the week, road pricing effects on retail can be mitigated.

Scenario C: Applying the same fee levels as in Scenario A to the whole city of Zurich, impacts on retail are comparably lower as more customers are living inside the cordon and enjoying reduced charges. Consequently, agglomeration effects in the city centre are stronger.

To test these scenarios one can build on current experiences with MATSim-T, as there were efforts to simulate a time-dependent toll scheme for the City of Zürich (Rieser *et al.*, 2007). It has to be observed that the micro-simulation approach does not guarantee that if a fixed point is reached this point is an optimum (in the sense of Nash: nobody can improve his situation by unilaterally doing something else), because the system is not deterministic. The iterative process might find a state that is locally stable, but there is no guarantee that running more iterations might not lead to completely different solutions. The system's complexity is even increased introducing the retailer agents. That gives further support to the idea that solutions might be path-dependent. For that reason not only the relaxed state will be considered, but attention will also be paid to initial condition and to the adaptation behavior of the agents.

6.6 Modelling retail markets: Agent-based approaches

The use of agent-based modelling in spatial issues is becoming more popular and is opening up new perspective in the understanding of retail markets. The popularity of agent-based systems in itself is a logical consequence of the increased calculation power of computers, but for the application to retail markets the availability of more and more precise data, it plays a crucial role. The introduction of new data collection technologies such as EPOS (electronic point-of-sale) systems and the increasing popularity of store cards simplify the task to obtain detailed information on consumers for retailers (Nakaya *et al.*, 2007). However, the use of the agent paradigm to describe retail markets is usually limited to the agent-based modelling of customers and not retailers. Agent-based implementations of retailers are rare so far and models where both retailers and customers are modelled as agents are even fewer (Arentze and Timmermans, 2007; Lombardo *et al.,* 2004).

Speaking about spatial choices in general some microscopic approaches can be found in the simulation of urban housing markets (Benenson, 1998; Bura *et al.*, 1996), traffic flows under different land use conditions (Miller *et al.*, 2004) and land use transformations (Parker *et al.*, 2003). More specifically, in retail geography micro-simulation approaches were initially used to merely overcome data deficits. The first attempt of this type was probably the one of Birkin and Clarke (1988) and one recent example can be found in Nakaya *et al.* (2007). Such models just solve a part of the problem that we confront, the set up of a micro-population, but they use it in the context of traditional methodologies. In this sense they can actually not be called agent-based, even if agents are somehow defined and used. Since then, only in recent years

fully implemented versions of retail market models appeared. Most of them use the agent paradigm to give a detailed representation of customer behaviour. More interesting are proper simulation experiments. The two sides of retail markets, supply and demand, are usually not both represented as agents, since typically only consumers are. In this vein of work, Heppenstall *et al.* (2006) use a hybrid approach where the petrol market is represented by combining an agent-based approach for the supply side and a spatial interaction model for the consumer side. Individual petrol stations are represented as agent-objects and supplied with knowledge of their initial starting price, production costs and the prices of stations within their neighbourhood. The location of an outlet is considered fixed and retailers are allowed to react to the sales only by adjusting their price to competitors' prices, which are considered to be known. The problem of the location is not directly tackled but the model successfully reproduces consumer's spatial choices observed in real markets and also the profitability of single retail outlets in the long term. In the market used as a case study in this work, the product sold can be considered homogeneous, which is an advantage but also a limitation. The lack of complex trade-offs typical for many other retail markets is of course a considerable simplification, but the representation of a composite retail market, where an entire array of products would be available to the customers, would imply a substantial modification of the model. Schenk *et al.* (2006) use a micro-simulation approach to model the shopping behaviour of inhabitants of an entire region of northern Sweden. In this case, the agents of the simulation are only representing the demand side at the household level. Nevertheless, the model is accurate and the supply side is modelled with a high spatial resolution. In particular, it is one of the few examples were prices are taken into account in order to characterise a retail store. The family agent is described with the socio-demographic attributes of its components and by some family specific attributes such as size and income. Agents can select stores by evaluating a bundle of attributes such as distance from home and work, location in the agglomeration, price, assortment, quality etc. The simulation is implemented using the SeSAm multi-agent simulation shell (www.simsesam.de), and the model is tested comparing the calculated with the simulated turnovers of the stores. Another example of application of multi-agent simulations to retail markets is Van Leeuwen *et al.* (2007). A discussion about possible applications of micro-simulations in spatial analysis is followed by a small scale simulation example. However, the simulation exercise is conducted "by hand" and, again, only the consumer side is described by agents, in this case at the household level.

In Lombardo *et al.* (2004), both sides of a retail market have agent-based descriptions. The multi-agent system is integrated in a GIS. The aim of the consumers is to reach the stores with minimal generalised costs. The aim of retailers is to maximize their profit. Shortcomings of this work are the summary description of both agents (basically within an agent type they are

completely undifferentiated) and the simple way in which the environment is depicted (only eight macro-zones with few links connecting them in which 80,000 consumers and 12,000 retailers are allowed to evolve).

Another stream of work tries to integrate land use with transport models and has already a relatively long tradition; some older works of that kind are DRAM/EMPAL (Putman 1983) and TRANUS (De la Barra *et al.* 1984). Recently land use and transport models have moved from aggregate models of various types to discrete choice logit models and more recently toward disaggregated, activity-based models. Recent examples are UrbanSim (Waddell 2002) ILUMASS (Beckmann *et al.*, 2007) and the work of Arentze and Timmermans (2000, 2007). In the case of ILUMASS for example, where all actors of the different markets (land, work, retail) were represented as agents, the project ended without matching the initial objectives. The simulation tool, especially the transport module, was too heavy. A simulation of the Dortmund region, the final goal of the project, was never performed and only small test scenarios were run. A complete report on its development, its achievements and its failure can be found in Wagner and Wegener (2007). UrbanSim applications have not experienced such problems, but a fully agent-based implementation is not yet realised and not the actual goal of the efforts. Arentze and Timmermans (2005) presented a multi-agent model of consumer behaviour which, besides the structural attributes of and the distance to the store, also includes opening hours as part of the institutional context of the shopping destination choice. Later (Arentze and Timmermans, 2007), they provide probably the only fully implemented example of an integrated land use and transport model where both, retailers and customers, are modelled as agents. For the customers' side, the model makes use of ALBATROSS (Arentze *et al.*, 2000), an activity-scheduling model that generates a day plan for each agent of the simulation. The supply side is made up of facilities, which are classified according to different demand types and sectors. Moreover, it distinguishes among three structurally different facilities: elementary (only one activity is possible), mixed (the sum of several different elementary facilities in the same place) and higher level (several elementary facilities organised in a "higher level" structure, allowing also activities not possible in elementary facilities). For each demand sector, an agent controls the development of the facilities network, with sub-agents controlling the development of a certain type of facility in the sector. The agents seek a location for the facility using a catchment area analysis. Once facilities are located, the simulation allocates the demand and supply agents are allowed to rearrange their facility. A test case, where a mid-size town is simulated, shows that the model is able to reproduce reasonably well location patterns of suppliers and a sensitivity test shows that the model correctly reacts to modification of main parameters.

6.7 An agent-based model of retailers in MATSim-T

In the current implementation of MATSim-T (Multi Agent Transport Simulation Toolkit) (www.matsim.org), only individuals are modelled as agents. The other actors interacting with individuals in the system (e.g. firms, retailers, planners, developers, public authorities, etc.) have a static representation, which is the typical approach in agent-based travel demand simulations. Nevertheless, in principle all those actors can be represented as agents and be allowed to interact dynamically with the individual agents in the simulation. This is not only possible, but also desirable since interactions between the different actors of the system are of dynamic nature and interdependent at various levels. In case of retail markets, suppliers (retailers) and individuals (consumers) are the players of this interaction. Customers, choosing the location of their shopping activities, determine economic viability of stores and congestion in the system. Retailers, by means of different location strategies (in the wider sense of the term, see Hernandez *et al.*, 1998), influence the location choices of customers (Lombardo *et al.*, 2004; Timermans and Arentze, 2007). Moreover, compared with traditional approaches, the microscopic one allows representing the heterogeneity of the decision makers in the system.

As a first step to a fully agent-based representation of the simulated world, a new agent type for retailers is introduced. The main scope is to correctly reproduce the location choices of retailers and to recognise how such choices and shopping location choices of individuals are mutually interdependent in the model system. In the future, other typical choices of retailers, such as price policies are modelled. The creation of a new module - implemented in Java code - specifies all attributes and the functionality of retailer agents. In addition, the representation of the customers has to be enriched with respect to shopping. The main goal of this section is to present a conceptual model of the agent-based retailers and to discuss the issues arising with their introduction into the MATSim-T toolkit.

6.7.1 MATSim-T overview

So far, MATSim-T is basically a fast microscopic transport model, where the supply side is modelled as fixed constraints of the system. The core idea of MATSim-T is that each single actor of the transport system, both on the supply and the demand side, can be simulated. In the current version of MATSim-T, each traveller of the real system is modelled as an individual agent in the simulation. The agents are able to make decisions, according to given information and coherent with a predetermined goal. An assumption is that transport is for individuals a derived necessity, in relation with the primary need of individuals to perform certain activities during the day. Therefore, for each agent a so-called *plan* is generated. One plan contains in-

formation on activities that are planned by an agent for a certain time span, typically one day. Not only activities are listed, but it is also specified where and when those activities will be performed, and which mode of transport will be used to reach the different locations. More details can be found in Ciari *et al.* (2007) and Meister *et al.* (2008). The plans are executed simultaneously in the traffic flow simulation. Agents are able to learn. Several plans for each agent are retained, given a score, and compared. The agents keep the plans with the highest scores, while creating new plans based on their previous experiences. Note that in this decision-execution process there is interdependency since decisions are conditioned by traffic conditions and traffic conditions are definitely dependent on agents' decisions in turn. This creates an iterative process in which the agents are learning from the simulation outcome in order to obtain better scores for their plans. The system iterates between plan generation and traffic flow simulation until a relaxed state is reached.

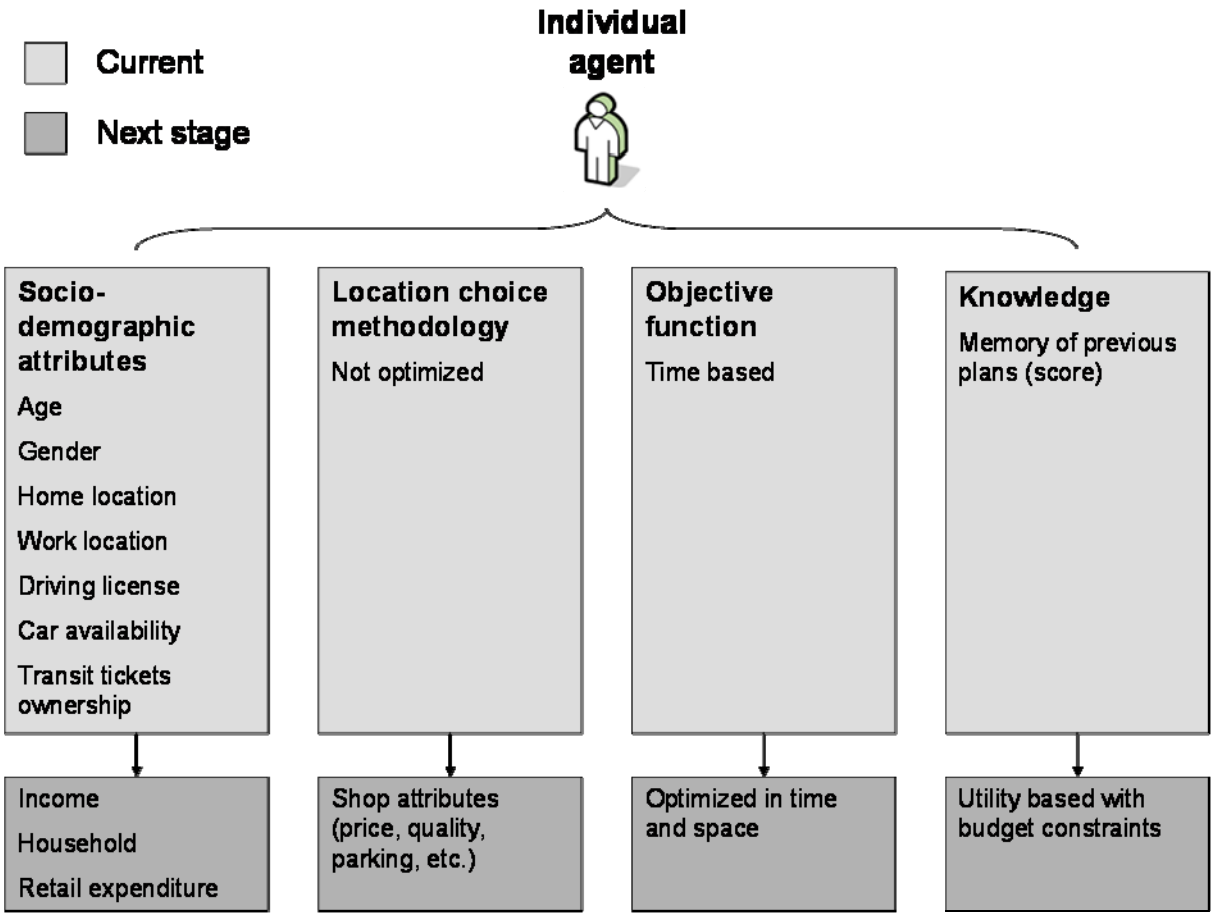
Among the literature on micro-simulations - increasingly abundant in recent years - a group of studies to which MATSim-T can be compared is the one based on activity-based demand generation. Several examples of implemented simulations of this kind can be found, such as (Abraham *et al.*, 2005; Arentze *et al.*, 2000; Bhat *et al.*, 2005; Bowman *et al.*, 1999; Vovsha *et al.*, 2002). However, most of them are able to produce results like origin-destination matrices, which are then used to dynamically assign the traffic to the network. An exception to that is TRANSIMS (2008), which generates individual activity plans as input to dynamic traffic assignment packages (see Axhausen and Gärling, 1992, for an older implementation of this approach). A more detailed comparison of TRANSIMS and MATSim-T can be found in Balmer (2007). Another vein of agent-based modelling works where similarities with MATSim-T can be found is agent-based land use models. An example is ILUTE (Salvini and Miller, 2005).

MATSim-T's most prominent application is a simulation of the travel behaviour of the whole Swiss population, which means application of the agent paradigm to as many as 7.6 million individuals. Socio-demographic attributes, like age, gender, driving license ownership, employment, etc., are generated and assigned based on the Swiss census data (ARE & BfS, 2001). The base network is the one of the Swiss national model (Vrtic *et al.*, 2007) with 24,000 nodes and 60,000 links, but networks with a higher resolution are also available. With a computer equipped with 4 dual core processors (2.2 GHz frequency each) a relaxed state is obtained with approximately 100 iterations, which means about 3.2 days (around 80 hrs) of computing time. Results are completely disaggregated and analysis can be performed at any level of resolution in space and time, and for any individual agent.

6.7.2 Individuals

All current agents of MATSim-T represent individuals. They are described by a group of sociodemographic attributes (age, gender, employment, driving license ownership, home location, etc.). Agents can choose the time when to leave from home, the transport mode to travel with and the route to concatenate all activities. Currently, there are five different activity types: home, work, education, shop and leisure. Note that some activities, namely home, work and education, are always performed in the same facility (for a given agent) in principle, while the others (shop and leisure) might be performed at any facility where the agent is allowed to do so. The location choice for these latter activities is now modelled in a very simple way and another ongoing work, being developed parallel to the one presented here, aims to improve it. There, the activity location choice will be optimised according to distance, time and capacity of the facility. The knowledge of an agent is limited to the plans already simulated and scored. Basically, the score of the current plan and those of a certain number of previous plans (simulated at a previous iteration) are compared. The objective of the agent is to maximise this score, where the only variable taken into account is time, with a positive value when it is spent in an activity and with a negative value when it is spent to travel. More details on the demand generation and agents in MATSim-T are described in Ciari *et al.* (2007) and Balmer *et al.* (2008) respectively. Thus, the current description of individuals in MATSim-T is already rich, but does not take into account their consumer aspect, and it will need to be modified. First, agents will have an income and, consequently, a monthly budget for retail expenditures. Their knowledge will be also enhanced accounting for the price level of the stores and for the presence of parking facilities. The objective function will be also modified, monetary cost explicitly considered, and agents will seek for the maximum satisfaction within the budget constraint. Furthermore, the specification of activities will be also refined and, in particular, shopping activities differentiated according to predefined retail sectors.

Figure 8 Current and future representation of individual agents



6.7.3 Retailers

The scope of the model is to reproduce the location behaviour of retailers. In the model system, activities may be performed at different locations called facilities. Each facility is an entity with following attributes: type, location, capacity, opening time, closing time. In a single facility one or more activities of different types can be performed (leisure, shopping, work). Each activity type of a facility contains a capacity - which defines the maximum number of agents which are permitted to perform a given activity in this facility at the same time. The focus here lies on *shop* facilities, interpreted as a retail store. Therefore, the retailer agent is represented as the decision maker having the control on a certain number of shop facilities. The retailer agent does not necessarily represent an individual (i.e. the owner of the shop) but for example it might also be the board of a retail chain. This entity will be provided with attributes, knowledge, one or multiple objectives, a strategy to pursue it, a methodology to implement this strategy and a group of allowed choices. Therefore, the retail agent is consistent

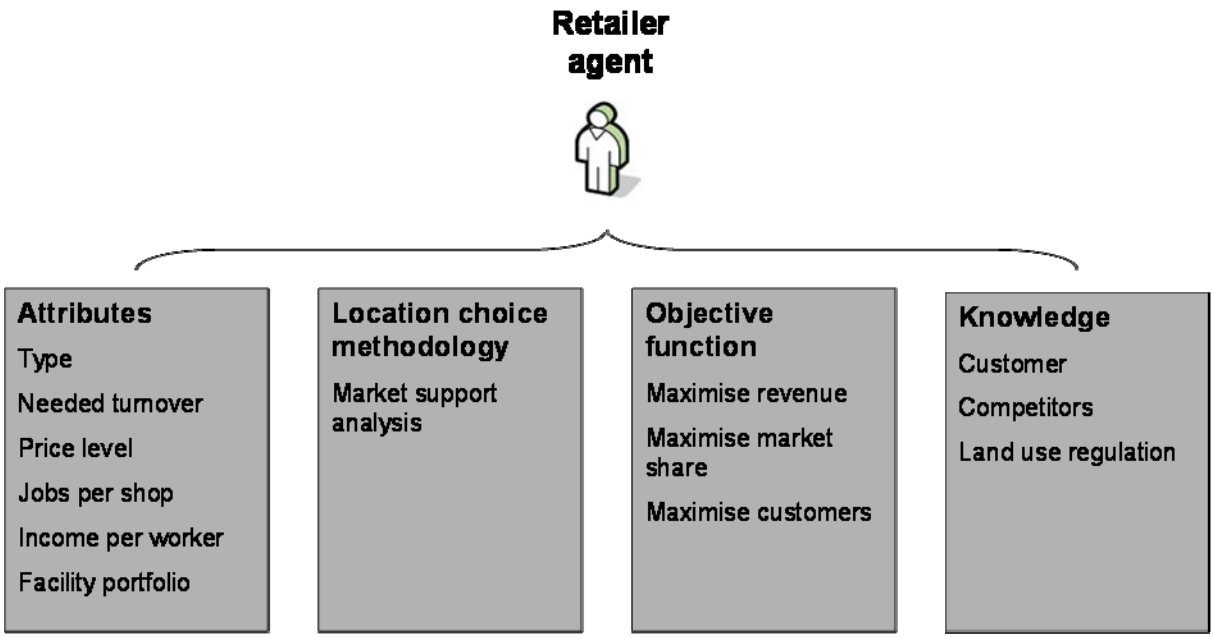
with most of definitions of agents in the artificial intelligence literature, for example the one given by Ferber (1999), where an agent is defined as a

> "real or virtual entity which is capable of acting in an environment, which can communicate directly with other agents, which is driven by a set of tendencies (in the form of individual objectives or of a satisfaction/survival function which it tries to optimize), which possesses resources of its own, which is capable of perceiving its environment (but to a limited extent), which has only a partial representation of its environment (and perhaps none at all), which possesses skills and can offer services, which may be possibly be able to reproduce itself, whose behaviour tends towards satisfying its objectives, taking account of the resources and skills available to it and depending on its perception, its representation and the communications it receives."

In the model system, all types of shops are undifferentiated. In the new representation they will be described with attributes like: type, amount of needed daily/weekly turnover per square meter, price level, jobs per shop/square meter, income per worker, facility portfolio. Each of the facilities controlled by one retailer will have some attributes in turn, such as: type, capacity, location, etc. The knowledge of a retail agent will be in principle of two types; knowledge of customers and knowledge of competitors. The model considers that the first is imperfect, the knowledge about customers will be limited to the number of individuals living in a determined area or passing through a given link of the network. However, as in Timmermans and Arentze (2007), the retail agent knows how many customers have shopped in one of his stores upon certain assumptions after each iteration. The retailer will be able to observe the competitors, their locations and price levels within a predefined neighbourhood. The choice of a new location will be effectuated taking into account such information, but it will be considered that choices of retailers will be simultaneous. Another information which will be added as retailer's knowledge is the land use regulation. Only land suitable for commercial use will be allowed as a new location. The objective for a retailer agent will be the maximisation of revenues, but we could also test other possibilities like the maximisation of the market share or simply the number of customers. The pattern the retailer intends to follow in order to meet his own objective in the long run constitution of the strategic level. Since the focus lies on the location choice, the strategic level coincides with the territorial expansion planned by the retailer. This has a meaning at the regional or at the national scale, while simulations will be run at the metropolitan level in this case, at least at an initial stage. Moreover, the interviews with German and Swiss retailers revealed that neither contagion nor hierarchic expansion patterns are particularly used in practice. Therefore, this level will be neglected at first. More generally, the strategic level could also be extended to other policies such as price or advertising level. The methodological level is the way the retailer will effectuate the choice,

in this case the location choice. The market support analysis will be the technique used by retailer agents. This technique is simple but still applied in practice by retailers (Birkin *et al.*, 2002; Rogers, 2007). Retailer's choices at a first stage will be limited to the location, but other choices will be allowed in the future work, such as changing the price level or opening times.

Figure 9 Retailer agent

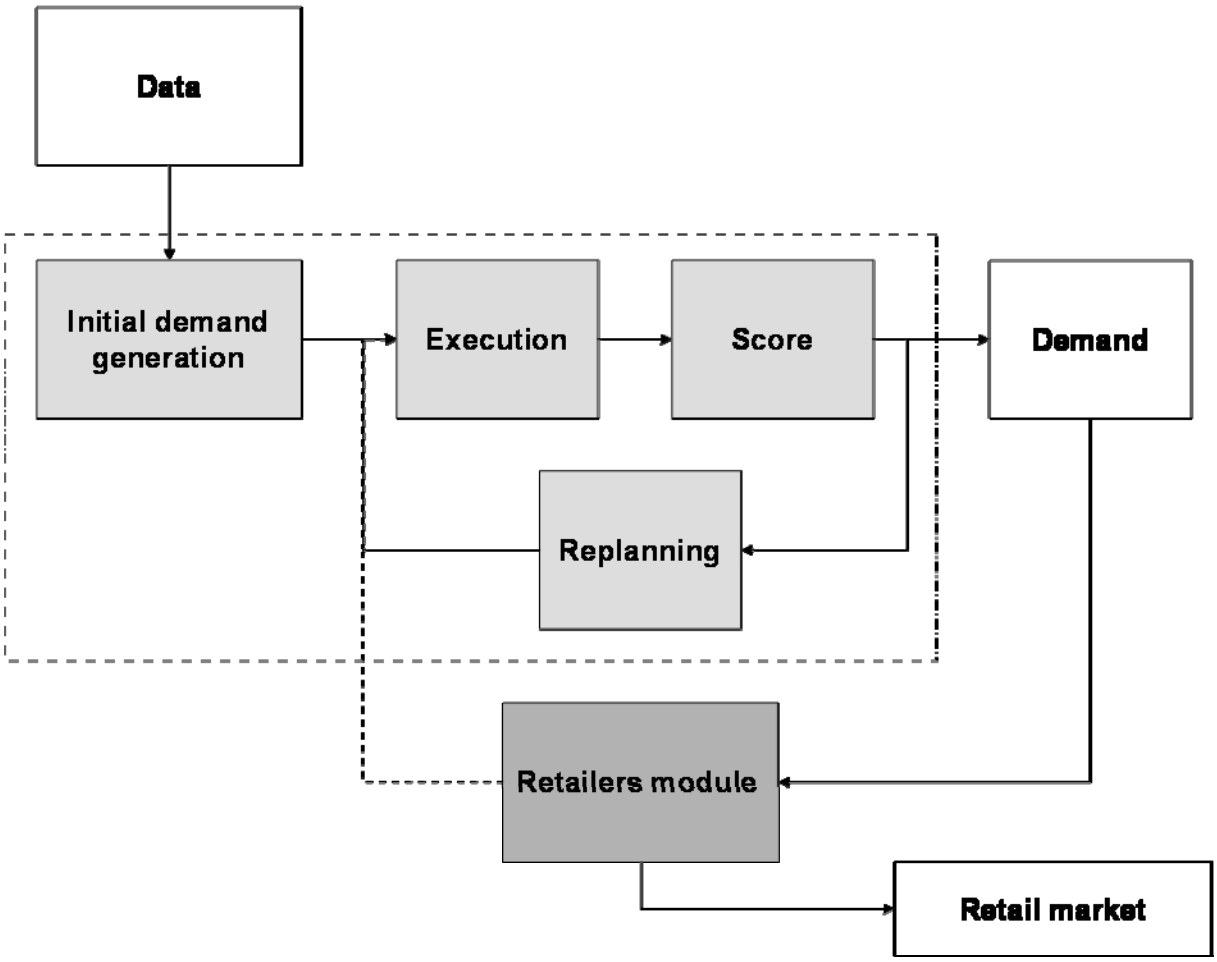


6.7.4 The new retailer module in MATSim-T

The retail market model is similar to the model defined by Timmermans and Arentze (2007), not only because the model is structurally similar, but also because it is conceived to seek a stable equilibrium in the facility location/facility usage interaction, while other integrated land use and transport models are seeking a time path. The optimisation of the supply/demand side will be separated, at least in the first stage. The retailer module will be integrated in MATSim-T outside the main loop, in which the demand is optimised. More precisely, we estimate the initial demand for a given scenario and then we feed the optimisation tool with this demand. The optimised demand is put in the retail module as input and the retailer agent will try to improve its location choice. This produces a new scenario, which is given back to the optimisation tool in a loop that stops when either a relaxed state or a fixed number of iterations is reached (see Figure 10).

Note that to implement the described cycle one does not need to overcome any particular difficulty related with the input/output formats, since both input and output are in the form of an xml file. In fact, this characteristic of MATSim-T makes it particularly easy to introduce new modules from a technical point of view.

Figure 10 Loop for the retailer module



Based on Balmer *et al.* (2008)

6.7.5 Conclusions

The model sketched provides a rich description of retailers and updates the modelling of individual agents taking into account their customer aspect. Results of the simulation are expected to reflect an equilibrium between retailers and individual agents' choices. With the implementation of this model, the interactions between the supply and demand side in the retail location

should emerge from the simulation. This aspect is of particular importance, the agent-based approach gives us the possibility to describe the agent of the system at the microscopic level and to detect behaviours at the macroscopic level which would be not expected considering single agents' actions. It is a powerful instrument in the hands of the modeller, but needs also to be used with caution. The construction of a comprehensive model of this type to describe real-life events is only one part of the work, although an important one. Other experiences, like the one of ILUMASS (Wagner and Wegener, 2007), tell that complex models are not only hard to implement, but also that results are completely reliable only if the model is accurately tested at each stage of its development. This is planned for the next stage of work, where the functionality of the model will be tested implementing a really simple version of it, and running it on a small test scenario. This approach ensures that the system reacts as expected at the small scale and that there is a reasonable calibrated model for applications at a larger scale.

6.8 Summary and Outlook

This chapter reports on an ongoing project aimed to introduce a model of retail markets in the agent-based traffic simulation toolkit MATSim-T. It will be done by introducing a new retailer module in the toolkit, representing retailers as agents. The new module's main goal is to reproduce the location choices of retailers and to recognise how such choices and shopping location choices of individuals are mutually interdependent. A review of the literature on retailers' location choice was presented and it is part of the theoretical background on which the new retail agent is constructed. Internationally, retailers have increasingly used more sophisticated computer-based location techniques over the last two decades. But older and simpler methodologies are far to be fallen into disuse at the same time. The literature review was enriched with the results of a series of in-depth interviews, which were conducted with retailers of various sizes and product groups in Germany and Switzerland in the winter of 2008. The results reveal that there is a huge variety of location strategies both between and within the different retail sectors. Moreover, location choices are heavily based on experience and intuition, particularly those decisions at the micro scale.

A model for retailers was sketched and the main issues integrating it in MATSim were discussed. The next stage of this work will focus on the implementation of the model and on its testing. First, an implementation will be run on a simple test scenario. Once the functionality of the model is demonstrated we will step up to (more sophisticated and) more realistic as well as larger scenarios, like the Zürich area. In later stages of the project, it is planned to run

simulation tests with different possible implementations of a road pricing scheme and to evaluate the results based on the hypothesis given in this chapter.

7 Synthesis

The applied techniques in this dissertation proved to be helpful as tools for analysing geographic variation and informing policy decisions and investment allocation. Conclusions were drawn individually at the end of Chapters 3 to 6. However, the chosen approaches are only partly evaluated there. In the following, this is broadened and combined with the discussion of alternative methods and potential further work.

The chosen modelling approach presented in Chapter 3 is considerably data driven. Although the available data form a valuable and unique base for the analysis, it would have been desirable to have more information about marketing measures such as channel budgeting or geographical differences in pricing and advertisement, etc. as well as more complete information about the competitor's agents. The availability of such data might improve the model fit and confirm the estimation results, which indicate that the agent density is not as important as one might expect. Moreover, it would have been interesting to be able to analyse if high distribution intensity impacts total insurance demand positively. But since the demand data at the municipality level is based on regional figures disaggregated by household age, income and size, absolute demand per municipality is an improper variable to investigate in this case.

A distribution intensity measure of the agents is suggested, which applies a uniform methodology across the country. It focuses on the location of the demand and assumes that each agent knows his local market perfectly within 30 minutes travel time and does spatially allocate his efforts to the demand accordingly. There must be considerable differences in the actual distribution activity patterns as municipalities with high agent density are adjacent to municipalities with low density as observable in Figure 3. Remarkable is the situation in the Chur region for example where almost all municipalities within 30 minutes travel time from Chur have high agent densities, while most other municipalities in Northern Grisons have low densities. Apparently, several agents choose Chur as their work location, but care for markets beyond the Chur area in Northern Grisons and have therefore an unusual large activity radius. Nevertheless, no nationwide uniform approach was found which explains these variations more appropriately. Moreover, the suggested agent density measure does neither consider the location of agents of the same company nor the location of the agents of competitors when the activity (distribution) area and strength of an individual agent is spatially associated to municipalities. In practice, both probably affect radius and shape of the activity space significantly, which could be investigated in further work.

An alternative to the chosen regression approach might be data envelopment analysis (DEA; Banker *et al.*, 1984; Charnes *et al.*, 1978), which is essentially a nonparametric linear programming approach to efficiency analysis. It has been widely applied predominantly in management sciences. In a comparison of DEA and regression analysis, Thanassoulis (1993) concludes that the DEA shows better accuracy of estimates but regression analysis offers greater stability of accuracy in his application. In general, the results might differ (Ferrier and Lovell, 1990). Moreover, DEA is described as good at identifying possible reasons for apparently poor performance (Cubbin and Tzanidakis, 1998, 84). On the other hand, hypothesis testing is more problematic with DEA (Cubbin and Tzanidakis, 1998, 80). Therefore, the statistical significance of competing explanatory variables as well as the appropriateness of the estimated functional form can be tested in regression analysis, but not in DEA. Prior to execution of DEA, it has to be decided about the relative importance of competing explanatory factors and even their sign. But no specification of the function is required and therefore, it avoids the danger of fitting the wrong functional form (Cubbin and Tzanidakis, 1998, 81). However, DEA does not consider spatial effects and therefore the results are biased. A possible solution might be the combination of DEA and the spatial regression approach. Felder and Tauchmann (2009) suggest a two-stage approach. In a first stage, they calculate efficiency scores of regional health care provision by the DEA approach and use the resulting scores afterwards in a spatial autoregressive model with autoregressive disturbances (Kelejian and Prucha, 1998) as the dependent variable. Although this seems to be a promising approach, it involves technical inconistencies since the variables in the second stage were obviously expected to affect the performance but were not considered in the first stage. Moreover, if the variables used in specifying the DEA are correlated with explanatory variables in the spatial regression, the regression results will be inconsistent and biased (see also Grosskopf, 1990, 165).

The hedonic estimation presented in Chapter 4 is proposed for the next UrbanSim application for Zürich. However, as long as UrbanSim does not represent structural variables of the building stock, the hedonic model has to rely on locational explanatory variables only and consequently on a reduced model with less detailed information and lower model fit.

As mentioned briefly in the introduction of Chapter 4, there was a household survey conducted in the project. The addresses were geocoded and matched with various locational variables. However, the hedonic modelling approach based on the referring rent and locational variables was not as successful as expected, indicated by a low model fit. The reason was probably the spatially clustered sampling of the households, as the survey was conducted in twenty municipalities in the Greater Zürich area and four districts of the City of Zürich. Apparently, there was not enough variance in the data to get sufficiently significant estimates.

Therefore, an important conclusion with regard to hedonic estimations based on solely locational explanatory variables is that it is only applicable in larger study areas with sufficient variation of the various explanatory factors.

With regard to the progress in hedonic modelling, a large amount of alternative approaches and procedures were suggested in the literature over the last years and new suggestions continue to appear. Many new model specifications have been considered, different test statistics proposed as well as novel estimation methods developed and their computational aspects assessed (Anselin *et al.,* 2004, 7). New developments in the spatial regression family include, for example, Bayesian approaches (LeSage, 1997), the generalised method of moments (Kelejian and Prucha, 1998 and 1999) and generalised maximum entropy approaches (Golan, Judge and Miller, 1996 and 1997). Moreover, new approaches and guidelines for constructing the spatial weights are proposed (Bavaud, 1998; Stakhovych and Bijmolt, 2008; Tiefelsdorf *et al.*, 1999). Dubin (2004, 97) suggests to combine the geostatistical and the weight matrix approach for considering spatial effects which better detect and relate problems of spatial lag and spatial error. Further alternative methods to model spatially autocorrelated data include non- and semiparametric methods (Pace, 1993 and 1995) as well as Bayesian hierarchical models (Banerjee *et al.*, 2004). Each of these approaches has its strengths and weaknesses and many of the suggestions remain to be further tested in various settings and circumstances. Since a comprehensive assessment of all these methods is beyond the scope of this dissertation, focus is given to one alternative approach.

Hierarchical linear modelling, sometimes also called multi-level modelling (Goldstein, 2003; Raudenbush and Bryk, 2002), might be an alternative to the spatial regression models presented in Chapter 4. Essentially, hierachical linear models recognise hierachical clusters where units of analysis (level 2) are nested within higher-level aggregate units (level 1). A typical example is that of students nested within schools. A hierachical model clearly identifies and differentiates between cluster heterogeneity and heterogeneity between units of analysis that are nested within aggregated clusters. Although the technique is popular in the social sciences (Raudenbush and Bryk, 2002, xxiii), applications in the field of real estate and housing are scarce (Habib and Miller, 2008, 184). However, exceptions are Brown and Uyar (2004), Gelfand *et al.* (2007), Goodman and Thibodeau (2003), Jones and Bullen (1994) and Orford (2000). Goodman and Thibodeau (1998) use hierarchical models to delineate housing submarket boundaries. Djurdjevic and Eugster (2008) apply the hierarchical approach to a residential rent data set from whole of Switzerland whereas level 1 corresponds to the dwelling unit and level 2 to the municipality. They find that the multilevel hedonic model provides more accurate predictions than the traditional OLS approach. Even after controlling for hous-

ing and locational attributes, a significant part of rental differences are still attributable to municipality differences in their data set.

The hierarchical linear model approach was not applied in the hedonic application in this dissertation due to several reasons. First, hierarchical linear models consider the context of each municipality, but they generally fail to measure the influence of this context as proximity or neighbourhood is not considered. Therefore, the method might be useful to improve hedonic prediction accuracy and a subset of parameters, but it does not improve the understanding of the observed variation of housing prices comprehensively. Additionally, spatial correlations are not considered in hierarchical linear models. Different than the nation-wide Swiss application by Djurdjevic and Eugster (2008), the study area is Canton Zürich with the City of Zürich as the economic and geographical center. Therefore, the municipalities are certainly more homogeneous. Finally, the models in Chapter 4.5 include tax level and car and public transport accessibility at municipality level[18], consequently considering submarkets to some extent.

The suggested hedonic model in Chapter 4.5 promises considerable improvement in the application of the next generation of the application of UrbanSim for the Greater Zürich area. In using the newly created explanatory variables and estimated rents, it might be possible to improve the discrete choice estimations presented in Chapter 5 as well. In particular, it might lower the negative impact of the rent price on residential location choice, which proved to be in particular problematic.

One lesson learned from the UrbanSim application experience, not mentioned in Chapter 5, is that modellers should start as early as possible to actually apply the UrbanSim program before the best available data can be used. This assures early familiarity with the program and prevents problems at the end of a project, when simulation results are not as expected as error search might be long. In the presented application, however, the early deployment of UrbanSim in the project was suspended because of an anticipated major version update of UrbanSim.

A general critique of integrated land use and transport simulations is that often the notion of integration is reduced to the principle that the calculated accessibility or travel time measures serve as one of the explanatory variables of the residential choice module (Timmermans, 2006, 237). Timmermans (2006, 238) describes the literature on residential location choice

[18] For the accessibility variables, the referring traffic zones are only in some parts of the study area at the municipality level.

behaviour has systematically shown that accessibility plays a marginal role in the residential choice decision. According to him, structural attributes of the house and physcial and social characteristics of the neighbourhood are more important. However, the hedonic estimations in Chapter 4 indicated that accessibility and travel time is by far the most important variable at least for explaining rents in Canton Zürich. A region specific reason might be the superior economic and social attraction of the City of Zürich in the geographic center of the canton. If prices are interpreted as willingness to pay or demand, one can state that accessibility and travel time indeed play a crucial role in the study area with regard to location choice. In this context, it would be interesting to update the residential choice model presented in Chapter 5 with the explanatory variables from Chapter 4 in order to be able to compare the parameters of the choice model with those of the hedonic rent model.

Another stream of work might be the accessibility measures, which are currently based on traffic zones and are therefore spatially aggregated. Consequently, an improvement would be a spatial disaggregation for example by considering distance from each public transport stop within the traffic zones. For the better integration of land use and transport interaction in general and the residential location choice model in particular, it would be beneficial to consider measures such as traffic volumes or traffic nuisances next to the residential buildings, in case the data are available for all locations[19]. Potentially, this can be achieved by using a very detailed street network in the transport model, providing traffic volume for any part of the network. In case the transport model is run after every simulation year with updated land use data, a more realistic setting of land use and transport might be achieved through the closer coupling of the two components.

In general, the discussion with regard to integrated land use and transport simulations centers around two issues. One is that of necessary and sufficient detail and complexity also involving methodological and technical advancement. This is predominantly discussed in the academic world. In the last years, there has been a trend towards the disaggregation of the representation of space, people and other entities of analysis (see also Iacono *et al.*, 2008, 337), although it is realised that there are theoretical, empirical, practical and ethical limits of disaggregation (Wagner and Wegener, 2007, 54). However, the question of how much detail is necessary to get sufficient simulation results is not adequately answered by the research community yet. The other issue is partly related to the aforementioned one and focuses on usability, policy sensitivity and therefore practical relevance of the the models. There is a debate

[19] The data were unavailable in the case for the models presented in Chapter 4 and Chapter 5. Tests with proximity measures by street category did not provide any significant results.

ongoing as to why spatial decision support systems, such as integrated land use and transport models, are hardly used in daily planning practice. Studies found that the gap between the modelling and planning communities is too large, since the spatial decision support systems are technology-oriented rather than planning-oriented, and these instruments do not fit the complex dynamics of real-world planning contexts (Te Brömmelstroet, 2009, 4). Moreover, the required resources and know-how to set up an integrated land use and transport model are substantial. There is no easy way around these obstacles, but any integrated model needs to be as accessible and easy as possible without loosing its capabilities to appropriately simulate the most important processes in spatial development (see also Timmermans, 2006, 239). Additionally, there is always a need for an assessment of required complexity and efforts for the implementation depending on the actual purpose as well as the necessary sensitivity and functionality. As noted in Chapter 5, data availability, collection and preparation forms a major challenge in the application of integrated land use and transport models. This will not change in the next years. In Switzerland for example, data availability will decrease in the near future as the comprehensive population census is reduced to an expansion based on a sub-sample, enhanced by information from local statistical offices.

There are voices which state that it does not seem realistic to expect that any integrated model of land use and transport could provide accurate land use forecasts at the level of individual cells but rather qualitative indications for wider areas (Timmermans, 2006, 239). Although the dissertation does not provide any results to argue against or in favor of this statement, it seems reasonable to assume that there are indeed limits of projection detail. Nevertheless, integrated land use and transport models can help to test theories and to develop knowledge about the behaviour of urban systems such as the analysis of certain interrelations. Moreover, they can be a useful tool for scenario testing and discussion, considering the impact of planned infrastructure measures or policies in order to draw informed location decisions. Finally they can be a helpful tool in location planning and resource deployment decision making.

The agent-based modelling approach suggested in Chapter 6 can not be judged at this point as it is a proposal and its execution is ongoing. However, the conducted interviews revealed that different than in the UK and the US, advanced spatial analysis and modelling techniques are rarely applied in retail location planning practice in Germany and in particular in Switzerland, where local decisions are predominantly based on intuition in the retail industry and involve less sophisticated methods. Nevertheless, the applicability of the agent-based modelling approach in a location planning context is manifold. It allows to identify optimal locations for the retailers under certain circumstances as well as to study public policies such as road pricing and its impacts on business location strategies.

The examples in this dissertation revealed that there are various spatial allocation decisions, which have to be solved by private actors. In Chapter 3, a market share model is developed which can be helpful in spatial resource allocation for example with regard to sales force, advertising and sales promotion. As Shankar (2008, 3) notes, there is a dearth of tools for such resource allocation decisions and he estimates that managers will be more likely to use simulators and decision support systems for resource allocation in the future than they did in the past (Shankar, 2008, 20). The hedonic models presented in Chapter 4 are not only useable in the next application of UrbanSim in the Greater Zürich area, but potentially offers also valuable information for tenants, home owners, consultants and banks. Integrated land use and transport simulations, such as the one presented in Chapter 5, are not only useful for analysing the land use and transport interaction and testing urban policies and investments for planners and decision makers in public planning authorities, but can also be a spatial decision support tool for large scale real estate developers and investors, such as banks, insurances and pension funds, which are interested to forecast the development of neighbourhoods or the spatial long-term effects of their investments. Nevertheless, the potential of these methods and tools are often not identified yet by these groups. There have to be some well established application cases, for which the usefulness and cost-effectiveness have to be proved and communicated well to spark interest. A key for success is certainly moderate complexity of methods and simulations which can be explained to and understood by decision makers with business backgrounds in the companies. The development and calibration of simulations can necessitate large resources which might not be worth for a single analysis. But once calibrated models and skilled people are available, the additional costs for a variety of further applications are small, which opens business opportunities for consultants who serve aforementioned potential customers.

Overall, the analysis of geographical variation of entities is closely related to local statistical methods. These methods were called a useful addition to the repertoire of spatial analysis and some of them were applied in this dissertation. However, several issues are not solved at this point and need further research. The application of GWR techniques in Chapter 4 clearly revealed a problem with local correlation (see also Wheeler and Tiefelsdorf, 2005). This must either be controlled or explicitly incorporated in the local model interpretations. Therefore, a clear identification of the unique local effect is required (Berry *et al.*, 2008, 235f). As described in Chapter 4, there are already some methodological suggestions, but research is ongoing (i.e. Waller *et al.*, 2007; Wheeler and Waller, 2009). As another research issue, Berry *et al.* (2008, 236) name the adoption of a joint statistical framework for testing local parameters to enable the significance of local patterns and the statistical power of the local estimates to be assessed. They see the development of a unified model structure as the ultimate objective,

which allows estimating and testing all local and global effects simultaneously. Beyond, Anselin (2000) notes that another challenge to spatial analysis is not only to develop new techniques of 'local' spatial analysis or more sophisticated models that formally express spatial effects, but also to provide the means to discover and understand the underlying social and behavioral mechanisms that yield the revealed spatial patterns.

Finally, it has been denoted that potential applications of spatial analysis methods are manifold. This is also true for the analysis of geographic variation of entities. Therefore, the applications in this dissertation can only give examples. Spatial analysis continues to develop and because of the interdisciplinary contributions to spatial analysis from various fields, there are already voices which see spatial analysis transferring to a geospatial science discipline of its own (Berry *et al.*, 2008, 236). Dekker and Rietveld use *spatial economics* as an umbrella term for urban and regional economics, transport economics, environmental and resource economics and spatial econometrics (Dekkers and Rietveld, 2009, 137). At the same time, interdisciplinary journals (i.e. Applied Spatial Analysis and Policy) as well as specialised conferences (i.e. World Conference of Spatial Econometrics) and associations (i.e. Spatial Econometrics Association) are appearing. There is no doubt that spatial analysis approaches will further diffuse into the various disciplines and establish new sub-disciplines as well as interdisciplinary approaches as there is still growing demand for spatial analysis in many disciplines.

8 Literature

Abraham, J. E., T. Weidner, J. Gliebe, C. Willison and J. D. Hunt (2005) Three methods for synthesizing base-year built form for integrated land use–transport models, *Transportation Research Record*, **1902**, 114-123.

Accenture & Universität St. Gallen (2005) *Assekuranz 2015 – Retailmärkte im Umbruch*, Zürich/St. Gallen, http://www.ivw.unisg.ch/org/ivw/web.nsf/SysWebRessources/Asse2015-Quovadis+pdf/$FILE/quovadis.pdf, retrieved 12th August 2008.

Agrawal, A. (2006) Engaging the inventor: exploring licensing strategies for university interventions and the role of latent knowledge, *Strategic Management Journal*, **27** (1) 63-79.

Ahearne, M., R. Jelinek and E. Jones (2007) Examining the effect of salesperson service behavior in a competitive context, *Journal of the Academy of Marketing Science*, **35** (4) 603-616.

Anglin, P.M. and R. Gençay (1996) Semiparametric estimation of a hedonic price function, *Journal of Applied Econometrics*, **11** (6) 633-648.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

Anselin, L. (1995) Local indicators of spatial association –LISA, *Geographical Analysis*, **27** (2) 93-115.

Anselin, L. (2000) The future of spatial analysis in the social sciences, *Geographic Information Sciences*, **5** (2) 67-76.

Anselin, L. (2001a) Spatial econometrics, in B.H. Baltagi (ed.) *A Companion to Theoretical Econometrics*, 310-330, Blackwell, Malden.

Anselin, L. (2001b) Spatial effects in econometric practice in environmental and resource economics, *American Journal of Agricultural Economics*, **83** (3) 705-710.

Anselin, L. (2002) Under the hood: Issues in the specification and interpretation of spatial regression models, *Agricultural Economics*, **27** (3) 247-267.

Anselin, L., A.K. Bera, R. Florax and M.J. Yoon (1996) Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics*, **26** (1) 77-104.

Anselin, L., R.J.G.M. Florax and S.J. Rey (2004) Econometrics for spatial models: Recent advances, in L. Anselin, R.J.G.M Florax and S.J. Rey (eds.) *Advances in Spatial Econometrics: Methodology, Tools and Applications*, 1-25, Springer, Berlin.

Arbia, G., D. Griffith and R. Haining (1998) Error propagation modelling in raster GIS: Overlay operations, *International Journal of Geographical Science*, **12** (2) 145-167.

ARE and BfS (2001) Mobilität in der Schweiz, Ergebnisse des Mikrozensus 2000 zum Verkehrsverhalten, Research Report, 6/91, Swiss Federal Office for Spatial Development and Swiss Federal Statistical Office, Bern.

Arentze, T., F. Hofman, H. van Mourik, and H. Timmermans (2000) ALBATROSS: A multiagent rule-based model of activity pattern decisions, *Transportation Research Record*, **1706**, 136-144.

Arentze, T.A. and H.J.P. Timmermans (2004) A micro-simulator of urban land use dynamics integrating a multi-agent model of land development and an activity-based model of transport demand, proceedings of the *83rd Annual Meeting of the Transportation Research Board*, Washington DC, January 2004.

Arentze, T.A. and H.J.P. Timmermans (2005) Impact of institutional change on shopping patterns: Application of a multi-agent model of activity-travel behaviour, paper presented at the *12th Recent Advances in Retailing and Services Science Conference*, Orlando, July 2005.

Arentze, T.A. and H.J.P. Timmermans (2007) A multi-agent activity-based model of facility location choice and use, *DISP*, **170** (3) 33-44.

Armstrong, A.T. (1986) Road pricing and user restraint: opportunities and constraints in developing countries, *Transportation Research A*, **20A** (2) 123-127.

Artikis, P.G., S. Mutenga and S.K. Staikouras (2008) A practical approach to blend insurance in the banking network, *The Journal of Risk Finance*, **9** (2) 106-124.

Ashby, D.I. and P.A. Longley (2005) Geocomputation, geodemographics and resource allocation for local policing, *Transactions in GIS*, **9** (1) 53-72.

Ataman, M.B. C.F. Mela and H.J. Van Heerde (2007) Consumer packaged good in France: National brands, regional chains, and local branding, *Journal of Marketing Research*, **44** (1) 14-20.

Atkinson, P.M. and N.J. Tate (2000) Spatial scale problems and geostatistical solutions: A review, *Professional Geographer*, **52** (4) 607-623.

Avant Management A/S (1992) *Handelsundersøkelsen 1992, Trondheim med omkringliggende kommuner*. [The trade inquiry of 1992, Trondheim with surrounding municipalities.]

Avery, J., T.J. Steenburgh, J. Deighton and M. Caravella (2009) Adding bricks to clicks: The contingencies driving cannibalization and complementarity in multichannel retailing, *Harvard Business School Working Paper*, **07-043**, Harvard Business School, Boston.

Avkiran, N.K. (1997) Models of retail performance for bank branches: Predicting the level of key business drivers, *International Journal of Bank Marketing,* **15** (6) 224-237.

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

Axelrod, R. (1997b) Advancing the art of simulation in the social sciences, in R. Conte, R. Hegselmann and P. Terna (eds.) *Simulating Social Phenomena*, 21-40, Springer, Berlin.

Axhausen, K.W. and T. Gärling (1992) Activity based approaches to travel analysis: Conceptual frameworks, models and research problems, *Transport Reviews*, **12** (4) 323-341.

Balmer, M. (2007) Travel Demand Modeling for Multi-Agent Traffic Simulations: Algorithms and Systems, Ph.D. Thesis, ETH Zürich, Zürich, May 2007.

Balmer, M., K.W. Axhausen and K. Nagel (2006) An agent-based demand modelling framework for large scale micro-simulations, *Transportation Research Record,* **1985**, 125-134.

Balmer, M., M. Rieser, K. Meister, D. Charypar, N. Lefebvre, K. Nagel and K. W. Axhausen (2008) MATSim-T: Architektur und Rechenzeiten, paper presented at the *Heureka '08*, Stuttgart, March 2008.

Balmer, M., M. Rieser, K. Meister, D. Charypar, N. Lefebvre and K. Nagel (2009) MATSim-T: Architecture and simulation times, in: A.L.C. Bazzan and F. Klügl (eds.) *Multi-Agent Systems for Traffic and Transportation Engineering*, 57-78, Information Science Reference, Hershey.

Banerjee, S., B.P. Carlin and A.E. Gelfand (2004) *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall/CRC, New York.

Banfi, S., A. Horehejaova and M. Filippini (2006) Hedonic price functions for Zürich and Lugano with special focus on electrosmog, paper presented at the *3rd World Congress of Environmental and Resource Economists*, Kyoto, Japan, June 2006.

Banker, R.D., A. Charnes and W.W. Cooper (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, **30** (9) 1078-1092.

Bannwarth, R. (2008) Internationale Trends im Multi-Channel-Mangement – Strategische Optionen und Schlüsselfaktoren, *IVW Management Information*, **30** (2) 7-11.

Baranzini, A. and J.V. Ramirez (2005) Paying for quietness: The impact of noise on Geneva rents, *Urban Studies*, **42** (4) 633-646.

Baranzini, A. and C. Schaerer (2007) *A sight for sore eyes: Assessing the value of view and landscape use on the housing market*, Geneva School of Business Administration, University of Applied Sciences Western Switzerland, Geneva.

Baranzini, A., C. Schaerer, J.V. Ramirez and P. Thalmann (2006*) Feel it or measure it. Perceived vs. measured noise in hedonic models*, Geneva School of Business Administration, University of Applied Sciences Western Switzerland, Geneva.

Baranzini, A., J.V. Ramirez, C. Schaerer and P. Thalmann (2008a) Introduction to this volume: Applying hedonics in the Swiss Housing Markets, *Swiss Journal of Economics and Statistics*, **144** (4) 543-559.

Baranzini, A., J.V. Ramirez, C. Schaerer and P. Thalmann (2008b) Introduction, in A. Baranzini, J.V. Ramirez, C. Schaerer and P. Thalmann (eds.) *Hedonic Methods in Housing Markets*, 1-12, Springer, New York.

Baranzini, A., C. Schaerer, J.V. Ramirez and P. Thalmann (2008c) Do foreigners pay higher rents for the same quality of housing in Geneva and Zurich?, *Swiss Journal of Economics and Statistics*, **144** (4) 703-730.

Basu, S. and T.G. Thibodeau (1998) Analysis of spatial autocorrelation in house prices, *Journal of Real Estate Finance and Economics*, **17** (1) 61-85.

Bateman, I.J. (1994) The contingent valuation and hedonic pricing methods: Problems and possibilities, *Landscape Research*, **19** (1) 30-32.

Baumann, J. (2007) Versicherer rüsten sich für den Preiskampf, *Finanz und Wirtschaft*, **60**, 17.

Bavaud, F. (1998) Models for spatial weight: A systematic look, *Geographical Analysis*, **30** (2) 153-171.

Beaumont, J.R. and K. Inglis (1989) Geodemographics in practice: Developments in Britain and Europe, *Environment and Planning A*, **21** (5) 587-604.

Beck, T. and I. Webb (2003) Economic, demographic, and institutional determinants of life insurance consumption across countries, *The World Economic Review*, **17** (1) 51-88.

Beckmann, K.J. (2006) Mikro-Simulation von Raum- und Verkehrsentwicklung – Stand der Kunst und Perspektiven zwischen Forschung, Entwicklung und Praxis, *Stadt Region Land*, 81, 7-31.

Benenson, I. (1998) Multi-agent simulations of residential dynamics in the city, *Computers, Environment and Urban Systems*, **22** (1) 25–42.

Berger, A.N. and A.A. Dick (2007) Entry into markets and early-mover advantage, *Journal of Money, Credit and Banking*, **39** (4) 775-807.

Berry, B.J.L., D.A. Griffith and M.R. Tiefelsdorf (2008) From spatial analysis to geospatial science, *Geographical Analysis*, **40** (3) 229-238.

Bhat, C.R., J.Y. Guo, S. Srinivasan and A. Sivakumar (2005) A comprehensive econometric microsimulator for daily activity-travel patterns, *Transportation Research Record*, **1894**, 57-66.

Bierlaire, M. (2006) BIOGEME 1.4 Homepage, http://roso.epfl.ch/biogeme, EPFL Lausanne, Lausanne, October 2006.

Birkin, M. and M. Clarke (1988) SYNTHESIS - A synthetic spatial information system for urban and regional analysis: methods and examples, *Environment and Planning A*, **20** (12) 1645 – 1671.

Birkin, M., G. Clarke and M. Clarke (2002) *Retail Geography and Intelligent Network Planning*, John Wiley, Chichester.

Bitter, C., G.F. Mulligan and S. Dall'erba (2007) Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method, *Journal of Geographical Systems*, **9** (1) 7-27.

Bivand, R. (2008) Implementing representations of space in economic geography, *Journal of Regional Science*, **48** (1) 1-27.

Bivand, R.S., E.J. Pebesma and V. Gmez-Rubio (2008) *Applied Spatial Data Analysis with R*, Springer, New York.

Biyalogorsky, E. and P. Naik (2003) Clicks and mortar: the effect of online activities on offline sales, *Marketing Letters*, **14** (1) 21-32.

Bodenmann, B. (2006) Lebenszyklusmodelle für Unternehmen in der Raumplanung, *Arbeitsberichte Verkehrs- und Raumplanung*, **393**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Boots, B. and A. Okabe (2007) Local statistical analysis: Inventory and prospect, *International Journal of Geographical Information Science*, **21** (4) 355-375.

Borning, A., P. Waddell and R. Förster (2008) UrbanSim: Using simulation to inform public deliberation and decision-making, in H. Chen, L. Brandt, V. Gregg, R. Traunmüller, S. Dawes, E. Hovy, A. Macintosh and C.A. Larson (eds.) *Digital Government: E-Government Research, Case Studies, and Implementation*, 439-466, Springer, New York.

Boufounou, P.V. (1995) Evaluating bank branch location and performance: A case study, *European Journal of Operational Research*, **87** (2) 389-402.

Bourassa, S.C., M. Hoesli and V.S. Peng (2003) Do housing submarkets really matter?, *Journal of Housing Economics*, **12** (1) 12-28.

Bourassa, S.C. and M. Hoesli (2010) Why do the Swiss rent?, *The Journal of Real Estate Finance and Economics*, **40** (3) 286-309.

Bowen, W.M., B.A. Mikelbank and D.M. Prestegaard (2001) Theoretical and empirical considerations regarding space in hedonic housing price model applications, *Growth and Change*, **32** (4) 466-490.

Bowman, J.L., M. Bradley, Y. Shiftan, T.K. Lawton and M. Ben-Akiva (1999) Demonstration of an activity-based model for Portland, in H. Meersman, E. Van De Voorde and W. Winkelmans (eds.) *World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research,* **3**, 171–184, Elsevier, Oxford.

Bradlow, E.T., B. Bronnenberg, G.J. Russell, N. Arora, D.R. Bell, S.D. Duvvuri, F.T. Hofstede, C. Sismeiro, R. Thomadsen and S. Yang (2005) Spatial models in marketing, *Marketing Letters*, **16** (3/4) 267-278.

Brandeau, M.L. and S.S. Chiu (1989) An overview of representative problems in location research, *Management Science*, **35** (6) 645-674.

Breusch, T.S. and A.R. Pagan (1980) The Lagrange Multiplier test and its applications to model specification in econometrics, *The Review of Economic Studies*, **47** (1) 239-253.

Bronnenberg, B.J. (2005) Spatial models in marketing research and practice, *Applied Stochastic Models in Business and Industry*, **21** (4/5) 335-343.

Bronnenberg, B.J., S.K. Dhar and J.P. Dubé (2007) Consumer packaged goods in the United States: National brands, local branding, *Journal of Marketing Research*, **44** (1) 4-13.

Bronnenberg, B.J. and P. Albuquerque (2003) Geography and marketing strategy in consumer packaged goods, in J.A.C. Baum and O. Sorenson (eds.) *Geography and Strategy*, 215-237, Elsevier Science, Oxford.

Bronnenberg, B.J. and V. Mahajan (2001) Unobserved retailer behavior in multimarket data: joint spatial dependence in market shares and promotion variables, *Marketing Science*, **20** (3) 284-299.

Bronnenberg, B.J. and C.F. Mela (2004) Market roll-out and retailer adoption for new brands, *Marketing Sciences*, **23** (4) 500-518.

Bronnenberg, B.J. and C. Sismeiro (2002) Using multimarket data to predict brand performance in markets for which no or poor data exist, *Journal of Marketing Research*, **39** (1) 1-17.

Brown, S. (1992) *Retail Location: A Micro-Scale Perspective*, Aldershot, Avebury.

Brown, K.H. and B. Uyar (2004) A hierarchical linear model approach for assessing the effects of house and neighbourhood characteristics on housing prices, *Journal of Real Estate Practice and Education*, **7** (1) 15-23.

Brunauer, W., S. Lang, P. Wechselberger and S. Bienert (2009) Additive hedonic regression models with spatial scaling factors: An application for rents in Vienna, *Journal of Real Estate Finance and Economics,* DOI 10.1007/s11146-009-9177-z.

Brunsdon, C., A.S. Fotheringham and M. Charlton (1998) Geographically weighted regression – modelling spatial non-stationarity, *The Statistician*, **47** (3) 431-443.

Brunsdon, C., J. McClatchey and D. Unwin (2001) Spatial variations in the average rainfall-altitude relationships in Great Britain: an approach using geographically weighted regression, *International Journal of Climatology*, **21** (4) 455-466.

Bucklin, R.E, S. Siddarth and J.M. Silva-Risso (2008) Distribution intensity and new car choice, *Journal of Marketing Research*, **45** (4) 473-486.

Bürgle, M. (2006a) Residential location choice model for the Greater Zurich area, paper presented at the *6th Swiss Transport Research Conference*, Ascona, March 2006.

Bürgle, M. (2006b) Modell der Wohnstandortwahl im Grossraum Zürich zur Verwendung in UrbanSim, *Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften*, **7**, NSL, ETH Zürich, Zürich.

Bürgle, M., M. Löchl, U. Waldner and K.W. Axhausen (2005) Land use and transport simulation: Applying UrbanSim in the Greater Zürich area, presentation, *9th Computers in Urban Planning and Urban Management Conference (CUPUM)*, London, June 2005.

Bundesamt für Raumentwicklung (2003) Räumliche Auswirkungen der Verkehrsinfrastrukturen „Lernen aus der Vergangenheit“, Methodologische Vorstudie, Teil II, Arbeitsbericht, Bern.

Bura, S., F. Guérin-Pace, H. Mathian, D. Pumain and L. Sanders (1996) Multiagent systems and the dynamics of a settlement system, *Geographical Analysis*, **28** (2) 161-178.

Burton, R.M. (2003) Computational laboratories for organization science: Questions, validity and docking, *Computational and Mathematical Organization Theory*, **9** (2) 91-108.

Butler, R.V. (1980) The specification of hedonic indexes for urban housing, *Land Economics*, **58** (1) 96-108.

Calvo, E. and M. Escolar (2003) The local voter: A geographically weighted approach to ecological inference, *American Journal of Political Science*, **47** (1) 189-204.

Cap Gemini (2008) World insurance report 2008, http://www.de.capgemini.com/m/de/tl/World_Insurance_Report_2008.pdf, retrieved 18th April 2008.

CEA Statistics (2008) European Insurance Figures, *CEA Statistics*, **36**, Brussels.

Chang, J.S. (2006) Models of the relationship between transport and land-use: A review, *Transport Reviews*, **26** (3) 325-350.

Charnes, A., W.W. Cooper and E. Rhodes (1978) Measuring the efficiency of decision making units, *European Journal of Operational Research*, **2** (6) 429-444.

Chatfield, C. (1995) *Problem solving: A statistician's guide*, Chapman & Hall, London.

Chen, G., M. S. Daskin, ZJ. M. Shen and S. Uryasev (2006) The α-reliable mean-excess regret model for stochastic facility location modelling, *Naval Research Logistics*, **53** (7) 617-626.

Cheshire, P. and S. Sheppard (1998) Estimating the demand for housing, land, and neighbourhood characteristics, *Oxford Bulletin of Economics and Statistics*, **60** (3) 357-382.

Chica-Olmo, J. (1995) Spatial estimation of housing prices and locational rents, *Urban Studies*, **32** (8) 1331-1344.

Chow, L.F., F. Zhao, X. Liu, M.T. Li and I. Ubaka (2006) Transit ridership model based on geographically weighted regression, *Transport Research Record*, **1972**, 105-114.

Christaller, W. (1933) *Die zentralen Orte in Süddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*, Fischer, Jena.

Church, R.L. (2002) Geographical information systems and location science, *Computers and Operations Research*, **29** (6) 541-562.

Church, R.L. and A.T. Murray (2009) *Business Site Selection, Location Analysis, and GIS*, John Wiley and Sons, Hoboken.

Ciari, F., M. Balmer and K.W. Axhausen (2007) Mobility tool ownership and mode choice decision processes in multi-agent transportation simulation, paper presented at the *7th Swiss Transport Research conference*, Ascona, September 2007.

Ciari, F., M. Löchl and K.W. Axhausen (2008) Location decisions of retailers: An agent-based approach, paper presented at the *15th International Conference on Recent Advances in Retailing and Services Science*, Zagreb, July 2008.

Clapp, J.M. (2003) A semiparametric method for valuing residential locations: Application to automated valuation, *Journal of Real Estate Finance and Economics*, **27** (3) 303-320.

Clapp, J.M. (2003) A semiparametric method for valuing residential locations: Application to automated valuation, *Journal of Real Estate Finance and Economics*, **27** (3) 303-320.

Clapp, J.M. (2004) A semiparametric method for estimating local house price indices, *Real Estate Economics*, **32** (1) 127-160.

Clapp, J.M., J.A. Fields and C. Ghosh (1990) An examination of profitability in spatial markets: The case of life insurance agency locations, *The Journal of Risk and Insurance*, **57** (3) 431-454.

Clark, S.D. (2007) Estimating local car ownership models, *Journal of Transport Geography*, **15** (3) 184-197.

Clarke, I. D. Bennison and J. Pal (1997) Towards a contemporary perspective of retail location, *International Journal of Retail and Distribution Management*, **25** (2) 59-69.

Clarke, G. and J. Stillwell (2004) Introduction, in: J. Stillwell and G. Clarke (eds.) *Applied GIS and Spatial Analysis*, 1-7, John Wiley and Sons, Chichester.

Cliff, A.D. and J.K. Ord (1973) *Spatial Autocorrelation*, Pion, London.

Cliff, A.D. and J.K. Ord (1981) *Spatial Processes – Models and Applications*, Pion, London.

Cliquet, G. and J.M. Josselin (2002) *Stratégies de localisation des enterprises commerciales et industrielles: de nouvelle perspectives*, De Boeck, Bruxelles.

Coelho, F.J., C. Easingwood and A. Coelho (2003) Exploratory evidence of channel performance in single vs multiple channel strategies, *International Journal of Retail & Distribution*, **31** (11) 561-573.

Coelho, F.J. and C. Easingwood (2004) Multiple channel systems in services: pros, cons and issues, *The Service Industries Journal*, **24** (5) 1-29.

Coelho, F.J. and C. Easingwood (2008) A model of the antecedents of multiple channel usage, *Journal of Retailing and Consumer Services*, **15** (1), 32-41.

Colwell, P.F. and G. Dilmore (1999) Who was first? An examination of an early hedonic study, *Land Economics*, **75** (4) 620-626.

Cornish, S.L. (1997) Strategies for the acquisition of market intelligence and implications for the transferability of information invputs, *Annals of the Association of American Geographers*, **87** (3) 451-470.

Craig, C.S., A. Ghosh and S. McLafferty (1984) Models of the retail location process: A review, *Journal of Retailing*, **60** (1) 5-35.

Credit Suisse Economic Research (2005) Die Schweizer Versicherungswirtschaft – Entwicklungen und Perspektiven, *Spotlight*, **March 2005**.

Crooks, A., C. Castle and M. Batty (2008) Key challenges in agent-based modelling for geo-spatial simulation, *Computers, Environment and Urban Systems*, **32** (6) 417-430.

Cropper, M.L, L.B. Deck and K.E. McConnell (1988) On the choice of functional form for hedonic price functions, *The Review of Economics and Statistics*, **70** (4) 668-675.

CUSPA – Center for Urban Simulation and Policy Analysis (2006) UrbanSim Homepage. http://www.urbansim.org, retrieved October 2006.

Daunfeldt, S.O., N. Rudholm and U. Rämme (2007) Road pricing and retail revenues: Results from the Stockholm road pricing trial, *HUI Working Papers*, **10**, The Swedish Retail Institute, Stockholm.

Davidson, W.R., D.J. Sweeney and R.W. Stampfl (1988) *Retailing Management*, Wiley, New York.

De Graaff, T., R.J.G.M. Florax and P. Nijkamp (2001) A general misspecification test for spatial regression models: Dependence, heterogeneity, and nonlinearity, *Journal of Regional Science*, **41** (2) 255-276.

De la Barra, T., B. Perez and N. Vera (1984) TRANUS-J: Putting large scale models into small computers, *Environment and Planning B*, **11** (1) 87–101.

Dekker, J. and P. Rietveld (2009) The adoption of geo-ICT in economics: Increasing opportunities for spatial research in economics, in H.J Scholten, N. van Manen and R. van de Velde (eds.) *Geospatial Technology and the Role of Location in Science*, 133-161, Springer, Dordrecht.

Deleersnyder, B., I. Geyskens, K. Gielens and M.G. Dekimpe (2002) How cannibalistic is the internet channel?, *International Journal of Research in Marketing*, **19** (4) 337-348.

Des Rosiers, F., M. Thériault and P.Y. Villeneuve (2000) Sorting out access and neighbourhood factors in hedonic price modelling, *Journal of Property Investment and Finance*, **18** (3) 291-315.

Dixon, L., N. Clancy, S.A. Seabury and A. Overton (2006) *The National Flood Insurance Program's Market Penetration Rate-Estimates and Policy Implications*, Rand Corporation, Santa Monica.

Djurdjevic, D., C. Eugster (2008) Estimation of hedonic models using a multilevel approach: An application for the Swiss rental market, *Swiss Journal of Economics and Statistics*, **144** (4) 679-701.

Doh, J.P. and E.D. Hahn (2008) Using spatial methods in strategy research, *Organizational Research Methods*, **11** (4) 659-681.

Du, H. and C. Mulley (2006) Relationship between transport accessibility and land value – Local model approach with geographically weighted regression, *Transportation Research Record*, **1977**, 197-205.

Dubé, J.P. and P. Manchanda (2005) Differences in dynamic brand competition across markets: An empirical analysis, *Marketing Science*, **24** (1) 81-95.

Dubin, R. (1998) Predicting house prices using multiple listings data, *Journal of Real Estate Finance and Economics*, **17** (1) 34-60.

Dubin, R. (2004) Spatial lags and spatial errors revisited: Some Monte Carlo evidence, in J.P. LeSage and R.K. Pace (eds.) *Spatial and Spatiotemporal Econometrics*, 75-98, Elsevier, Amsterdam.

Duckham, M., K. Mason, J. Stell and M. Worboys (2001) A formal approach to imperfection in geographic information, *Computers, Environment and Urban Systems*, **25** (1) 89-103.

Dumm, R.E. and R.E. Hoyt (2003) Insurance distribution channels: markets in transition, *Journal of Insurance Regulation*, **22** (1) 27-47.

DYNASMART, http://mctrans.ce.ufl.edu/featured/dynasmart, accessed January 2008.

Eastman, K.L., J.K. Eastman and A.D. Eastman (2002) Issues in marketing online insurance products: An exploratory look at agents' use, attitudes, and views of the impact of the internet, *Risk Management and Insurance Review*, **5** (2) 117-134.

Ekeland, I., J.J. Heckman and L. Nesheim (2004) Identification and estimation of hedonic models, *Journal of Political Economy*, **112** (1) 60-109.

Enz, R. (2000) The S-curve relation between per-capita income and insurance penetration, *The Geneva Papers on Risk and Insurance*, **25** (3) 396-406.

Ettema, D. and H. Timmermans (2006) Multi-agent modelling of urban systems: A progress report of PUMA system, *Stadt Region Land*, **81**, 165-171.

Farber, S. and M. Yeates (2006) A comparison of localized regression models in a hedonic house price context, *Canadian Journal of Regional Science*, **29** (3) 405-420.

Fahrländer, S. (2006) Semiparametric construction of spatial generalized hedonic models for private properties, *Swiss Journal of Economics and Statistics*, **142** (4) 501-528.

Federal Office of Private Insurance (2008) Vermittlerportal, www.vermittleraufsicht.ch, retrieved 7th January 2008.

Felder, S. and H. Tauchmann (2009) Regional differences in the efficiency of health production: An artefact of spatial dependence?, *Ruhr Economic Papers*, **112**, RWI, Essen.

Ferber, J. (1999) *Multi-agent Systems: An Introduction to Distributed Artificial Intelligence*, Addison-Wesley, Harlow.

Ferrell, O.C. and M.D. Hartline (2005) *Marketing Strategy*, Thomson South-Western, Mason.

Ferrier, G.D. and C.A.K. Lovell (1990) Measuring cost efficiency in banking – Econometric and linear programming evidence, *Journal of Econometrics*, **46** (1/2) 229-245.

Fik, T.J., D.C. Ling and G.F. Mulligan (2003) Modeling spatial variation in housing prices: A variable interaction approach, *Real Estate Economics*, **31** (4) 623-646.

Fischer, M.M (2006) *Spatial Analysis and Geocomputation: Selected Essays*, Springer, Berlin.

Fletcher, M., P. Gallimore and J. Mangan (2000) Heteroscedasticity in hedonic house price models, *Journal of Property Research*, **17** (2) 93-108.

Florax, R.J.G.M. and A.J. Van Der Vlist (2003) Spatial econometric data analysis: Moving beyond traditional models, *International Regional Science Review*, **26** (3) 223-243.

Förster, R. and S. Kytzia (2004) How can integrative modelling contribute to sustainable regional development in practice? The potential of group model building - Results of a case study, in C. Pahl-Wostl, S. Schmidt and T. Jakeman (eds.) *iEMSs 2004 International Congress "Complexity and Integrated Resources Management"*, Osnabrück, June 2004.

Fotheringham, A.S., C. Brunsdon and M. Charlton (2000) *Quantitative Geography – Perspectives on Spatial Data Analysis*, Sage Publications, London.

Fotheringham, A.S., C. Brunsdon and M. Charlton (2001) Spatial variation in school performance: a local analysis using geographically weighted regression*, Geographical and Environmental Modelling,* **5** (1) 176-193.

Fotheringham, A.S., C. Brunsdon and M. Charlton (2002) *Geographically Weighted Regression – The Analysis of Spatially Varying Relationships*, John Wiley & Sons, West Sussex.

Fotheringham, A.S. and T.C. Pitts (1995) Directional variation in distance decay, *Environment and Planning A*, **27** (5) 715-729.

Fotheringham, A.S. and P.A. Rogerson (1993) GIS and spatial analytical problems, *International Journal of Geographical Information Systems*, **7** (1) 3-19.

Fotheringham, A.S. and P.A. Rogerson (2009) Introduction, in A.S. Fotheringham and P.A. Rogerson (eds.), *The Sage Handbook of Spatial Analysis*, 1-4, Sage, London.

Fox, J. (2002) *An R and S-PLUS Companion to Applied Regression*, Sage Publications, Thousand Oaks.

Franses, P.H. (2005) On the use of econometric models for policy simulation in marketing, *Journal of Marketing Research*, **42** (1) 4-14.

Frazier, G.L. and W.M. Lassar (1996) Determinants of distribution intensity, *Journal of Marketing*, **60** (4) 39-51.

Frazier, G.L. (1999) Organizing and managing channels of distribution, *Journal of the Academy of Marketing Science*, **27** (2) 226-240.

Freeman, A.M. (1993) *The Measurement of the Environmental and Resource Values – Theories and Methods*, Resources for the Future, Washington, DC.

Free Software Foundation (1991) GNU General Public License, Version 2. http://www.gnu.org/licenses/gpl.txt., October 2006.

Galabova, T. and R. Lester (2001) Is insurance a luxury?, *The Actuary*, **2001** (12) 32-34.

Gao, X, Y. Asami and C.J.F. Chung (2006) An empirical evaluation of spatial regression models, *Computers & Geosciences*, **32** (8) 1040-1051.

Gaski, J.F. (1984) The theory of power and conflict in channels of distribution, *Journal of Marketing*, **48** (3) 9-29.

Gelfand, A.E. (1998) Spatio-temporal modeling of residential sales data, *Journal of Business and Economic Statistics*, **16** (2) 312-321.

Gelfand, A.E., S. Banerjee, C.F. Sirmans, Y. Tu and S.E. Ong (2007) Multilevel modelling using spatial processes: Application of the Singapore housing market, *Computational Statistics and Data Analysis*, **51** (7) 3567-3579.

Gelfand, A.E., H.J. Kim, C.F. Sirmans and S. Banerjee (2003) Spatial modelling with spatially varying coefficient processes, *Journal of the American Statistical Association*, **98** (462) 387-396.

Gensler, S., M.G. Dekimpe and B. Skiera (2007) Evaluating channel performance in multi-channel environments, *Journal of Retailing and Consumer Services*, **14** (1) 17-23.

Getis, A. (2008) A history of the concept of spatial autocorrelation: A geographer's perspective, *Geographical Analysis*, **40** (3) 297-309.

Ghosh, A. and S.L. McLafferty (1987) *Location Strategies for Retail and Service Firms*, D.C. Heath, Lexington.

Ghosh, A. and G. Rushton (1987) Progress in location-allocation modelling, in A. Ghosh and G. Rushton (eds.) *Spatial Analysis and Location-Allocation Models*, 1-18, Van Nostrand Reinhold, New York.

Ghosh, A. (1990) *Retail Management*, Dryden Press, Chicago.

Gilbert, D. (2003) *Retail Marketing Management*, 2nd edition, Prentice Hall, Harlow.

Gilbert, N. and K.G. Troitzsch (2005) *Simulation for the Social Scientist*, 2nd edition, McGraw-Hill Publishing, Columbus, OH.

Golan, A. G.G. Judge and D. Miller (1996) *Maximum Entropy Econometrics: Robust Information with Limited Data*, Wiley, New York.

Golan, A. G.G. Judge and D. Miller (1997) The maximum entropy approach to estimation and inference: An overview, in T.B. Fomby and R.C. Hill (eds.) *Advances in Econometrics: Applying Maximum Entropy to Econometric Problems*, **12**, 3-24, JAI Press, Greenwich.

Goldstein, H. (1987) *Multilevel Models in Educational and Social Research*, Griffin & Co, Oxford.

Goldstein, H. (1995) *Multilevel Statistical Models*, 3rd edition, Arnold, London.

Golledge, R.G. (2009) The future for spatial analysis, in A.S. Fotheringham and P.A. Rogerson (eds.), *The Sage Handbook of Spatial Analysis*, 481-486, Sage, London.

Goodchild, M.F. (2009) Challenges in spatial analysis, in A.S. Fotheringham and P.A. Rogerson (eds.), *The Sage Handbook of Spatial Analysis*, 465-477, Sage, London.

Goodman, A.C. (1998) Andrew Court and the invention of hedonic price analysis, *Journal of Urban Economics*, **44** (2) 291-298.

Goodman, A.C. and T.G. Thibodeau (1998) Housing market segmentation, *Journal of Housing Economics*, **7** (2) 121-143.

Goodman, A.C. and T.G. Thibodeau (2003) Housing market segmentation and hedonic prediction accuracy, *Journal of Housing Economics*, **12** (3) 181-201.

Gotway, C.A. and L.J. Young (2002) Combining incompatible spatial data, *Journal of the American Statistical Asscociation*, **97** (458) 632-648.

Graf, A. and P. Maas (2008) Kundenzugangswege bei Versicherern: Herausforderungen und Strategien im Multi-Channel-Management, *IVW Management Information*, **30** (2**)** 19-21.

Grether, E.T. (1983) Regional-spatial analysis in marketing, *Journal of Marketing*, **47** (4) 36-43.

Greve, H.R. (2000) Market niche entry decisions: Competition, learning, and strategy in Tokyo banking, 1894-1936, *Academy of Management Journal*, **43** (5) 816-836.

Griffith, D.A. (1996) Some guidelines for specifying the geographic weights matrix contained in spatial statistical models, in S.L. Arlington and D.A. Griffith (eds.) *PHB Practical Handbook of Spatial Statistics*, 65-82. CRC Press, Boca Raton.

Griffith, D. A. (2003) *Spatial Autocorrelation and Spatial Filtering*, Springer, Berlin.

Griffith, D.A. (2008) Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR), *Environment and Planning A*, **40** (11) 2751-2769.

Griliches, Z. (1971) *Price Indexes and Quality Change: Studies in New Methods of Measurement*, Harvard University Press, Camebridge.

Haas, T. (1995) Local prediction of a spatio-temporal process with an application to wet sulphate deposition, *Journal of the American Statistical Association*, **90** (432) 1189-1199.

Haas, T. (1996) Multivariate spatial prediction in the presence of nonlinear trend and covariance nonstationarity, *Environmetrics*, **7** (2) 145-166.

Habib, M.A. and E.J. Miller (2008) Influence of transportation access and market dynamics and property values – Multilevel spatiotemporal models of housing price, *Transportation Research Record*, **2076**, 183-191.

Hadayeghi, A., A.S. Shalaby and B.N. Persaud (2003) Macrolevel accident prediction models for evaluating safety of urban transportation systems, *Transportation Research Record*, **1840**, 87-95.

Hägerstrand, T. (1968) *Innovation Diffusion as a Spatial Process*, University of Chicago Press, Chicago.

Haining, R. (2003) *Spatial Data Analysis – Theory and Practice*, Cambridge University Press, Cambridge.

Haining, R. (2009) The spatial nature of spatial data, in A.S. Fotheringham and P.A. Rogerson (eds.), *The Sage Handbook of Spatial Analysis*, 5-39, Sage, London.

Hale, T.S. and C.R. Moberg (2003) Location science research: A review, *Annals of Operations Research*, **123** (1-4) 21-35.

Halvorsen, R. and H.O. Pollakowski (1981) Choice of functional form for hedonic price equations, *Journal of Urban Economics*, **10** (1) 37-49.

Hanssens, D.H., L.J. Parsons and R.L. Schultz (2001) *Market Response Models, Econometric and Time Series Analysis*, Kluwer Academic, Dordrecht.

Harris, R., P. Sleight and R. Webber (2005) *Geodemographics, GIS and Neighbourhood Targeting*, Wiley, Chichester.

Harrison, J.R., Z. Lin, G.R. Carroll and K.M. Carley (2007) Simulation modelling in organizational and management science, *Academy of Management Review*, **32** (4) 1229-1245.

Hattemer, K. (2008) Am liebsten möglichst billig, aber mit Vertreter – Neue Markt-Media-Studien legen Trends für Vertrieb und Marketing offen, *Versicherungswirtschaft*, **63** (2) 89-93.

Hauri, H.R., K.W. Axhausen and M. Löchl (2008) Standortwahl von Generalagenturen in der Versicherungsbranche, *IVW Management Information*, **30** (1) 27-31.

Henneberry, J. (1998) Transport investment and house prices, *Journal of Property Valuation and Investment*, **16** (2) 144-158.

Heppenstall, A., A. Evans and M. Birkin (2006) Using Hybrid Agent-Based Systems to Model Spatially-Influenced Retail Markets, *Journal of Artificial Societies and Social Simulation*, **9** (3) 2.

Hernandez, T., D. Bennison and S. Cornelius (1998) The organisational context of retail locational planning, *GeoJournal*, **45** (4) 299–308.

Hernandez, T. and D. Bennison (2000) The art and science of retail location decisions, *International Journal of Retail & Distribution Management*, **28** (8) 357-67.

Hernandez, T. and M. Biasiotto (2001) Retail location decision-making and store portfolio management, *Canadian Journal of Regional Science*, **24** (3) 399-418.

Hess, R.L., R.S. Rubin and L.A. West (2004) Geographic information systems as a marketing information system technology, *Decision Support Systems*, **38** (2) 197-212.

Holmes, T. (2008) The diffusion of Wal-Mart and economies of density, *NBER Working Paper Series*, **13783**, National Bureau of Economic Research, Cambridge.

Hotelling, H. (1929), Stability in competition, *Economic Journal*, **39** (153) 41-57.

Huang, Y and Y. Leung (2002) Analysing regional industrialisation in Jiangsu province using geographically weighted regression, *Journal of Geographical Systems*, **4** (2) 233-249.

Hunt, J.D. (1994) Calibrating the Naples land use and transport model, *Environment and Planning B: Planning and Design*, **21** (5) 569-590.

Hunt, J.D., D.S. Kriger and E.J. Miller (2005) Current operational urban land-use-transport modelling frameworks: A review, *Transport Review*, **25** (3) 329-376.

Hunziker, P., A. Maier, A. Nippe, M. Tresch, D. Weers and P. Zemp (1998) Data mining at a major bank: Lessons from a large marketing application, in J.M. Zytkow and M. Quafafou (eds.) *Principles of Data Mining and Knowledge Discovery*, 345-351, Springer, Berlin.

Hussels, S., D. Ward and R. Zurbruegg (2005) Stimulating the demand for insurance, *Risk Management and Insurance Review*, **8** (2) 257-278.

Iacono, M., D. Levinson and A. El-Geneidy (2008) Models of transportation and land use change: A guide to the territory, *Journal of Planning Literature*, **22** (4) 323-340.

Janelle, D.G. and M.F. Goodchild (2009) Location across disciplines: Reflections on the CSISS experience, in H.J Scholten, N. van Manen and R. van de Velde (eds.) *Geospatial Technology and the Role of Location in Science*, 15-29, Springer, Dordrecht.

Jank, W. and P.K. Kannan (2005) Understanding geographical markets of online firms using spatial models of customer choice, *Marketing Science*, 24 (4) 623-634.

Japelli, T. and L. Pistaferri (2002) Tax incentives and the demand for life insurance: Evidence from Italy, *Journal of Public Economics*, **87** (7/8) 1779-1799.

Jones, J.P., E. Casetti (1992) *Applications of the Expansion Methods*, Routledge, London.

Jones, K. (1991) Specifying and estimating multi-level models for geographical research, *Transactions of the Institute of British Geographers*, **16** (2) 148-159.

Jones, K. and N. Bullen (1994) Contextual models of urban prices: A comparison of fixed- and random-coefficient models developed by expansion, *Economic Geography*, **70** (3) 252-272.

Johansson, B. and U. Forslund (2009) The analysis of location, collocation and urbanization economies, in C. Karlsson (ed.) *Handbook of Research on Cluster Theory*, 39-66, Edward Elgar, Cheltenham.

Kain, J.F. and J.M. Quigley (1970) Measuring the value of housing quality, *Journal of the American Statistical Association*, **65** (330) 532-548.

Kanaroglou, P. and D.M. Scott (2002) Integrated urban transportation and land-use models for policy analysis, in M. Dijst, W. Schenkel and I. Thomas (eds.) *Governing Cities on the Move – Functional and Management Perspectives on Transformations of European Urban Infrastructures*, 42-72, Ashgate, Aldershot.

Karande, K. and J.R. Lombard (2005) Location strategies of broad-line retailers: An empirical investigation, *Journal of Business Research*, **58** (5) 687-695.

Kelejian, H. and I.R. Prucha (1998) A generalized spatial two-stage least square procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate and Finance Economics*, **17** (1) 99-121.

Kelejian, H. and I.R. Prucha (1999) A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review*, **40** (2) 509-533.

Kelejian, H. and D.P. Robinson (2000) The influence of spatially correlated heteroskedasticity on tests for spatial correlation, in L. Anselin and R.J.G.M Florax (eds.) *Advances in Spatial Econometrics – Methodology, Tools and Applications*, 79-97, Springer, Berlin.

Kestens, Y., M. Thériault and F. Des Rosiers (2006) Heterogeneity in hedonic modelling of house prices: looking at buyers' household profiles, *Journal of Geographic Systems*, **8** (1) 61-96.

Kim, W.J., D. Mayers and C.W. Smith (1996) On the choice of insurance distribution systems, *The Journal of Risk and Insurance*, **63** (2) 207-227.

Kim, C.W., T.T. Phipps and L. Anselin (2003) Measuring the benefits of air quality improvement: A spatial hedonic approach, *Journal of Environmental Economics and Management*, **45** (1) 24-39.

Kimsey, M.J., J. Moore and P. McDaniel (2008) A geographically weighted regression analysis of Douglas-Fir site index in North Central Idaho, *Forest Science*, **54** (3) 356-366.

King, S.F. and J.S. Liou (2004) A framework for internet channel evaluation, *International Journal of Information Management*, **24** (6) 473-488.

Kissling, W.D. and G. Carl (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models, *Global Ecology and Biogeography*, **17** (1) 59-71.

Klinkenberg, B. (2003) The true cost of spatial data in Canada, *The Canadian Geographer*, **47** (1) 37-49.

Knospe, J. (2008) Direktversicherer: Mühsam nährt sich das Eichhörnchen, *Versicherungswirtschaft*, **63** (13) 1097.

Krafft, M. and M. Matalla (2006) *Retailing in the 21th Century – Current and Future Trends*, Springer, Berlin.

Kriström, B. (2008) Applying hedonics in the housing market: An illustration, in: in A. Baranzini, J.V. Ramirez, C. Schaerer and P. Thalmann (eds.) *Hedonic Methods in Housing Markets*, 247-263, Springer, New York.

Lancaster, K. J. (1966) A new approach to consumer theory, *Journal of Political Economy*, **74** (2) 132-157.

Lancaster, K.J. (1971) *Consumer Demand: A New Approach*, Columbia University Press, New York.

Laulajainen, R. (1987) *Spatial Strategies in Retailing*, Reidel, Dordrecht.

Lee, D.B. Jr. (1973) Requiem for large-scale models, *Journal of the American Institute of Planners*, **39** (3) 163-178.

Lenzi, A., G. Millo (2005) Regional heterogeneity and spatial spillovers in the Italian insurance market, paper presented at the *Workshop on Applied Economics of Insurance*, Rome, March 2005.

LeSage, J.P. (1997) Bayesian estimation of spatial autoregressive models, *International Regional Science Review*, **20** (1/2) 113-129.

LeSage, J.P. and R.K. Kelly (2004) Introduction, in: J.P. LeSage and R.K. Pace (eds.) Spatial and Spatiotemporal Econometrics, *Advances in Econometrics*, **18**, 1-32, Springer, Amsterdam.

LeSage J.P. and O. Parent (2007) Bayesian model averaging for spatial econometric models, *Geographical Analysis*, **39** (3) 241-267.

Leung, Y., C.L.Mei and Wen-Xiu Zhang (2000) Statistical tests for spatial nonstationarity based on the geographically weighted regression model, *Environment and Planning A*, **32** (1) 9-32.

Li, D., F. Moshirian, P. Nguyen and T. Wee (2007) The demand for life insurance in OECD countries, *The Journal of Risk and Insurance*, **74** (3) 637-652.

Lloyd, C.D. (2007) *Local Models for Spatial Analysis*, CRC Press, Boca Raton.

Lloyd, C. and I. Shuttleworth (2005) Analysing commuting using local regression techniques: scale, sensitivity, and geographical patterning, *Environment and Planning A*, **37** (1) 81-103.

Lodish, L.M. (2007) Another reason academics and practitioners should communicate more, *Journal of Marketing Research*, **44** (1) 23-25.

Löchl, M. (2006) Real estate and land price models for the Greater Zürich application of UrbanSim, *Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften*, **6**, NSL, ETH Zürich, Zürich.

Löchl, M. (2006) Land use effects of road pricing – A literature review, paper presented at the *6th Swiss Transport Research conference*, Ascona, March 2006.

Löchl, M. (2007) Considering spatial dependence in hedonic rent price regression, paper presented at the *7th Swiss Transport Research Conference*, Ascona, September 2007.

Löchl, M. (2008) Standortplanung im Detail-/Einzelhandel – Auswertung von Interviews mit Unternehmen in Deutschland und der Schweiz, *Arbeitsberichte Verkehrs- und Raumplanung*, **492**, IVT, ETH Zürich, Zürich.

Löchl, M. and K.W. Axhausen (forthcoming) Modelling hedonic residential rents for land use and transport simulation while considering spatial effects, *Journal of Transport and Land Use*.

Löchl, M., M. Bürgle and K.W. Axhausen (2007) Implementierung des integrierten Flächennutzungsmodells UrbanSim für den Grossraum Zürich – ein Erfahrungsbericht, *DISP*, **168**, 13-25.

Löchl, M., H.R. Hauri and K.W. Axhausen (2009) Agents, space and market shares: A spatial analysis of the Swiss Insurance market, *Arbeitsberichte Verkehrs- und Raumplanung*, **557**, IVT, ETH Zürich, Zürich.

Löning, H. and M. Besson (2002) Can distribution channels explain differences in marketing and sales performance measurement systems?, *European Management Journal*, **20** (6) 597-609.

Lombardo, S., M. Petri and D. Zotta (2004). Intelligent GIS and retail location dynamics: A multi agent system integrated with ArcGis, *Lecture Notes in Computer Science*, **3044**, 1046-1056.

Longley, P.A. (2008) Geographic information systems (GIS), in K.K. Kemp (ed.) *Encyclopedia of Geographic Information Science*, 190-194, Sage, Los Angeles.

Longley, P.A. and Carolina Tobón (2004) Spatial dependence and heterogeneity in patters of hardship: An intra-urban analysis, *Annals of the Association of American Geographers*, **94** (3) 503-519.

Macy, M.W. and R. Willer (2002) From factors to actors: Computational sociology and agent-based modeling, *Annual Review of Sociology*, **28** (1) 143-166.

Malpezzi, S. (2003) Hedonic pricing models: A selective and applied review, in T. O'Sullivan and K. Gibb (eds.) *Housing Economics and Public Policy*, 67-89, Blackwell Science, Oxford.

Manson, S. (2008) Agent-based models, in K.K. Kemp (ed.) *Encyclopedia of Geographic Information Science*, 4-6, Sage, Thousand Oaks.

Martínez, F.J. (2000) Towards a land use and transport interaction framework, in D.A. Hensher and K. Button (eds.) *Handbooks in Transport – Handbook 1: Transport Modelling*, 154-164, Elsevier, The Hague.

Martins-Filho, C. and O. Bin (2005) Estimation of hedonic price functions via additive nonparametric regression, *Empirical Economics*, **30** (1) 93-114.

McFadden, D. (1978) Modelling the choice of residential location, in A. Karlquist, L. Lundquist, F. Snickars and J.W. Weibull (eds.) *Spatial Interaction Theory and Planning Models*, 75-96, North-Holland, Amsterdam.

McGoldrick, P. (2002) *Retail Marketing*, 2nd edition, McGraw-Hill, London.

McMillen, D.P. (2003) Spatial autocorrelation or model misspecification, *Internatioal Regional Science Review*, **26** (2) 208-217.

Meister, K., M. Rieser, F. Ciari, A. Horni, M. Balmer and K. W. Axhausen (2009) Anwendung eines agentenbasierten Modells der Verkehrsnachfrage auf die Schweiz, *Strassenverkehrstechnik*, **53** (5) 269-280.

Militino, A.F., M.D. Ugarte and L. García-Reinaldos (2004) Alternative models for describing spatial dependence among dwelling selling prices, *Journal of Real Estate and Economics*, **29** (2) 193-209.

Miller, E.J. (2006) Integrated Urban Models: Theoretical Prospects, paper presented at the *11th International Conference on Travel Behaviour Research*, Kyoto, August 2006.

Miller, M., J.D. Hunt, J.E. Abraham and P.A. Salvini (2004) Microsimulating urban systems, *Computers, Environment and Urban Systems*, **28** (1/2) 9-44.

Mittal, V., W.A. Kamakura and R. Govind (2004) Geographic patterns in customer service and satisfaction: An empirical investigation, *Journal of Marketing,* **68** (3) 48-62.

Moeckel, R., B. Schwarze and M. Wegener (2006) Das Projekt ILUMASS – Mikro-Simulation der räumlichen, demographischen und wirtschaftlichen Entwicklung, *Stadt Region Land*, **81**, 53-62.

Modelistica (2006) Modelistica – Tranus. http://www.modelistica.com, October 2006.

Monmonier, M. (1996) *How to Lie with Maps*, 2nd edition, University of Chicago Press, Chicago.

Moran, P.A.P. (1950) Notes on continuous stochastic phenomena, *Biometrika*, **37** (1-2) 17-23.

Mossin, J. (1968) Aspects of rational insurance purchasing, *The Journal of Political Economy*, **76** (4) 553-568.

Mosteller, F. and J.W. Tukey (1977) *Data Analysis and Regression – A Second Course in Statistics*, Addison-Wesley Publishing, Reading.

MySQL AB (2006) MySLQ Community Edition. http://www.mysql.com/products/database/mysql/community_edition.html, October 2006.

Nakata, H. and Y. Sawada (2007) Demand for non-life insurance: a cross-country analysis, *CIRJE Discussion Papers*, **461**, Center for International Research on the Japanese Economy, University of Tokyo, Tokyo.

Nakaya, T. (2001) Local spatial interaction modelling based on the geographically weighted regression approach, *GeoJournal*, **53** (4) 347-358.

Nakaya, T., A.S. Fotheringham, K. Hanaoka, G. Clarke, D. Ballas and K. Yano (2007) Combining microsimulation and spatial interaction models for retail location analysis, *Journal of Geographical Systems,* **9** (4) 345-369.

Nedović-Budić, Z. and J.K. Pinto (1999) Interorganizational GIS: Issues and prospects, *The Annals of Regional Science*, **33** (2) 183-195.

Neslin, S.A., D. Grewal, R. Leghorn, V. Shankar, M.L. Teerling, J.S. Thomas and P.C. Verhoef (2006) Challenges and opportunities in multichannel customer management, *Journal of Service Research*, **9** (2) 95-112.

Nwogugu, M. (2006) Site selection in the US retailing industry, *Applied Mathematics and Computation*, **182** (2) 1725-1734.

Olteanu, A.M., S. Mustière and A. Ruas (2006) Matching imperfect spatial data, in: M. Caetano and M. Painho (eds.) *Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 694-704, Instituto Geográphico Português, Lisbon.

Ord, J.K. (1975) Estimating methods for models of spatial interaction, *Journal of the American Statistical Association*, **70**, 120-126.

Orford, S. (1999) *Valuing the Built Environment – GIS and House Price Analysis*, Ashgate, Brookefield.

Orford, S. (2000) Modelling spatial structures in local housing market dynamics: A multilevel perspective, *Urban Studies*, **37** (9) 1643-1671.

O'Sullivan, A. (2007) *Urban Economics*, 6th edition, McGraw-Hill, New York.

O'Sullivan, D. and D.J. Unwin (2002) *Geographic Information Analysis*, Wiley and Sons, Hoboken.

Pace, R.K., R. Barry, J.M. Clapp and M. Rodriquez (1998) Spatialtemporal autoregressive models of neighbourhood effects, *Journal of Real Estate Finance and Economics*, **17** (1) 15-33.

Pace, R.K. (1993) Nonparametric methods with applications to hedonic models, *Journal of Real Estate Finance and Economics*, **7** (3) 185-204.

Pace, R.K. (1995) Parametric, semiparametric, and nonparametric estimation of characteristic values within mass assessment and hedonic pricing models, *Journal of Real Estate Finance and Economics*, **11** (3) 195-217.

Pace, R.K. (1998) Appraisal using generalized additive models, *Journal of Real Estate Research*, **15** (1) 77-100.

Páez, A. (2005) Local analysis of spatial relationships: A comparison of GWR and the expansion method, *Lecture Notes in Computer Science*, **3482**, 162-172.

Páez, A., F. Long and S. Farber (2008) Moving window approaches for hedonic price estimation: an empirical comparison of modelling techniques, *Urban Studies*, **45** (8) 1565-1581.

Parker, D.C., S.M. Manson, M.A. Janssen, M.J. Hoffmann and P. Deadman (2003) Multi-Agent systems for the simulation of land-use and land-cover change: A review, *Annals of the Association of American Geographers*, **93** (2) 314-337.

Parmeter, C.F., D.J. Henderson and S.C. Kumbhakar (2007) Nonparametric estimation of a hedonic price function, *Journal of Applied Econometrics*, **22** (3) 695-699.

PartnerRe (2001) *Distribution in Insurance and new business models – European Outlook*, PartnerRe, Bermuda.

Pfeil, C.M., T. Posselt and N. Maschke (2008) Incentives for sales agents after the advent of the internet, *Marketing Letters*, **19** (1) 51-63.

Pioch, E. and J. Byrom (2004) Small independent retail firms and locational decision-making: outdoor leisure retailing by the crags, *Journal of Small Business and Enterprise Development*, **11** (2) 222-232.

Psychonomics (2007) *Kundenmonitor e-Assekuranz 2007*, Eigenverlag, Köln.

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

Putnam, S.H. (1994) Integrated land use and transportation models: an overview of progress with DRAM and EMPAL, with suggestions for further research, paper presented at the *73rd Annual Meeting of the Transportation Research Board*, Washington, D.C., January 1994.

Python Software Foundation (2006) The Python Programming Language. http://www.python.org/, October 2006.

Quddus, M., A. Carmel and M.G.H. Bell (2006) The impact of the congestion charge on retail: the London experience, paper presented at the *85th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2006.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna.

Raudenbush, S.W. and A. S. Bryk (2002) *Hierarchical Linear Models – Applications and Data Analysis Methods*, 2nd edition, Sage Publications, Thousand Oaks.

Regan, L. and S. Tennyson (2000) Insurance distribution systems, G. Dionne (ed.) *Handbook of Insurance*, 709-748, Kluwer, Boston.

Reibstein, D.J. and P.W. Farris (1995) Market share and distribution: A generalization, a speculation, and some implications, *Marketing Science*, **14** (3) 190-202.

ReVelle, C.S. and H.A. Eiselt (2005) Location analysis: A synthesis and survey, *European Journal of Operational Research*, **165**, (1) 1-19.

Rey, S.J., L. Anselin (2006) Recent advances in software for spatial analysis in the social sciences, *Geographical Analysis*, **38** (1) 1-4.

Rieser, M., U. Beuck, K. Nagel and K.W. Axhausen (2007) Multi-agent transport simulations and economic evaluation, *Arbeitsbericht Verkehrs- und Raumplanung*, **457**, IVT, ETH Zurich, Zurich.

Rietveld, P. and F. Bruinsma (1998) *Is Transport Infrastructure Effective? Transport Infrastructure and Accessibility: Impacts on The Space Economy*, Springer, Berlin.

Rogers, D.S. (2007) Retail location analysis in practice – Retail site analysis is much more than "location, location, location", *Research Review*, **14** (29) 73-78.

Rosen, S. (1974) Hedonic prices and implicit markets: product differentiation in price competition, *Journal of Political Economy*, **82** (1) 35-55.

Salvini, P.A. and E.J. Miller (2005) ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems, *Networks and Spatial Economics*, **5** (2) 217-234.

Schenk, T.A., G. Löffler and J. Rauh (2008) Agent-based simulation of consumer behavior in grocery shopping on a regional level, *Journal of Business Research*, **60** (8) 894-903.

Schlag, C.H. (2004) Determinants of demand for life insurance products – theoretical concepts and empirical evidence, paper presented at the *Annual Meeting of the Swiss Society of Economics and Statistics*, Basel, March 2004.

Schletti, B. (2008) Allianz kitzelt Platzhirsch – Zurich findet es nicht lustig, *Tagesanzeiger*, 23th July 2008, 21.

Schulz, B. and F. Dosch (2005) Trends der Siedlungsentwicklung und ihre Steuerung in der Schweiz und Deutschland, *DISP*, **160**, 5-15.

Ševčiková, H., A. Raftery and P. Waddell (2007) Assessing Uncertainty in Urban Simulations Using Bayesian Melding, *Transportation Research Part B*, **41** (6) 652-669.

Shankar, V. (2008) Strategic allocation of marketing resources: Methods and managerial insights, *Marketing Science Institute Special Report*, 08-207, Texas A&M University, College Station.

Shearmur, R., P. Apparicio, P. Lizion and M. Polèse (2007) Space, time, and local employment growth: An application of spatial regression analysis, *Growth and Change*, **38** (4) 696-722.

Shin, K, S. Washington and K. Choi (2007) Effects of transportation accessibility on residential property values – application of spatial hedonic price model in Seoul, South Korea, Metropolitan Area, *Transportation Research Record*, **1994**, 66-73.

Simmonds, D. and O. Feldman (2005) Land-use modelling with DELTA: Update and experience, paper presented at the *9th Computers in Urban Planning and Urban Management Conference* (CUPUM), London, June 2005.

Sirmans, G.S., C.F. Sirmans and J.D. Benjamin (1989) Determining apartment rent: The value of amenities, services and external factors, *The Journal of Real Estate Research*, **4** (2) 33-44.

Sirmans, G.S., D.A. Macpherson and E.N. Zietz (2005) The composition of hedonic pricing models, *Journal of Real Estate Literature*, **13** (1) 3-43.

Sorenson, O. and P.G. Audia (2000) The social structure of entrepreneurial activity: Geographic concentration of footwear production in the United States, 1940-1989, *American Journal of Sociology*, **106** (2) 424-462.

Sorenson, O. and J.A.C. Baum (2003) Editor's introduction: geography and strategy: the strategic management of space and place, in J.A.C. Baum and O. Sorenson (eds.) Geography and strategy, *Advances in Strategic Management*, **20**, 1-19, Elsevier Science Ltd., Oxford.

Stakhovych, S. and T.H.A. Bijmolt (2008) Specification of spatial models: A simulation study on weights scenarios, *Papers in Regional Science*, **88** (2) 389-408.

Statistical Office of Canton Zürich (2004) Genossenschaftlich Wohnen, *Statistik.Info*, 20, Canton Zürich, Zürich. http://www.statistik.zh.ch/themenportal/themen/down.php?id=394&fn=2004_20.pdf.

Statistical Office of Canton Zürich (2008) Mikrozensus 2005: Kilometerleisung. http://www.statistik.zh.ch/themenportal/themen/daten_detail.php?id=722&tb=&mt=2.

Swiss Insurance Association (2009) *Facts and Figures*, Swiss Insurance Association, Zürich. http://www.svv.ch/index.cfm?id=9707 retrieved 9th March 2009.

SwissRe (2007) Statistical appendix, *Sigma,* **2007** (4) 29-40.

Taylor, L.O. (2008) Theoretical foundations and empirical developments in hedonic modeling, in: A. Baranzini, José Ramirez, Caroline Schaerer and Philippe Thalmann (eds.) *Hedonic Methods in Housing Markets – Pricing Environmental Amenities and Segregation*, Springer, New York.

Te Brömmelstroet (2009) The relevance of research in planning support systems: A response to Janssen et al., *Environment and Planning B*, **36** (1) 4-7.

Ter Hofstede, F., M. Wedel and J.B.E.M. Steenkamp (2002) Identifying spatial segments in international markets, *Marketing Science*, **21** (2) 160-177.

Thanassoulis, E. (1993) A comparison of regression analysis and data envelopment as alternative methods for performance assessment, *Journal of the Operational Research Society*, **44** (11) 1129-1144.

Thériault, M., F. Des Rosiers, P. Villeneuve and Y. Kestens (2003) Modelling interactions of location with specific value of housing attributes, *Property Management*, **21** (1) 25-62.

Tiefelsdorf, M., D. Griffith and B. Boots (1999) A variance-stabilizing coding scheme for spatial link matrices, *Environmental and Planning A*, **31** (1) 165-180.

Timmerman, H.J.P. (2006) The saga of integrated land use and transport modelling: How many more dreams before we wake up?, in K.W. Axhausen (ed.) *Moving through Nets: The Physical and Social Dimension of Travel*, 219-248, Elsevier, Oxford.

Tobler, W.R. (1970) A computer movie simulating urban growth in the Detroit region, *Economic Geography*, **46** (2) 234-240.

Torrens, P.M. (2000) How land-use-transportation models work, *Working Paper Series*, **20**, Centre for Advanced Spatial Analysis, University College London, London.

TRANSIMS, http://transims.tsasa.lanl.gov/, accessed February 2008.

Transport for London (2005) *Congestion Charging: Impacts Monitoring – Third Annual Report*, TfL, London.

Transport for London (2006) *Congestion Charging: Impacts Monitoring – Fourth Annual Report*, TfL, London.

Transport for London (2007) *Congestion Charging: Impacts Monitoring – Fifth Annual Report*, TfL, London.

Tretvik, T. (1999) The EUROPRICE project: The Trondheim toll ring and the effects on retailing, *SINTEF Memo N-12/99*, SINTEF, Trondheim.

Tretvik, T. (2003) Urban road pricing in Norway: Public acceptability and travel behaviour, in J. Schade and B. Schlag (eds.) *Acceptability of Transport Pricing Strategies*, 77-92, Elsevier, Oxford.

Trigg, D.W. and D.H. Leach (1967) Exponential smoothing with an adaptive response rate, *Operational Research Quarterly*, **18** (1) 53-59.

Trigo-Gamarra, L. (2008) Reasons for the coexistence of different distribution channels: An empirical test for the German insurance market, *The Geneva Papers*, **33** (3) 389-407.

Tsay, A.A. and N. Agrawal (2004) Channel conflict and coordination in the e-commerce age, *Production and Operations Management*, **13** (1) 93-110.

Tschopp, M. and K.W. Axhausen (2007) Transport infrastructure and spatial development in Switzerland between 1950 and 2000, paper presented at the *86th Annual Meeting of the Transportation Research Board,* Washington, D.C., January 2007.

Unwin, D.J. (2008) Spatial Analysis, in K.K. Kemp (ed.) *Encyclopedia of Geographic Information Science*, 392-396, Sage, Los Angeles.

Unwin, A. and D. Unwin (1998) Exploratory spatial data analysis with local statistics, *The Statician*, **47** (3) 415-421.

Upton, G. and B. Fingleton (1985) *Spatial Data Analysis by Example: Volume 1 – Point Pattern and Quantitative Data*, John Wily and Sons, Chichester.

Valente, J., S.S. Wu, A. Gelfand and C.F. Sirmans (2005) Apartment rent prediction using spatial modelling, *Journal of Real Estate Research*, **27** (1) 105-136.

Valos, M.J. and A. Vocino (2006) An integrative marketing channel performance measurement framework, *Database Marketing & Customer Strategy Management*, **14** (1) 17-28.

Van Dyke Parunak, H., R. Savit and R.L. Riolo (1998) Agent-based modeling vs. equation-based modeling: A case study and users' guide, *Lecture Notes in Computer Science*, **1534**, 10-25.

Van Leeuwen, E.S., J.E. Hagens and P. Nijkamp (2007) Multi-agent systems: A tool in spatial planning, *DISP*, **170** (3) 19-32.

Verhoef, P.C. and B. Donkers (2005) The effect of acquisition channels on customer loyalty and cross-buying, *Journal of Interactive Marketing*, **19** (2) 31-43.

Villeneuve, B. (2000) Life insurance, in G. Dionne (ed.) *Handbook of Insurance*, 901-932, Kluwer, Boston.

Von Thünen, J.H. (1842) *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Leopold, Rostock.

Voss, P. (2008) Spatial econometrics, in K.K. Kemp (ed.) *Encyclopedia of Geographic Information Science*, 409-412, Sage, Thousand Oaks.

Vovsha, P., E. Petersen and R. Donnelly (2002) Microsimulation in travel demand modelling: lessons learned from the New York best practice model, *Transportation Research Record*, **1805**, 68-77.

Vrtic, M. and P. Fröhlich (2008) *Aktualisierung des nationalen Personenverkehrsmodells für den öffentlichen und privaten Verkehr*, report to the Federal Department of the Environment, Transport, Energy and Communications, Bern.

Vrtic, M., D. Lohse, P. Fröhlich, C. Schiller, N. Schüssler, H. Teichert and K.W. Axhausen (2007) Simultaneous two-dimensionally constrained disaggregate trip generation, distribution and mode choice model: Theory and application for a Swiss national model, *Transportation Research Part A*, **41** (9) 857–873.

Vuorinen, I., R. Järvinen and U. Lehtinen (1998) Content and measurement of productivity in the service sector – A conceptual analysis with an illustrative case from the insurance business, *International Journal of Service Industry Management*, **9** (4) 377-396.

Waddell, P. (1998) An urban simulation model for integrated policy analysis and planning: Residential location and housing market components of UrbanSim, *Proceedings of the 8th World Conference on Transport Research*, Antwerp, Belgium, July 1998.

Waddell, P., B.J.L. Berry and I. Hoch (1993) Residential property values in a multimodal urban area: New evidence on the implicit price of location, *Journal of Real Estate Finance and Economics*, **7** (2) 117-141.

Waddell, P. (2002) UrbanSim: modelling urban development for land use, transportation and environmental planning, *Journal of the American Planning Association*, **68** (3) 297-314.

Waddell, P., A. Borning, M. Noth, N. Freier, M. Becke, and G. Ulfarsson (2003) Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim, *Networks and Spatial Economics*, **3** (1) 43-67.

Waddell, P., H. Ševčíková, D. Socha, E.J. Miller and K. Nagel (2005) Opus: An open platform for urban simulation, paper presented at the *9th Computers in Urban Planning and Urban Management Conference (CUPUM)*, London, Juni 2005.

Waddell, P. and G.F. Ulfarsson (2003) Dynamic simulation of real estate development and land prices within an integrated land use and transportation system, paper presented at the *83th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2003.

Wagner, P. and M. Wegener (2007) Urban land use, transport and environment models, *DISP*, **170**, 45-56.

Waldner, U., M. Löchl, M. Bürgle and K.W. Axhausen (2005) Haushaltsbefragung zur Wohnsituation im Grossraum Zürich – Feldbericht, *Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften*, **1**, NSL, ETH Zürich, Zürich.

Wall, M.M. (2004) A closer look at the spatial structure implied by the CAR and SAR models, *Journal of Statistical Planning and Inference*, **121** (2) 311-324.

Waller, L.A., L. Zhu, C.A. Gotway, D.M. Gorman, P.J. Gruenewald (2007) Quantifying geographic variations in association between alcohol distribution and violence: A comparison of geographically weighted regression and spatially varying coefficient models, *Stochastic Environmental Research and Risk Assessment*, **21** (5) 573-588.

Wang, N., C.L. Mei and X.D. Yan (2008) Local linear estimation of spatially varying coefficient models: An improvement on the geographically weighted regression technique, *Environment and Planning A*, **40** (4) 986-1005.

Webb, K.L. and J.E. Hogan (2002) Hybrid channel conflict: causes and effects on channel performance, *Journal of Business & Industrial Marketing*, **17** (5) 338-356.

Webber, R. (2008) Geodemographics, in K.K. Kemp (ed.) *Encyclopedia of Geographic Information Science*, 169-171, Sage, Los Angeles.

Weber, A. (1909) *Theory of the Location of Industries*, C.J. Friedrich (trans.), Charles Scribner's Sons, New York.

Wegener, M. (2004) Overview of land-use and transport models, in D.A. Hensher, K.J. Button, K.E. Haynes and P.R. Stopher (eds.) *Handbook of Transport Geography and Spatial Systems*, 127-146, Elsevier, Oxford.

Wegener, M. (2009) Energy scarcity and climate change: The challenge for urban models, paper presented at the *11th International Conference on Computers in Urban Planning and Urban Management* (CUPUM 2009), University of Hong Kong, June 2009.

Wheeler, D.C. (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression, *Environment and Planning A*, **39** (10) 2464-2481.

Wheeler, D.C. (2009) Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso, *Environmental and Planning A*, **41** (3) 722-742.

Wheeler, D.C. and M. Tiefelsdorf (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression, *Journal of Geographical Systems*, **7** (2) 161-187.

Wheeler, D.C. and L.A. Waller (2009) Comparing spatially varying coefficient models: A case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests, *Journal of Geographical Systems*, **11** (1) 1-22.

Widrow, G. and M.E. Hoff (1960) Adaptive switching circuits*, Institute of Radio Engineers Western Electric Show and Convention Record*, part 4, 96-104.

Willhelmsson, M. (2002) Spatial models in real estate economics, *Housing, Theory and Society*, **19** (2) 92-101.

Wood, S. and S. Browne (2007) Convenience store location planning and forecasting – A practical research agenda, *International Journal of Retail and Distribution Management*, **35** (4) 233-255.

Wolk, A. and B. Skiera (2009) Antecedents and consequences of internet channel performance, *Journal of Retailing and Consumer Services*, **16** (3) 163-173.

Wüest & Partner (2007) *Immo-Monitoring 2008*, Wüest & Partner, Zürich.

Xie, Y. and D.G. Brown (2007) Simulation in spatial analysis and modeling, *Computers, Environment and Urban Systems*, **31** (3) 229-231.

Young, L.J., C.A. Gotway, J. Yang, G. Kearny and C. DuClos (2008) Assessing the association between environmental impacts and health outcomes: A case study from Florida, *Statistics in Medicine*, **27** (20) 3998-4015.

Yu, D. (2004) *Modeling Housing Market Dynamics in the City of Milwaukee: A Geographically Weighted Regression Approach*, Department of Geography, University of Wisconsin, Milwaukee.

Yu, D., Y.D. Wei and C. Wu (2007) Modeling spatial dimensions of housing prices in Milwaukee, WI, *Environment and Planning B: Planning and Design*, **34** (6) 1085-1102.

Zentes, J., D. Morschett and H. Schramm-Klein (2007) *Strategic Retail Management*, Gabler, Wiesbaden.

Zhang, L. and H. Shi (2004) Local modeling of tree growth by geographically weighted regression, *Forest Science*, **50** (2) 225-244.

Zhao, L. and J Harris (2006) Market research design on modelling propensity to purchase and market potential: Using GIS and data mining as the tools, paper presented at the *6th IEEE International Conference on Data Mining*, Hong Kong, December 2006.

Zietz, E.N. (2003) An examination of the demand for life insurance, *Risk Management and Insurance Review*, **6** (2) 159-191.

Annexe 1 Additional tables

Table 13 Variables tested but not included in the models due to insignificance

Variable
Share of home owners
Population older than 18 years
Share of male inhabitants
Share of inhabitant younger than 19 years
Share of inhabitants between 19 and 39 years
Share of inhabitants older than 59 years
Inhabitants younger than 19 years divided by inhabitants aged 25 to 49
Inhabitants younger than 19 years divided by inhabitants aged 25 to 54
Inhabitants younger than 19 years divided by inhabitants aged 25 to 59
Language regions (dummies: French, Italian, Rhaeto-Romanic)

Table 14 Estimated OLS parameters for models with spatial explanatory variables only

	Model 3	Sign.	Model 4	Sign.
(Constant)	9.515	***	8.366	***
Ln(CARTT_CBD)	-0.258	***		
CARACC05			0.064	***
PTACC05	-0.012	***	-0.004	
Ln(RAILSTATION)	-0.019	***	-0.015	**
AUTOBAHN	-0.061	**	-0.064	**
AIRNOISE	-0.009		-0.053	***
Ln(HOTREST_JOBS)	0.000		0.020	***
Ln(POP_DENS)	-0.093	***	-0.091	***
FOREIGNERS (logit)	-0.087	***	-0.091	***
Ln(TAXLEVEL)	-0.289	***	-0.368	***
SLOPE (logit)	0.018	***	0.027	***
Ln(VIEW_LAKE)	0.015	***	0.018	***
Ln(VIEW_ALL)	0.007		0.005	
Ln(SOLAR_EVE)	0.071	***	0.056	***
Adjusted R-square	0.235		0.215	
F-test	208.800	***	182.200	***
Moran's I	0.165	***	0.185	***

Probability of rejecting H_0 = *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Annexe 2 Curriculum Vitae

Name	Michael Löchl
Date of birth	29th July 1975
Place of birth	Düsseldorf/Germany
Marital status	Single

Education

Since 06/2004	PhD student, Institute of Transport Planning and Systems (IVT), ETH Zurich
08/2000 – 05/2001	Studies in the Joint Master's Programme in City and Regional Planning/Civil Engineering, Georgia Institute of Technology, Atlanta
10/1996 – 11/2002	Master (Dipl.-Ing.) in Spatial Planning, University of Dortmund, Dortmund

Working experience

Since 10/2009	Project Manager, Volkswirtschaftsdirektion Kanton Zürich, Amt für Verkehr
01/2004 – 09/2009	Research Assistant, Institute of Transport Planning and Systems (IVT), ETH Zurich
12/2002 – 11/2003	Research Assistant, Institute of Spatial Planning (IRPUD), University of Dortmund
09/2001 – 08/2002	Research Assistant, Institut für Landes- und Stadtentwicklungsforschung des Landes NRW (ILS), Dortmund
01/2001 – 07/2001	Research Assistant, Geostats, Atlanta
08/2000 – 12/2000	Research Assistant, Hughes, Good, O'Leary and Ryan (HGOR), Atlanta
04/2000 – 05/2000	Intern, Kommunikative Stadt- und Regionalentwicklung Sinning und Knieling (KoRiS), Hannover
10/1998 – 04/2000	Research Assistant, Institut für Landes- und Stadtentwicklungsforschung des Landes NRW (ILS), Dortmund

Publications

Löchl, M. and K.W. Axhausen (forthcoming) Modeling hedonic rent prices for land use and transport simulation while considering spatial effects, *Journal of Land Use and Transport*.

Erath, A., M. Löchl and K.W. Axhausen (2009) Graph-theoretical analysis of the Swiss road and railway networks over time, *Networks and Spatial Economics*, **9** (3) 379-400.

Löchl, M., H.R. Hauri and K.W. Axhausen (2009) Agents, space and market shares: A spatial analysis of the Swiss Insurance market, Arbeitsberichte Verkehrs- und Raumplanung, 557, IVT, ETH Zürich, Zürich.

Ciari, F., M. Löchl and K.W. Axhausen (2008) Location choice of retailers - An agent-based approach, paper presented at the *15th International Conference on Recent Advances in Retailing and Services Science*, Zagreb, July 2008.

Hauri, H.R., K.W. Axhausen und M. Löchl (2008) Standortwahl von Generalagenturen in der Versicherungsbranche, *IVW Management Information – St. Galler Trendmonitor*, **30** (1) 27-31.

Löchl, M. (2008) Standortplanung im Detail-/Einzelhandel: Auswertung von Interviews mit Unternehmen in Deutschland und der Schweiz, *Arbeitsberichte Verkehrs- und Raumplanung*, **492**, IVT, ETH Zürich, Zürich.

Hauri, H.R., M. Löchl und K.W. Axhausen (2008) Aufgaben des Corporate Real Estate und Facility Managements (FM), *in puncto Fachzeitschrift für Facility Management*, **2**, 2-4.

Löchl, M. (2007) Considering spatial dependence in hedonic rent price regression, paper presented at the *7th Swiss Transport Research Conference*, Ascona, September 2007.

Löchl, M., M. Bürgle und K.W. Axhausen (2007) Implementierung des integrierten Flächennutzungsmodells UrbanSim für den Grossraum Zürich – ein Erfahrungsbericht, *DISP*, **168**, 13-25.

Löchl, M., M. Bürgle und U. Waldner (2007) Handbuch Simulationsmodell Grossraum Zürich, Arbeitsberichte Polyprojekt *"Zukunft urbane Kulturlandschaften"*, **10**, NSL, ETH Zürich, Zürich.

Axhausen, K.W., M. Löchl, R. Schlich, T. Buhl and P. Widmer (2007) Fatigue in long-duration travel diaries, *Transportation*, **34** (2) 143-160.

Löchl, M. (2006) Land use effects of road pricing – a literature review, paper presented at *6th Swiss Transport Research Conference*, Ascona, March 2006.

Kohte M., M. Bürgle und M. Löchl (2005) Zur Zukunft urbaner Kulturlandschaften: *Halbzeittagung des ETH-Forschungsprojektes am Netzwerk Stadt und Landschaft*, *DISP*, **163**, 78-80.

Löchl, M., U. Waldner und M. Bürgle (2005) Haushaltsbefragung zur Wohnsituation im Grossraum Zürich – Tabellenband, *Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften*, **2**, NSL, ETH Zürich, Zürich.

Waldner, U., M. Löchl, M. Bürgle und K.W. Axhausen (2005) Haushaltsbefragung zur Wohnsituation im Grossraum Zürich – Feldbericht, *Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften*, **1**, NSL, ETH Zürich, Zürich.

Bürgle, M., M. Löchl, U. Waldner and K.W. Axhausen (2005) Land use and transport simulation: Applying UrbanSim in the Greater Zurich area, Vortrag, *Computers in Urban Planning and Urban Management*, London, June 2005.

Löchl, M., S. Schönfelder, R. Schlich, T. Buhl, P. Widmer und K.W. Axhausen (2005) Stabilität des Verkehrsverhaltens, Schlussbericht, SVI 2001/514, *Schriftenreihe*, **1120**, Bundesamt für Strassen, UVEK, Bern.

Bürgle, M., M. Löchl und U. Waldner (2005) Entwicklung eines Simulationsmodells - Infrastruktur, Erreichbarkeit und Raumentwicklung, *DISP*, **160**, 94-95.

Löchl, M., K.W. Axhausen and S. Schönfelder (2005) Analysing Swiss longitudinal travel data, paper presented at *5th Swiss Transport Research Conference*, Ascona, March 2005.

Bürgle, M., M. Löchl and U. Waldner (2005) Integrated Transport and Land Use Simulation for Zurich, paper presented at *5th Swiss Transport Research Conference*, Ascona, March 2005.

Machguth, M., M. Löchl und M. Bürgle (2004) Berechnung von Routen- und Verkehrsmittelalternativen für den Datensatz Thurgau 2003, *Arbeitsberichte Verkehrs- und Raumplanung*, **231**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Machguth H. und M. Löchl (2004) Geokodierung 6-Wochenbefragung Thurgau 2003, *Arbeitsberichte Verkehrs- und Raumplanung*, **219**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Gløersen, E., J. Grimm, A. Hill, Ch. Lindner, M. Löchl and C. Schürmann (2004) Mountain Areas in Europe: Delineation, Database and Analysis. In: School of Environmental Sciences (eds.) *Proceedings of the GIS Research UK 12th Annual Conference*, 92-97, University of East Anglia, Norwich.

Dittrich-Wesbuer, A., M. Frehn und M. Löchl (2004) Verkehrliche Orientierungen und ÖPNV-Nutzung in der Stadtregion Münster, *Verkehr und Technik*, **57** (2) 39-45; (4) 115-122; (5), 166-172.

Gløersen, E., J. Grimm, A. Hill, Ch. Lindner, M. Löchl and C. Schürmann (2003) Comprehensive analysis of European mountain areas by using GIS techniques, *Proceedings of the 18th ESRI European User Conference*, Innsbruck, October 2003.

Wolf, J., M. Loechl, M. Thompson and C. Arce (2003) Trip rate analysis in GPS-enhanced personal travel surveys, in P. Stopher und P. M. Jones (Hrsg.) *Transport Survey Quality and Innovation*, 483-498, Pergamon, Oxford.

Bäumer, D., M. Löchl und S. Rabe (2000) Mobilitätsdienstleistungen für Jugendliche und junge Erwachsene, U.MOVE. Jugend und Mobilität, *ILS-Schriften*, **150**, 47-59, Institut für Landes- und Stadtentwicklungsforschung, Dortmund.

Presentations:

Löchl, M. und R. Haase (2009) Mietzinsen und Erreichbarkeit bei Büro- und Wohnimmobilien: Neue Ergebnisse für den Grossraum Zürich, IVT-Seminar „Wie schafft Erreichbarkeit Werte?“, ETH Zürich, Mai 2009.

Löchl, M. (2009) Modellierung von Mietpreisen und Flächennutzung im Kanton Zürich, Gastvortrag an der Fakultät für Raumplanung, Technische Universität Dortmund, Februar 2009.

Löchl, M. and K.W. Axhausen (2008) The Zürich Experience, European UrbanSim Users' Conference, Zürich, March 2008.

Löchl, M. (2007) Modelling hedonic residential rent prices in Canton Zurich, Symposium Hedonic Methods in Real Estate, Geneva, June 2007.

Löchl, M. (2006) Verkehr und Raumentwicklung - Konzepte und Modelle, Urbaplan, Lausanne, September 2006.

Löchl, M. (2006) Erkenntnisse der Auswirkungen und Ansätze zur Abschätzung von räumlichen Effekten des Road Pricing, 4. Verkehrsreform-Workshop, SRL-Tagung, Fachgruppe Forum Mensch und Verkehr, Hofgeismar, Februar 2006.

Löchl, M. (2005) Modelling businesses and employment: UrbanSim and its application in Zurich, Workshop of land use transportation (LUT) modelling group, London, July 2005.

Löchl, M. (2004) Mikrosimulation der Flächennutzung und des Verkehrsverhaltens - Eine Anwendung für den Grossraum Zürich, Deutsche Hochschultagung der Verkehrsplaner, St. Marienthal, September 2004.

Löchl, M., M. Price, C. Schürmann and E. Gløersen (2004) Developing typologies to comprehend the diversity of Europe's mountains, 30th Congress of the International Geographical Union Conference, Glasgow, August 2004.

Löchl, M. and E. Gløersen (2003) Pan-European data gathering at NUTS 5 level – experiences from the DG Regio mountain study, Eurostat, Working group "Rural Development Statistics", Luxembourg, November 2003.