Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Institut für Verkehrswirtschaft
Lehrstuhl für Betriebswirtschaftslehre, insb. Verkehr
Schwerpunkt: Operations & Supply Chain Management

Advanced Spatial Analytics and Management

Models, Methods and Applications

Sven Müller*

18. Juli 2014

Habilitationsleistungen

Kumulative Habilitation im Fach Betriebswirtschaftslehre
an der Fakultät Betriebswirtschaft
der Universität Hamburg

*Wiesenhof 24, D-20143 Hamburg, Germany, mueller_sven@gmx.net, www.geospatial-science.com

Contents

1 Summary

Geographical or spatial factors - such as location - play an important role in everyday decision making of organizations since the data used for decision making has a geographic component in nearly all cases (Day et al. (1987), Crossland et al. (1995), Benoit and Clarke (1997), Vlachopoulou et al. (2001), Grimshaw (2000), Porter and Stern (2001), Graf and Mudambi (2005), and Miller et al. (2006)). For many operations, such as health care, financial, energy, insurance, communications, transportation, logistics and retail, location intelligence or spatial analytics provide very specific benefits, which translate to increased revenues, reduced costs, and improved efficiency for any organization. For example, a retail chain may be interested in analyzing the shopping destination choice behavior of their customers to estimate branch patronization of existing and potential branch locations. Based on these estimates the retail chain might modify their network of branches to increase patronage, revenues and profits. Using tailored spatial models and methods helps the management to make better locational decisions. In particular, the endogenous incorporation of the (spatial) choice behavior of customers in quantitative decision models seems to be beneficial for many decision processes in many operations (line planning in public transportation, for example).

In this thesis, I incorporate econometric models, that describe the spatial choice behavior, into mathematical programs for spatial decision making (facility location planning, for example). I consider this approach as *advanced spatial analytics and management*. The thesis comprehends a series of papers published in peer-reviewed journals. The first part (Chapter 2.1) is dedicated to customer demand analysis. I distinguish between spatially aggregated data (Chapter 2.1.1) and disaggregated data (Chapter 2.1.2). In the former case spatial econometric models and methods are applied (geographically weighted regression, for instance), while in the latter one discrete choice models are applied. In the second part (Chapter 2.2), I show how the demand models (based on aggregated or disaggregated data) can be incorporated in mathematical programs for districting (Chapter 2.2.2) and facility location planning (Chapter 2.2.1) and how the resulting problems might be solved. The applications comprehend the analysis of transport mode choice behavior, destination choice behavior, service quality, and innovation diffusion as well as the planning of facility locations of schools, retail branches, and health care units or sales territory alignment and sales resource allocation.

2 Articles

2.1 Spatial Demand Analysis

2.1.1 Aggregate Data

1. Müller, S. Wilhelm, P., Haase, K. (2013): **Spatial dependencies and spatial drift in public transport seasonal ticket revenue data**. *Journal of Retailing and Consumer Services* **20**(3), 334-348.

2. Müller, S. (2012): **Identifying spatial nonstationarity in German regional firm start-up data**. *Review of Regional Research* **32**(2), 113-132.

3. Müller, S., Haase, K.: **Local revenue response to service quality: spatial effects in seasonal ticket revenue data**. *European Journal of Marketing*, accepted for publication.

Journal of Retailing and Consumer Services 20 (2013) 334–348

Contents lists available at SciVerse ScienceDirect

Journal of Retailing and Consumer Services

journal homepage: www.elsevier.com/locate/jretconser

Spatial dependencies and spatial drift in public transport seasonal ticket revenue data

Sven Müller *, Pascal Wilhelm, Knut Haase

Institute for Transport Economics, University of Hamburg, D-20146 Hamburg, Germany

ARTICLE INFO

Article history:
Received 11 September 2012
Received in revised form
29 November 2012
Accepted 2 January 2013
Available online 27 February 2013

Keywords:
Spatial econometrics
Geographically weighted regression
Spatial autocorrelation
Spatial non-stationarity
Seasonal ticket revenue
Public transport
Revenue response function

ABSTRACT

When firms' customers are located in spatially dispersed areas, it can be difficult to manage service quality on a geographically small scale because the relative importance of service quality might vary spatially. Moreover, standard approaches discussed so far in the marketing science literature usually neglect spatial effects, such as spatial dependencies (spatial autocorrelation for example) and spatial drift (spatial non-stationarity). We propose a comprehensive approach based on spatial econometric methods that covers both issues. Based on the real company data on seasonal ticket revenue of a local public transport service company, we show that addressing such spatial effects of service data can improve management's ability to implement programs aimed at enhancing seasonal ticket revenue. In particular, the article shows how a spatial revenue response function might be specified.

1. Introduction

Revenue management is essential to the success of a firm. Hence, reliable information about the influencing factors that drive revenue is needed. The coherences between these factors and revenue might be mapped by a revenue response function (Tollefson and Lessig, 1978) – that is, the revenue response of customers expressed as a function of price, service quality and/or advertising activities, for example. Managers might use such a function in order to predict the increase of revenue due to improvements in service or the expenditures in advertising and promotion. However, especially at firms that provide spatially varying service (quality), managers face several challenges in implementing revenue management strategies (Rust et al., 1995). For spatially dispersed service areas, revenue and the importance placed on service quality will vary over locations. Moreover, the firm's ability to provide sufficient service may also vary spatially. Here we consider two types of spatial effects: (i) *spatial dependencies*: locations proximate to one another usually share resources, history, and sociodemographic and economic make-up. Therefore, consumer culture, lifestyle, values, attitudes, benefits, and consumption tend to be spatially associated as well. Empirical support for such local similarities in cultural, attitudinal, and behavioral patterns can be found in several studies (e.g., Foster and Gorr, 1986; Garber et al., 2004; Bronnenberg and Mahajan, 2001; Bronnenberg, 2005; and Anselin, 2003). (ii) *spatial drift*: Let us say we find a formal link between revenue and a certain promotion. Then, would it be likely that this relationship is constant over the whole study area? The answer might be "no" since there are reasons why this formal link could be different at different locations. For example, due to (unobserved) attitudinal differences between the customers of two locations A and B, the customers located in A might be alienated by a specific promotion while those located in B are stimulated in order to consume a given product. On a global scale (neglecting these spatial differences) the effect of the promotion might be averaged out. For the incidence of spatial drift there exists strong evidence in the literature in general (e.g., Leung et al., 2000; Yu, 2006; Huang and Leung, 2002; Bitter et al., 2007) and in the marketing science literature in particular (Mittal et al., 2004). As such, incorporation of spatial dependencies and spatial drift in revenue response functions is important. In contrast, the key assumption in the traditional marketing science literature is that the behavior of a consumer is conditionally independent of the behavior of another consumer and that this behavior is spatially homogeneous (Bradlow et al., 2005). Fortunately, there exist tailored models and methods, namely spatial statistics and spatial econometrics, in order to account for spatial effects (Wheeler and Paez, 2010). Moreover, nowadays these models

* Corresponding author. Tel.: +49 40 42838 9021; fax: +49 40 42838 2924.
E-mail addresses: sven.mueller@wiso.uni-hamburg.de, mueller_sven@gmx.net (S. Müller).

0969-6989/$ - see front matter
http://dx.doi.org/10.1016/j.jretconser.2013.01.005

and methods are easily accessible, because they are implemented in standard state-of-the-art, public domain software packages.

In this study, we employ revenue and service data of the year 2005 of a monopolistic public transport company located in the city of Dresden, Germany. We consider a specific segment of the public transport revenue: seasonal ticket revenue.[1] A seasonal ticket of the considered public transport company is valid for a month or a year and costs 40 Euro or 425 Euro, respectively in 2005. The seasonal ticket enables the customer to use all public transport services within the city of Dresden. Revenue and service data of public transport companies are particularly appealing, because public transport companies operate in a spatially dispersed service area and they provide spatially varying service quality. Interestingly, there is very sparse literature on cross-sectional data of seasonal tickets in general (Brown, 2002; Forrest et al., 2002; McDonald, 2010) and on public transport seasonal tickets in particular (FitzRoy and Smith, 1999).

This paper contributes to the literature in two ways. First, we propose a novel but intelligible modeling approach in order to deal with spatial dependencies and spatial drift simultaneously. Second, we present a study based on real company seasonal revenue data and service data. Based on this, we specify a revenue response function of service variables. For the requirements of response functions see Albers (2011), for example. We control for several socio-demographic and socio-economic variables as well as land-use variables. Altogether we present an interesting marketing science application of spatial models as demanded by Bradlow et al. (2005).

The remainder of this paper is organized as follows. In Section 2 we provide a brief discussion of econometric models that adequately deal with spatial dependencies (Section 2.1) and spatial drift (Section 2.2). In Section 2.3 we propose an approach that accounts (at least partially) for spatial dependencies and spatial drift simultaneously. This is followed by a discussion of the data and the model specification (Section 3). In Section 4 we discuss the results of the estimation and present managerial insights. Conclusions are drawn in the final section.

2. Spatial econometric models

The discussion in this section is rather brief. For a more detailed discussion of spatial models we refer to Anselin (1988) and LeSage and Pace (2009); for a specific marketing science view we refer to Bradlow et al. (2005) and Bronnenberg (2005).

Roughly speaking, spatial econometric models assume that individuals (or, more generally, units of analysis, such as postal codes) can be located in a space. Typically, responses by individuals are assumed to be correlated in such a manner that individuals near one another in space generate similar outcomes[2] – as stated in Tobler's First Law of Geography (Tobler, 1970): "Everything is related to everything else, but near things are more related than distant things". The methodology can integrate complex spatial correlations between entities into a model in a parsimonious and flexible manner. As a typical marketing related application we might imagine a consumer, whose decision to adopt a new Internet service is affected by interactions with other consumers who live or work in the same district. This kind of spatial pattern is called spatial dependency or more specific spatial autocorrelation. We discuss models that account for such kind of spatial effects in Section 2.1. Although these models are able to cope with spatial autocorrelation, they still imply a global relationship between the dependent variable (here: revenue) and the independent variables (service quality for example). Therefore, we call such models, *global spatial models*. Exemplarily, Mittal et al. (2004) argue that geography dictates the parameters of a satisfaction rating regression model due to differences in lifestyle and climate. This kind of spatial effect can be regarded as a representation of unobserved spatial heterogeneity (sometimes called spatial non-stationarity) in which the parameters (as opposed to the dependent variable per se) follow a spatial process: spatial drift. The geographically weighted regression – as discussed in Section 2.2 – is based on the idea, that the model parameters are a function of the observation's location in space. Hence, geographically weighted regression models are able to unmask spatial drift. Unfortunately, these models are (i) complex in terms of interpretation and it might be difficult to deduce a coherent management strategy from the results of a geographically weighted regression. (ii) geographically weighted regression models only deal accidentally with spatial autocorrelation. Based on the results of the geographically weighted regression we try to find spatial clusters in which revenue is similarly responsive to service variables. Then the clusters are incorporated in the global spatial models yielding *spatially heterogeneous models* (Section 2.3).

2.1. Global spatial models

Spatial dependence states, that observations in a spatial data set depend on other observations at other locations. More specifically, observations with similar values coincide with similar locations. Formally, spatial autocorrelation can be defined as:

$$\mathrm{Cov}[z_i, z_j] = \mathrm{E}[z_i z_j] - \mathrm{E}[z_i]\mathrm{E}[z_j] \neq 0, \quad \text{for } i \neq j \tag{1}$$

where i, j refer to individual observations (locations) and z_i and z_j are the values of a random variable of interest at that location. As pointed out by Anselin (1988), this spatial dependence can be either caused by model misspecification, measurement problems like spill-over effects or result from the spatial organization and structure of the phenomena. Since the premise of independence of observations cannot be held in the presence of spatial autocorrelation, estimates can be inefficient, biased and/or inconsistent. In contrast to temporal autocorrelation, spatial autocorrelation can potentially go in any direction in space, increasing the complexity of this influence. The basis for global spatial models is the spatial relationship between observations defined in a spatial weights matrix $\boldsymbol{W}$. To get $\boldsymbol{W}$, we start with a binary matrix $\boldsymbol{B}$ indicating the neighborhood of locations i and j by setting the matrix element:

$$b_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are neighbors and } i \neq j \\ 0 & \text{else} \end{cases} \tag{2}$$

The definition of the neighborhood of locations can be achieved in a lot of different ways. In this paper two locations are set as neighbors when they have contiguous boundaries, meaning that they share at least two distinct pairs of coordinates. An overview of standard and alternative methods of constructing $\boldsymbol{W}$ is given by Harris et al. (2011). We use the most common method of row-standardization, where every matrix element b_{ij} is divided by the respective sum of its row:

$$w_{ij} = \frac{b_{ij}}{\sum_{j'=1}^{n} b_{ij'}} \tag{3}$$

with w_{ij} being the elements of $\boldsymbol{W}$ and n being the number of observations (locations). Having built the spatial weights matrix $\boldsymbol{W}$, we are now able to use spatial models that can deal with spatial dependencies. There are two ways to achieve this: Either a spatially lagged dependent variable is used (spatial lag model) or

[1] Other segments of revenue are single tickets and student passes for example.

[2] In a competitive context, individuals might generate dissimilar (negatively correlated) outcomes.

spatial dependence is included in the error term (spatial error model). It is also possible to extend a spatial lag model by a set of spatially lagged independent variables resulting in a spatial Durbin model.

Spatial Lag Model: The spatial lag model (SLM) incorporates spatial autocorrelation by using a spatially lagged dependent variable as independent variable. It is based on a global linear regression model (GM) of the form:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{4}$$

For n observations and p parameters, $\boldsymbol{y}$ is an n by 1 vector of dependent observations, $\boldsymbol{X}$ is an n times p matrix of independent explanatory variables with the elements of the first column set to 1, and $\boldsymbol{\beta}$ is a p by 1 vector of respective coefficients. $\boldsymbol{\epsilon}$ is an n by 1 vector of independent identically normal distributed error terms with zero mean and constant variance σ^2 (i.i.d. normal). Using the spatial weights matrix $\boldsymbol{W}$, the spatial lag for y at location i is then built as

$$[\boldsymbol{Wy}]_i = \sum_{j=1}^{n} w_{ij} \cdot y_j \tag{5}$$

and added to the model. The spatial lag model is then expressed as

$$\boldsymbol{y} = \rho \boldsymbol{Wy} + \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{6}$$

where ρ represents a spatial autoregressive coefficient. As Anselin 1988, p. 85 states, estimating the parameters for the SLM with an ordinary least squares (OLS) approach leads to biased and inconsistent results, so maximum likelihood estimation is used to avoid these problems (Anselin, 1988, pp. 60–65).

Spatial Error Model The spatial error model (SEM) is also based on the global regression model of Eq. (4), but here the spatial dependence is connected to the error term:

$$\boldsymbol{y} = \boldsymbol{X\beta} + (\boldsymbol{I} - \lambda \boldsymbol{W})^{-1} \boldsymbol{\epsilon} \tag{7}$$

where λ is the coefficient of the spatially lagged autoregressive errors (Bivand et al., 2008, pp. 289–296). According to Anselin (1988), applying least squares to estimate the SEM leads to unbiased, but inefficient parameter estimates. Therefore, as for the SLM, maximum likelihood estimation is adapted.

Spatial Durbin Model: The spatial Durbin model is basically a spatial lag model with an additional set of spatially lagged independent variables. The term $\boldsymbol{WX}$ adds average neighboring observation values of the independent variables to the equation. γ is a $(p-1)$ by 1 vector (the intercept is not lagged) measuring the marginal impact of the independent variables from neighboring observations on $\boldsymbol{y}$. The spatial Durbin model (SDM) can be written as (Beer and Riedl, 2012)

$$\boldsymbol{y} = \rho \boldsymbol{Wy} + \boldsymbol{X\beta} + \boldsymbol{WX}\gamma + \boldsymbol{\epsilon} \tag{8}$$

The same issues as with the spatial lag model arise here and therefore we use maximum likelihood for the estimation.

2.2. *Geographically weighted regression*

While the global spatial models presented so far are successful in dealing with spatial dependence, they lack the ability to handle spatial drift. Their statistical outputs are of global nature, meaning that the relationships uncovered by the regression are assumed to be the same at every point of the study area. Since in most cases spatial data is in fact varying throughout the region examined, valuable information can be missed and existing relationships are not correctly reflected. This kind of divergence can be random or inherent (Huang and Leung, 2002).

To identify and handle spatial drift, we use a geographically weighted regression (GWR). In this section we refer to Fotheringham et al. (1997) as not stated otherwise. GWR permits local relationships to exist and therefore making them observable. Starting point is the standard global regression model of Eq. (4). To allow the estimation of local parameters, GWR extends this basic model so that the parameters can be estimated using a weighted least squares approach. The GWR model is expressed as

$$\boldsymbol{y} = (\boldsymbol{\beta} \otimes \boldsymbol{X})\boldsymbol{1} + \boldsymbol{\epsilon} \tag{9}$$

where $\otimes$ is a logical multiplication operator in which each element of $\boldsymbol{\beta}$ is multiplied by the corresponding element of $\boldsymbol{X}$. If there are n data points and p independent variables (including the intercept), both $\boldsymbol{\beta}$ and $\boldsymbol{X}$ will have dimensions $n \cdot p$ and $\boldsymbol{1}$ is a $p \cdot 1$ vector of 1s. The matrix $\boldsymbol{\beta}$ now consists of n sets of local parameters. Each row of $\boldsymbol{\beta}$ is represented by location i. The parameters of each row i are estimated by

$$\hat{\boldsymbol{\beta}}_i = (\boldsymbol{X}'\tilde{\boldsymbol{W}}_i\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{W}}_i\boldsymbol{y} \tag{10}$$

where the weighting matrix $\tilde{\boldsymbol{W}}_i$ is an n by n matrix whose off-diagonal elements are zero and whose diagonal elements are the weights of each observation, i.e., $\tilde{\boldsymbol{W}}_i = \mathrm{diag}(\tilde{w}_{i1}, \tilde{w}_{i2} \ldots \tilde{w}_{in})$. The estimation process is that of fitting a spatial kernel to the data. As the main part of this spatial kernel, the weighting matrix $\tilde{\boldsymbol{W}}_i$ defines the relationship between a regression point at location i and the data points at location j surrounding it. The premise in spatial analysis states that closer data points have greater impact on the regression point than those farther away. This premise is reflected in $\tilde{\boldsymbol{W}}_i$, where the data based on observations closer to i is weighted more than data based on observations farther from i. The specification of the elements of $\tilde{\boldsymbol{W}}_i$ also needs to account for the fact that most spatial processes are continuous ("drift"). A lot of techniques can achieve continuity, but the most commonly used method is the Gaussian function of the form:

$$\tilde{w}_{ij} = \exp\left(\frac{-q_{ij}^2}{r_i^2}\right), \quad i,j = 1,2 \ldots n \tag{11}$$

where q_{ij} is the distance between regression point i and data point j, and r_i describes a bandwidth at regression point i. With increasing distance between regression point i and data point j, the weight of these data points decreases corresponding to a Gaussian curve. If i and j share the same location, the weight of the data at this location will be one. Then again, the weight for data points j will practically fall to zero if they exceed a certain distance, so that the information at these locations is ignored in the parameter estimation of i.

The spatial kernels also need to incorporate the fact, that data points may not be evenly distributed over an examined area. In regions where the density of data points is high, changes in the relationships over relatively small distances might be missed with fixed kernels that are larger than needed, leading to possibly biased estimates. In regions where data are scarce, the number of data points will be too small for fixed kernels, leading to unreliable estimates. To avoid these problems, spatially varying kernels are used. By using a fixed number of observations (nearest neighbors) for every local estimation, the bandwidth of the spatial kernels varies over space according to the density of the data (see Fig. 1). To determine the optimal number of nearest neighbors and the corresponding bandwidths r_i, we use the cross-validation approach defined as

$$\min \mathrm{CV} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{y}_{\neq i}(r_i)]^2}. \tag{12}$$

The cross-validation score (CV) is minimized by computing several GWR with different bandwidths r_i. The optimal bandwidth is chosen where (12) is minimized. Using a bandwidth r_i, $\hat{y}_{\neq i}(r_i)$ is the fitted value of y_i without the observations at point i. This is necessary to avoid the cross-validation score turning to zero by

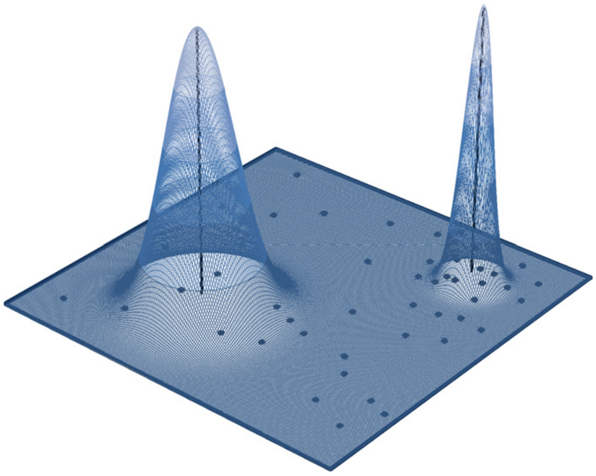


Fig. 1. Adaptive kernels: The kernels (and the respective r_i) are larger in regions with scarce data and smaller in regions with dense data. Each point on the surface represents one data point.

using a bandwidth so small, that the fitted values become the actual values. This would lead to undefined parameters and a wild fluctuation of estimates over the study area. By omitting i for the fitted values in the cross-validation process (**not** the final parameter estimation!), a small bandwidth only accounts for points near to i, but not i itself.

Following Wheeler and Paez (2010) GWR has matured to a well accepted methodology today. However, as they state GWR has its shortcomings as well. One issue is that GWR does not propose a base model for the source of the spatial drift and is thus more appropriately seen as a heuristic approach. The GWR model yields local estimates for every observation (location) of every independent variable. While the high number of parameters increases the amount of information about relationships, it also makes the GWR results hard to interpret (particularly in terms of management strategies) as stated by Mittal et al. (2004). They propose to search manually for regions of homogeneous model parameters in order to implement management strategies. As we aim to account for both, spatial dependency and spatial drift, we bear on the results of the geographically weighted regression and define spatial clusters of as homogeneous as possible model parameters by a cluster analysis. In contrast to Mittal et al. (2004) this procedure is particularly appealing, because the definition of the spatial clusters is not arbitrary or manual but data driven and empirical. Then, the spatial clusters are incorporated as additional variables in the global spatial models, resulting in "semi-global" spatial models: *spatially heterogeneous models*.

2.3. A spatially heterogeneous model

Remember that for a given independent variable the GWR yields coefficients for each observation as shown in (10). Now, in order to exploit the information about the spatial drift and to use this information within a global spatial model, we first have to define spatial clusters of as homogeneous as possible coefficient estimates based on the results of the GWR. For each independent variable we aggregate the spatial drift of the numerous GWR coefficients into few spatial clusters (of locations). The coefficients of locations that are assigned to the same cluster need to be preferably similar to each other, while they should be as different as possible to coefficients of locations assigned to other clusters. For this task, the k-means algorithm is an appealing clustering method (Lloyd, 1982). Given an independent variable $m=1\dots p$ the algorithm seeks to partition the n coefficient estimates β_{im} with $i=1\dots n$ into k_m disjoint clusters $c_{1_m}, c_{2_m}\dots c_{l_m}\dots c_{k_m}$. For each independent variable m, the optimization criterion is to minimize over all k_m clusters the sum of squared distances between each coefficient β_{im} for which the corresponding observation i is assigned to a given cluster c_{l_m} and the corresponding cluster centroid. Since the k-means method is heuristic, we cannot know whether we have found the cluster solution that represents a global optimum. To reduce this uncertainty we employ the most commonly used hierarchical agglomerative clustering methods, which we then compare to our k-means solutions. In general, the hierarchical agglomerative methods start by treating each observation i as a cluster, which are then merged step by step to form larger clusters. To determine which clusters should be merged, the (dis)similarity between them is computed (Cardoso and de Carvalho, 2009; Halkidi et al., 2001).

In the way the spatial clusters are derived from the GWR results we expect the relationship between dependent and independent variables to be constant for observations within a cluster but this relationship is likely to vary between clusters. Based on the spatial clusters c_{l_m}, we are now able to derive artificial dummy variables

$$z_{iml} = \begin{cases} 1 & \text{if location } i \text{ is assigned to cluster } c_{l_m} \\ 0 & \text{else} \end{cases} \quad \forall\, i,m,\ l=1\dots k_m \tag{13}$$

with $\sum_{l=1}^{k_m} z_{iml}=1$ given i and m. In order to define spatially heterogeneous models, we replace $\boldsymbol{X}$ of (4), (6), (7) and (8) by the matrix

$$\boldsymbol{H} = \begin{pmatrix} x_{10}z_{101} & x_{10}z_{102} & \cdots & x_{10}z_{10l_0} & x_{11}z_{111} & \cdots & x_{11}z_{11l_1} & \cdots & x_{1p}z_{1pl_p} \\ x_{20}z_{201} & x_{20}z_{202} & \cdots & x_{20}z_{20l_0} & x_{21}z_{211} & \cdots & x_{21}z_{21l_1} & \cdots & x_{2p}z_{2pl_p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \dots \\ x_{n0}z_{n01} & x_{n0}z_{n02} & \cdots & x_{n0}z_{n0l_0} & x_{n1}z_{n11} & \cdots & x_{n1}z_{n1l_1} & \cdots & x_{np}z_{npl_p} \end{pmatrix} \tag{14}$$

and the vector $\boldsymbol{\beta}$ is replaced by

$$\tilde{\boldsymbol{\beta}} = \begin{pmatrix} \beta_{01_0} \\ \beta_{02_0} \\ \vdots \\ \beta_{0l_0} \\ \beta_{11_1} \\ \vdots \\ \beta_{1l_1} \\ \vdots \\ \beta_{pl_p} \end{pmatrix} \tag{15}$$

With β_{ml_m} as the coefficient of independent variable m corresponding to spatial cluster c_{l_m}. Since each observation (location) is assigned to exactly one spatial cluster, we are able to measure different relationships according to the spatial clusters. Note, we replace $\boldsymbol{X}\gamma$ of (8), correspondingly.

3. Research setting and data

In this paper we employ the annual seasonal ticket revenue of a public transport company located in the city of Dresden, Germany. Dresden is the capital of the state of Saxony located in the far east of Germany. It spans an area of approximately 328 square kilometers and has a population of nearly 490,000 in 2005. Since annual seasonal tickets are personalized, the locations of the customers are known. We obtain aggregated revenue data on the scale of the 400 statistical districts of the city of Dresden,

2005. Note that revenue includes the seasonal tickets for pupils, while student tickets are excluded. Due to inconsistent or missing data this analysis is based on the data of 391 districts (see Fig. 2).

Public transport seasonal ticket sales – and thus revenue – is determined by the demand for public transport services. This demand in turn is mainly influenced by the fare (or the price) a customer has to pay and the service quality (travel-time, for example). Now, in our cross-sectional data on public transport seasonal ticket revenue, the price for a seasonal ticket does not differ over space. Hence, for our revenue response function only service quality variables are of interest.

Since we lack information about the trips of the customers (and the related travel-times) we rely on the headway: we assume that a higher frequency of departures from public transport stops per period improves the attractiveness of public transport services, leading to a positive effect on (seasonal) ticket revenue. Therefore, we consider the average and the total number of departures of all stops located in a statistical district per year. Moreover, we assume that the accessibility of the services has a positive effect on seasonal ticket revenue as well: We consider a district i and an address located in this district denoted by a_i with A_i as the total number of addresses located in i. Further, we consider the average Euclidean distance from a_i to the three most proximate stops denoted by $\mathfrak{d}_{a_i}$. Then, we define for each district i the accessibility variable *average distance to stops*

$$\mathrm{AvgDist2Stops}_i = \frac{\sum_{a_i} \mathfrak{d}_{a_i}}{A_i} \tag{16}$$

Besides the service variables we assume that – based on the work of Hawkins et al. (1981) and Parker and Tavassoli (2000) – several socio-demographic and socio-economic as well as land-use variables do have an impact on the seasonal ticket revenue.

- Population and population density have positive influence on the seasonal ticket revenue because the more people live in a given district the higher the potential for public transport demand. In areas with high population density the relative attractiveness of public transport services might increase due to a higher propensity of traffic congestion and lack of parking space.
- Age structure: different age groups are assumed to have different mobility patterns and thus they are expected to have different demand for public transport services. For example elderly (and retired) persons might not make so many trips a day as a young (and employed) person. And hence annual seasonal tickets might not be efficient.
- We expect the employment status to have impact on the seasonal ticket revenue as well. On the one hand, unemployed persons are less able to afford a car and thus are more depended on public transport service. On the other hand, these persons might not afford an annual seasonal ticket. In contrast, employed persons are expected to be able to afford both, a car and an annual seasonal ticket. Altogether, we cannot make a distinct assumption about the expected sign. We additionally control for the percentage of students located in a district as there are special passes for students. The corresponding revenue is not included in our data. Therefore, we assume a negative impact of a high number of students on revenue. We expect the higher the number of cars per capita within a district the more likely it is for the persons located in this district to own a car. Car ownership is assumed to reduce the demand for public transport and hence we assume a negative impact on revenue.
- Because the demand for public transportation is a derived demand, we assume that land-use patterns influence revenue. The average area of a district is 0.82 sq km. So we assume that activities within a district are accessible by non-motorized transport modes. For example, a high amount of schools in a district yields short travel distances for pupils, making it more likely for them to travel on foot or by bicycle, and not by public transport. Therefore, we assume that a higher amount of schools lowers seasonal ticket revenue. Due to its high concentration of jobs, we deem the central business district (CBD) a good indicator for the influence of work on transport demand. With very low or very high distances to the CBD,

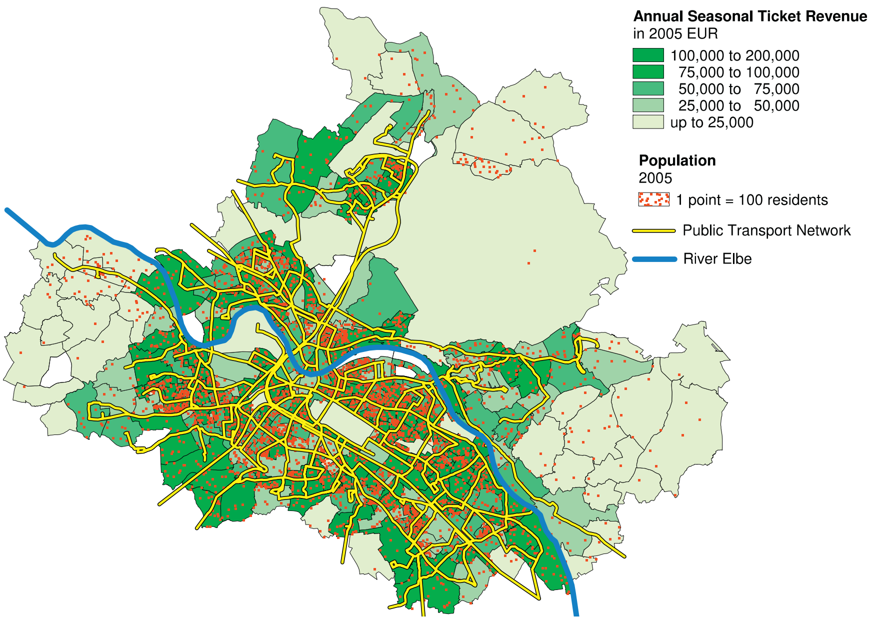


Fig. 2. The city of Dresden, Germany.

Table 1
Overview of the data used in regression models. The maximum value of the percentage of students exceeds 1 due to a city district with a high concentration of dormitories. Since many students only have a secondary residence at their study place, they are not registered as part of the population. This leads to an inadequacy between the amount of students and the population within that district.

Nr.	Variable	Description	Min.	Q_1	Median	Mean	Q_3	Max.
	TotSTRev	Total annual seasonal ticket revenue per year (in EUR)	106	44,840	71,770	71,940	97,670	195,100
1	TotPop	Total population	9	881.5	1247.0	1302.0	1690.0	3185.0
2	PopSqKM	Population per square kilometer	0.559	1631	5065	6480	9802	26900
3	PctAge6	Percentage of people up to 6 years	0.000	0.071	0.086	0.106	0.118	0.896
4	PctAge6to17	Percentage of people between 6 and 17 years	0.000	0.071	0.092	0.091	0.108	0.219
5	PctAge18to24	Percentage of people between 18 and 24 years	0.000	0.073	0.092	0.097	0.114	0.281
6	PctAge25to44	Percentage of people between 25 and 44 years	0.000	0.220	0.272	0.271	0.316	0.598
7	PctAge45to59	Percentage of people between 45 and 59 years	0.000	0.140	0.172	0.175	0.208	0.349
8	PctAge60to64	Percentage of people between 60 and 64 years	0.000	0.050	0.069	0.072	0.090	0.224
9	PctEmpl	Percentage of employed people	0.000	0.398	0.439	0.432	0.473	0.871
10	PctUnempl	Percentage of unemployed people	0.000	0.081	0.107	0.114	0.141	0.417
11	PctStudents	Percentage of students	0.000	0.024	0.026	0.054	0.054	1.104
12	CarsPC	Number of cars per capita	0.011	0.357	0.419	0.488	0.481	9.548
13	TotSchools	Total number of schools	0	0	0	0.432	1	5
14	Dist2CBD	Distance to the central business district (in meters)	369.5	2957	4670	5163	6861	14,660
15	WorkersPerJob	Average number of workers per job	0.000	0.732	1.880	2.678	3.729	40.000
16	FloorspMperSqKM	Average amount of sales floor per square kilometer (in meters)	0	0	1007	6302	5190	250,600
17	TotDepartStop	Total number of departures from stops per year	0	0	68,700	123,600	179,100	1,054,000
18	AvgDist2Stops	Average distance to the next three stops (in meters)	152.3	259.4	308.8	331.0	364.7	1263.0

public transport is either unnecessary or inefficient. Otherwise, on medium distances, we assume that public transport is in fact an efficient transport mode. We therefore expect mixed, location dependent influences of the distance to the CBD on revenue. We further assume that a lower amount of jobs in a district increases the propensity of its residents to travel to other places in order to work. Hence, a higher ratio of workers per job should increase transport demand and thus the sales and revenues of public transport seasonal tickets. Like work, consumption drives the transport demand of people. Therefore, we suppose that a high extent of sales-floor-space in a district reduces the need to travel elsewhere in order to fulfill consumer needs, and thus lowering public transport seasonal ticket revenue in that district.

All land-use, socio-economic and socio-demographic data are obtained from the local statistics authority of the city of Dresden. All variables used in our analyses can be found in Table 1

4. Results

In this section we first discuss the process of model building in order to find an adequate model specification (Section 4.1). Based on this model we then discuss empirical findings in Section 4.2 and managerial insights including the specification of the revenue response function in Section 4.3.

4.1. Model building

Based on the variables shown in Table 1 and some previously performed model estimations we propose the following global (spatial) model[3]:

$$\begin{aligned}\log(\text{TotSTRev}) = {} & \beta_0 + \beta_1 \cdot \text{TotPop} + \beta_2 \cdot \text{PopSqKM} \\ & + \beta_3 \cdot \text{PctAge6} + \beta_4 \cdot \text{PctAge6to17} \\ & + \beta_5 \cdot \text{PctAge18to24} + \beta_6 \cdot \text{PctAge25to44} \\ & + \beta_7 \cdot \text{PctAge45to59} + \beta_8 \cdot \text{PctAge60to64} \\ & + \beta_9 \cdot \text{PctEmpl} + \beta_{10} \cdot \text{PctUnempl} \\ & + \beta_{11} \cdot \text{PctStudents} + \beta_{12} \cdot \text{CarsPC} \\ & + \beta_{13} \cdot \text{TotSchools} + \beta_{14} \cdot \text{Dist2CBD} + \beta_{15} \cdot \text{WorkersPerJob} \\ & + \beta_{16} \cdot \log(\text{FloorspMperSqKM}+1) + \beta_{17} \cdot \log(\text{TotDepartStop}+1) \\ & + \beta_{18} \cdot \text{AvgDist2Stops}\end{aligned}$$

The estimation results of the GM, SEM, SLM and the SDM are given in Table 2. A tentative interpretation of the results roughly confirms our hypotheses of Section 3. There is a positive effect of population and population density on revenue. Socio-economic and socio-demographic variables have statistical significant impact on the seasonal ticket revenue as well. The land-use variables show unexpected effects while the service variables confirm our assumptions. More important, the assumption that spatial dependencies underlie the relationship between dependent and independent variables is impressively verified: The coefficients λ of the SEM and ρ of the SLM and SDM as well as some $\hat{\gamma}_m$ are statistically significant. Moreover, the Moran's I for the GM shows that the residuals are spatially correlated.[4] This statistic tells us, that in our case only SEM and SDM treat spatial dependencies adequately. AICc points the SDM as the best model so far. While there is no issue with (perfect) multicollinearity (all variance inflation factors (VIF) are less than 5), we witness that heteroscedasticity seems to lead to inefficient estimates (Breusch–Pagan test).

Since we have shown that spatial dependencies have an influence on our data and the corresponding models, we are now interested in whether spatial drift takes place as well. Therefore, we consider the results of the GWR of (9) shown in Table 3. In order to test whether the GWR model is more appropriate compared to a global linear model (GM), Leung et al. (2000) have constructed several pertinent statistics. Here, we employ two selected test statistics:

1. $\mathcal{F}_1$-test: The null hypothesis of this test statistic is that there is no significant difference between a given GM and the GWR model under consideration in describing the coherences of interest. Basically this test consists of the ratio of the residual

[3] We omit the subscript i for convenience reasons.

[4] Moran's I is the standard test statistic to identify spatial autocorrelation in residuals (Anselin, 1988).

Table 2
Estimation results for the global (spatial) models. Standard errors are displayed in brackets. The estimates and standard errors of the variables TotPop, PopSqKM, Dist2CBD and AvgDist2Stops are multiplied by 1000 for visualization purpose. VIF < 5 for all parameters. The significance levels are: ′0.1, * 0.05, ** 0.01,*** 0.001.

Variable m	GM	SEM	SLM	SDM	
	Estimate $\hat{\beta}_m$	Estimate $\hat{\beta}_m$	Estimate $\hat{\beta}_m$	Estimate $\hat{\beta}_m$	Estimate $\hat{\gamma}_m$
(Intercept)	9.901***	10.148***	6.261***	5.666***	–
	(0.421)	(0.414)	(0.587)	(0.915)	–
TotPop	0.839***	0.809***	0.863***	0.786***	−0.354 *
	(0.067)	(0.057)	(0.060)	(0.058)	(0.139)
PopSqKM	0.029***	0.029***	0.022**	0.028***	−0.012
	(0.008)	(0.008)	(0.007)	(0.008)	(0.016)
PctAge6	−2.219**	−1.765**	−1.970**	−2.326***	−2.079′
	(0.702)	(0.628)	(0.630)	(0.632)	(1.156)
PctAge6to17	2.036	2.555*	2.160*	1.759	−2.939
	(1.307)	(1.145)	(1.173)	(1.134)	(2.530)
PctAge18to24	−1.102	−0.542	−0.596	−0.003	−2.720
	(1.397)	(1.284)	(1.255)	(1.291)	(2.653)
PctAge25to44	−0.597	0.090	−0.350	−0.194	−2.813*
	(0.664)	(0.654)	(0.595)	(0.685)	(1.247)
PctAge45to59	3.504***	3.746***	3.786***	3.007**	−4.208*
	(0.994)	(0.935)	(0.892)	(0.942)	(1.828)
PctAge60to64	2.548′	3.589**	2.493′	3.282*	−6.737*
	(1.493)	(1.389)	(1.341)	(1.374)	(2.692)
PctEmpl	−0.055	−1.337*	−1.123*	−1.235*	5.624***
	(0.628)	(0.563)	(0.571)	(0.567)	(1.182)
PctUnempl	1.406′	1.339′	−1.123	−1.235	5.624
	(0.823)	(0.784)	(0.571)	(0.567)	(1.182)
PctStudents	−0.885′	−1.237**	−1.102**	−1.102**	2.120*
	(0.459)	(0.406)	(0.412)	(0.401)	(0.910)
CarsPC	−0.306***	−0.342***	−0.327***	−0.364***	0.110
	(0.064)	(0.058)	(0.058)	(0.057)	(0.102)
TotSchools	0.051	0.033	0.040	0.038	0.103
	(0.042)	(0.035)	(0.038)	(0.038)	(0.090)
Dist2CBD	−0.115***	−0.128***	−0.072***	−0.402***	0.346***
	(0.016)	(0.022)	(0.015)	(0.094)	(0.100)
WorkersPerJob	−0.022′	−0.031**	−0.032**	−0.031***	0.029
	(0.012)	(0.010)	(0.011)	(0.011)	(0.018)
log(FloorspMperSqKM+1)	0.024*	0.015′	0.014	0.015′	0.029
	(0.010)	(0.008)	(0.009)	(0.008)	(0.019)
log(TotDepartStop+1)	0.023***	0.017**	0.020**	0.021***	0.026′
	(0.007)	(0.006)	(0.006)	(0.006)	(0.015)
AvgDist2Stops	−1.334***	−0.964**	−0.848**	−0.621*	0.507
	(0.329)	(0.305)	(0.299)	(0.308)	(0.659)
λ	–	0.531***	–	–	
ρ	–	–	0.342***	0.401***	
AICc	751.512	687.103	697.417	680.539	
n	391	391	391	391	
Moran's I statistic	0.276	−0.032	0.083	−0.022	
p-value	0.000	0.838	0.002	0.743	
Breusch–Pagan statistic	63.922	70.228	71.552	92.020	
p-value	0.000	0.000	0.000	0.000	

sum of squares of the GWR model and the GM weighted by the number of parameters to be estimated. The distribution of the test statistic $\mathcal{F}_1$ is approximated by an F-distribution.

2. $\mathcal{F}_3$-test: The null hypothesis of this test statistic is that all GWR coefficients for a given independent variable are equal. That is

$$\hat{\beta}_{1m} = \hat{\beta}_{2m} = \cdots = \hat{\beta}_{im} = \cdots = \hat{\beta}_{nm}, \quad \forall m = 1 \ldots p.$$

This test is based on the sample variance of the n estimated values of $\hat{\beta}_{im}$. Again, the distribution of $\mathcal{F}_3(k)$ is approximated by an F-distribution.

In general, the AICc of the GWR has a value of 657.348, which is lower than the AICc value of the GM (751.512). Since the $\mathcal{F}_1$ value is significant to the level of $p < 0.01$, H_0 is rejected and the $\mathcal{F}_1$-test confirms that the GWR outperforms the global model significantly. Concerning the estimates, we identify a remarkable variation of the parameter values. Particularly, the estimates of several parameters include negative and positive values (for example PctEmpl). This kind of spatial variation would not only be invisible in global models, it would also distort the results since negative and positive values may cancel each other out to some degree (compare the estimate of PctEmpl of GM in Table 2). If we take a look at the $\mathcal{F}_3$-test in Table 3, we see that the intercept and the coefficients of 14 out of 18 independent variables do in fact vary significantly over space. To get a better view of their spatial variation, we map these coefficients (see Fig. 4 in the appendix). In order to account for the spatial drift in our global spatial models we try to identify spatial clusters for these coefficients by the approach outlined in Section 2.3. The possible number of clusters ranges from 1 to 391. A single cluster would include all city districts and be equivalent to the global models we already have. Therefore, we set a minimum of two clusters. The possible maximum amount of 391 clusters (one cluster for every city district) would make the whole clustering process unnecessary, since our goal is to merge statistical districts, resulting in a spatially heterogeneous model. For each independent variable that shows significant spatial variation we start with two clusters

Table 3
Estimation results for the GWR. Standard errors are displayed in brackets. The estimates and standard errors of the variables TotPop, PopSqKM, Dist2CBD and AvgDist2Stops are multiplied by 1000 for a better display. $\mathcal{F}_1$ and $\mathcal{F}_3$ refer to the statistical tests of Leung et al. (2000). The significance levels are: ′0.1,*0.05,**0.01,***0.001. The adaptive quantile of 0.202 corresponds to 78 nearest neighbors for determining the optimal bandwidth.

Variable m	$\hat{\beta}_{im}$						
	Min	Q_1	Median	Mean	Q_3	Max	$\mathcal{F}_3$
(Intercept)	9.000	9.653	10.030	10.090	10.590	11.130	4.784***
	(0.363)	(0.426)	(0.457)	(0.457)	(0.478)	(0.599)	
TotPop	0.710	0.748	0.776	0.781	0.797	0.905	0.893
	(0.059)	(0.074)	(0.078)	(0.077)	(0.082)	(0.088)	
PopSqKM	0.004	0.013	0.018	0.018	0.022	0.031	1.334*
	(0.007)	(0.009)	(0.009)	(0.009)	(0.010)	(0.011)	
PctAge6	−4.748	−3.996	−3.427	−3.058	−2.115	−0.359	6.259***
	(0.608)	(0.714)	(0.821)	(0.802)	(0.873)	(1.002)	
PctAge6to17	−2.810	0.213	1.693	1.356	2.903	3.807	2.992***
	(1.153)	(1.457)	(1.586)	(1.552)	(1.658)	(1.781)	
PctAge18to24	−2.468	−1.204	−0.392	−0.413	0.415	1.418	1.031
	(1.214)	(1.442)	(1.547)	(1.565)	(1.669)	(2.037)	
PctAge25to44	−2.338	−1.510	−0.947	−0.288	0.829	3.373	9.027***
	(0.583)	(0.696)	(0.833)	(0.819)	(0.917)	(1.099)	
PctAge45to59	−0.203	1.275	2.155	2.848	3.917	7.876	8.001***
	(0.878)	(1.120)	(1.231)	(1.224)	(1.352)	(1.531)	
PctAge60to64	1.161	1.736	2.288	2.519	3.365	4.240	0.845
	(1.292)	(1.546)	(1.635)	(1.617)	(1.693)	(1.880)	
PctEmpl	−3.781	−0.568	0.337	0.113	1.311	2.746	8.397***
	(0.554)	(0.723)	(0.850)	(0.843)	(0.968)	(1.113)	
PctUnempl	−1.122	−0.019	1.330	1.139	2.272	3.161	4.509***
	(0.719)	(0.887)	(0.937)	(0.934)	(0.982)	(1.133)	
PctStudents	−1.401	−1.015	−0.830	−0.784	−0.582	−0.087	1.522**
	(0.395)	(0.442)	(0.458)	(0.465)	(0.479)	(0.603)	
CarsPC	−0.669	−0.474	−0.360	−0.378	−0.280	−0.195	6.648***
	(0.055)	(0.060)	(0.062)	(0.073)	(0.079)	(0.161)	
TotSchools	−0.001	0.021	0.033	0.034	0.046	0.069	0.370
	(0.037)	(0.044)	(0.047)	(0.047)	(0.050)	(0.059)	
Dist2CBD	−0.247	−0.121	−0.096	−0.111	−0.083	−0.066	7.067***
	(0.014)	(0.019)	(0.021)	(0.022)	(0.025)	(0.035)	
WorkersPerJob	−0.042	−0.024	−0.013	−0.011	−0.004	0.026	1.756**
	(0.011)	(0.015)	(0.020)	(0.018)	(0.022)	(0.024)	
log(FloorspMperSqKM+1)	0.008	0.012	0.015	0.018	0.020	0.041	1.421**
	(0.009)	(0.010)	(0.011)	(0.011)	(0.011)	(0.013)	
log(TotDepartStop+1)	0.007	0.009	0.011	0.014	0.020	0.033	2.110***
	(0.006)	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	
AvgDist2Stops	−1.663	−1.186	−0.787	−0.830	−0.527	0.220	1.696***
	(0.295)	(0.451)	(0.514)	(0.493)	(0.555)	(0.608)	
$\mathcal{F}_1$	0.745**						
AICc	657.348						
n	391						
Adaptive quantile	0.202						
Moran's I Statistic	0.224						
p-value	0.000						

and then we increase the number of clusters step-by-step. For each step we verify whether the model fit (AICc) of the corresponding models SH-GM, SH-SEM, SH-SLM, and SH-SDM improves or not. It turns out that the best number of clusters (in terms of model fit) is different for the considered independent variables. In general, more than ten clusters always yield poor model fits. Unfortunately, we observe strong multicollinearity issues with several variables such that we do allow the corresponding coefficients to vary spatially. Finally, the new model resulting from the described procedure contains 7 modified variables in order to account for spatial drift (PctAge6, PopSqKM, Dist2CBD, FloorsMperSqKM, TotDepartStop, CarsPC and WorkersPerJob). The other 7 variables with spatially varying GWR coefficients (Pct-Age6to17, PctAge25to44, PctAge45to59, PctEmpl, PctUnempl, PctStudents and Avg-Dist2Stops) either don't improve the model fit or lead to strong multicollinearity when modified. The different cluster solutions of the incorporated modified independent variables are visualized in Fig. 5 in the appendix. We see that the spatial variation of the coefficients seems to be adequately represented by the different clusters. We therefore assume that a large part of the spatial drift is incorporated in our spatially heterogeneous model.

4.2. *Empirical insights*

The estimation results for the spatially heterogeneous versions of the linear model (SH-GM), spatial error model (SH-SEM), spatial error model (SH-SEM) and spatial Durbin model (SH-SDM) can be found in Table 4. As expected, population and population density positively affect seasonal ticket revenue. However, the effect of population density diminishes with increasing distance to the city center. Particularly, the coefficient for cluster two (city center) is statistically insignificant. Concerning the demographic structure we witness (SH-SEM) that a high percentage of young children (PctAge6) has a negative impact on revenue - particularly in the eastern parts of the city of Dresden, while a high percentage of persons older than 6 years has a positive effect on revenue. Only PctAge18-24 shows a non-

Table 4
Estimation results for the spatially heterogeneous models. Standard errors are displayed in brackets. The estimates and standard errors of the (modified) variables TotPop, PopSqKM, Dist2CBD and AvgDist2Stops are multiplied by 1000, and the estimates and standard errors of the modified variable WorkersPerJob are multiplied by 10 for a better display. VIF < 5 for all parameters. The significance levels are: ′0.1, * 0.05, ** 0.01, *** 0.001. Coefficients displayed in bold font, are statistically significant different from the corresponding coefficient estimate of the GM of Table 2 (95% confidence level).

Variable m	SH-GM	SH-SEM	SH-SLM	SH-SDM	
	Estimate $\hat{\beta}_{ml}$	Estimate $\hat{\beta}_{ml}$	Estimate $\hat{\beta}_{ml}$	Estimate $\hat{\beta}_{ml}$	Estimate $\hat{\gamma}_{ml}$
(Intercept)	**10.974*****	**11.072*****	9.160***	8.768***	–
	(0.386)	(0.370)	(0.612)	(1.153)	–
TotPop	0.772***	0.783***	0.790***	0.768***	−0.261′
	(0.057)	(0.050)	(0.053)	(0.051)	(0.134)
PopSqKM cluster $l=1$	**0.053*****	0.050***	0.047***	0.041**	−0.019
	(0.012)	(0.012)	(0.011)	(0.013)	(0.030)
PopSqKM cluster $l=2$	**0.003**	**−0.002**	**0.003**	**−0.008**	0.030′
	(0.010)	(0.010)	(0.009)	(0.012)	(0.018)
PopSqKM cluster $l=3$	0.016′	0.016′	0.015′	0.022*	0.006
	(0.009)	(0.008)	(0.008)	(0.009)	(0.022)
PctAge6 cluster $l=1$	−1.089	**−0.074**	**−0.889**	**−0.743**	−6.670***
	(0.697)	(0.642)	(0.649)	(0.663)	(1.377)
PctAge6 cluster $l=2$	−3.096***	−2.722***	−2.909***	−3.325***	−2.071
	(0.728)	(0.684)	(0.678)	(0.721)	(1.556)
PctAge6to17	1.577	1.827′	1.578	1.962′	1.112
	(1.146)	(1.028)	(1.067)	(1.020)	(2.434)
PctAge18to24	−0.482	−1.566	−0.524	−0.653	5.404*
	(1.242)	(1.154)	(1.158)	(1.158)	(2.587)
PctAge25to44	0.421	**1.148***	**0.550**	0.346	−5.052***
	(0.601)	(0.583)	(0.559)	(0.628)	(1.264)
PctAge45to59	4.037***	4.402***	4.078***	3.813***	−4.859*
	(0.900)	(0.845)	(0.838)	(0.873)	(1.934)
PctAge60to64	2.408′	2.686*	2.623*	1.713	−3.614
	(1.334)	(1.234)	(1.243)	(1.230)	(2.826)
PctEmpl	**−1.988****	**−2.979*****	**−2.388*****	**−2.642*****	6.336***
	(0.611)	(0.550)	(0.571)	(0.565)	(1.323)
PctUnempl	**−0.302**	0.322	**−0.196**	**−0.134**	−3.889*
	(0.739)	(0.692)	(0.688)	(0.739)	(1.622)
PctStudents	−0.995*	−1.077**	−1.053**	−1.215**	0.530
	(0.414)	(0.367)	(0.385)	(0.379)	(0.944)
CarsPC cluster $l=1$	0.009	−0.003	−0.072	−0.132	0.039
	(0.251)	(0.278)	(0.235)	(0.343)	(0.471)
CarsPC cluster $l=2$	−0.162	−0.254	−0.337	−0.062	0.217
	(0.286)	(0.321)	(0.269)	(0.433)	(0.591)
CarsPC cluster $l=3$	−0.053	**0.037**	−0.052	**0.015**	−0.229
	(0.134)	(0.123)	(0.125)	(0.122)	(0.248)
CarsPC cluster $l=4$	**−0.780*****	**−0.909*****	**−0.763*****	**−1.024*****	0.806**
	(0.132)	(0.125)	(0.124)	(0.130)	(0.273)
CarsPC cluster $l=5$	**−0.415*****	**−0.410*****	**−0.417*****	**−0.418*****	−0.020
	(0.060)	(0.054)	(0.056)	(0.053)	(0.114)
CarsPC cluster $l=6$	−0.369*	−0.379*	−0.387*	−0.519**	−0.154
	(0.182)	(0.182)	(0.169)	(0.191)	(0.330)
TotSchools	0.036	0.023	0.033	0.040	0.164′
	(0.036)	(0.031)	(0.034)	(0.034)	(0.087)
Dist2CBD cluster $l=1$	−0.124***	−0.124***	−0.099***	**−0.340*****	0.225*
	(0.018)	(0.021)	(0.018)	(0.099)	(0.110)
Dist2CBD cluster $l=2$	**−0.165*****	**−0.163*****	−0.133***	**−0.386*****	0.245*
	(0.020)	(0.023)	(0.021)	(0.087)	(0.101)
Dist2CBD cluster $l=3$	**−0.360*****	**−0.379*****	**−0.309*****	**−0.538*****	0.289**
	(0.025)	(0.030)	(0.027)	(0.096)	(0.109)
WorkersPerJob cluster $l=1$	−0.086	−0.125	0.117	−0.375	−0.405
	(0.297)	(0.282)	(0.277)	(0.286)	(0.613)
WorkersPerJob cluster $l=2$	**−1.632***	**−1.680***	**−1.657****	−1.341′	0.857
	(0.682)	(0.660)	(0.636)	(0.695)	(1.405)
WorkersPerJob cluster $l=3$	−0.001	0.155	−0.024	0.294	−1.015
	(0.252)	(0.246)	(0.235)	(0.264)	(0.641)
WorkersPerJob cluster $l=4$	−0.040	0.106	−0.110	−0.185	−0.583
	(0.286)	(0.269)	(0.268)	(0.278)	(0.617)
WorkersPerJob cluster $l=5$	**−0.651*****	**−0.624*****	**−0.667*****	**−0.660*****	−0.089
	(0.138)	(0.120)	(0.129)	(0.127)	(0.239)
WorkersPerJob cluster $l=6$	−0.100	−0.036	−0.159	−0.124	−0.219
	(0.195)	(0.179)	(0.183)	(0.186)	(0.449)
log(FloorspMperSqKM+1) cluster $l=1$	**−0.019**	**−0.026**′	**−0.022**	**−0.012**	0.038
	(0.015)	(0.014)	(0.014)	(0.015)	(0.034)
log(FloorspMperSqKM+1) cluster $l=2$	0.057**	0.051**	0.046**	0.058**	0.058
	(0.019)	(0.017)	(0.018)	(0.019)	(0.046)
log(FloorspMperSqKM+1) cluster $l=3$	0.014	0.012	0.012	0.006	0.007
	(0.009)	(0.008)	(0.009)	(0.008)	(0.020)
log(TotDepartStop+1) cluster $l=1$	0.028**	0.023**	0.025**	0.029**	0.025

Table 4 (*continued*)

Variable m	SH-GM	SH-SEM	SH-SLM	SH-SDM	
	Estimate $\hat{\beta}_{ml}$	Estimate $\hat{\beta}_{ml}$	Estimate $\hat{\beta}_{ml}$	Estimate $\hat{\beta}_{ml}$	Estimate $\hat{\gamma}_{ml}$
	(0.009)	(0.008)	(0.008)	(0.009)	(0.018)
log(TotDepartStop+1) cluster l=2	**0.063*****	**0.075*****	**0.058*****	**0.072*****	−0.073*
	(0.015)	(0.014)	(0.014)	(0.015)	(0.030)
log(TotDepartStop+1) cluster l=3	0.030**	0.016′	0.026**	0.017′	0.067***
	(0.010)	(0.009)	(0.009)	(0.010)	(0.020)
log(TotDepartStop+1) cluster l=4	**0.003**	**0.005**	**0.004**	**0.008**	0.003
	(0.007)	(0.006)	(0.006)	(0.006)	(0.017)
AvgDist2Stops	−0.883**	−**0.807****	−**0.766****	−**0.613***	0.641
	(0.295)	(0.271)	(0.276)	(0.273)	(0.666)
$\hat{\lambda}$	–	0.429***	–	–	
ρ	–	–	0.162***	0.202**	
AICc	634.705	605.638	624.365	638.520	
n	391	391	391	393	
Moran's I Statistic	0.169	−0.004	0.102	−0.007	
p-value	0.000	0.519	0.000	0.552	
Breusch–Pagan Statistic	133.827	132.448	134.813	192.318	
p-value	0.000	0.000	0.000	0.000	

significant negative estimate. Obviously, persons aged between 45 and 59 are a valuable target group for annual seasonal tickets. One possible explanation for this might be, that the car ownership levels in Dresden in the late 1970s—when most of these persons would have acquired their driving licenses - were comparably low (Heitland, 2007). Hence, these persons are used to using public transportation for a large part of their lives.

As expected, the percentage of students decreases seasonal ticket revenue. For the employment status we do not find a distinct relationship with revenue. Both, unemployed and employed persons seem to use other modes of transport or at least other tickets than an annual seasonal ticket. Unemployed persons might not be able to afford the ticket price, while employed persons might use public transport only occasionally and thus an annual seasonal ticket is not efficient. Of course, the number of cars per capita (CarsPC) has a negative influence on annual seasonal ticket revenue. This effect is most pronounced in the city center as well as the northern and the southern districts. Concerning the outskirts this result is appealing: it might be difficult for persons located in the outskirts to access their specific trip destination using public transport. Therefore, these persons are more dependent on cars.

While the number of schools (TotSchools) has no significant effect, the negative impact of the distance to the CBD (Dist2CBD) increases with increasing distance to the city center. That is, each additional meter lowers revenue more in the outskirts than in inner city districts. Although there are some positive (non-significant) estimates for the number of workers per job (WorkersPerJob), the significant estimates obtain negative signs. This is in contrast to our assumption. Because employed persons seem to use other ticket alternatives or even other modes of transport (see PctEmpl in Table 4) it is reasonable that an increase in the ratio workers per job results in a decrease of seasonal ticket revenue. While in the south-eastern and the north-eastern parts of Dresden the effect of the floorspace (FloorspMperSqKM) is negative as expected according to our hypothesis, we observe an obverse effect in the far east. Maybe this is an artifact due to the location of the corresponding districts at the borders of the study region. Moreover, these districts are rarely accessible by public transport (see Fig. 2).

As assumed, the accessibility variable of (16) has a global – that is spatially constant – negative effect on seasonal ticket revenue: The farther away a public transport stop the less attractive is the use of public transport for frequent trips and hence revenue declines. Concerning the second service variable - the number of departures from all stops of a district (TotDepartStop)—has a positive impact on revenue. However, this effect is not constant in space: the most positive effect is obtained in districts in the far east. As Fig. 2 reveals these districts are badly covered by public transport services. The effect of TotDepartStop is, however, insignificant for central districts. These districts are quite good accessible by public transport. From this point it seems plausible, that improvements in service quality are most efficient (all other constant) in districts where service is relatively poor. At this point the importance of using a spatially heterogeneous model becomes obvious while the global models (see Table 2) overestimate the effect of TotDepartStop for central districts the effect for outskirts is underestimated.

The advantage of spatially heterogeneous models compared to the global (spatial) models is confirmed by the goodness-of-fit measure AICc (compare Tables 2 and 4). As with the global (spatial) models the SH-SEM and SH-SDM are the models that are able to cope with spatial autocorrelation adequately (Moran's I). Finally, we recognize that besides the spatially varying coefficient estimates some of the globally constant coefficient estimates are statistically different from the linear global model (GM) as well (see AvgDist2Stops, PctEmpl and PctAge25to44 in Table 4 for example). Based on the criteria goodness-of-fit and spatial-autocorrelation we propose model SH-SEM as the preferred model for the specification of the revenue response function.

4.3. Managerial insights

Now, in terms of a revenue response function the independent variables 1–16 of Table 1 and the intercept are constants from the managerial perspective, because managers are not able to immediately influence these variables by managerial decisions. Therefore, for each district i we define a constant value

$$RP_i = \sum_{m=0}^{16} \left(\sum_{l=1}^{k_m} \hat{\beta}_{ml} z_{iml} \right) x_{im} \tag{17}$$

as the *revenue potential* of district i. Then, we propose

$$\log(\mathrm{TotSTRev}_i) = RP_i + \left(\sum_{l=1}^{k_{17}} \hat{\beta}_{17,l} z_{i,17,l} \right) \log(\mathrm{TotDepartStop}_i + 1) + \hat{\beta}_{18} \mathrm{AvgDist2Stops}_i$$

as the log revenue response function. Now, by exponentiation we get the revenue response function

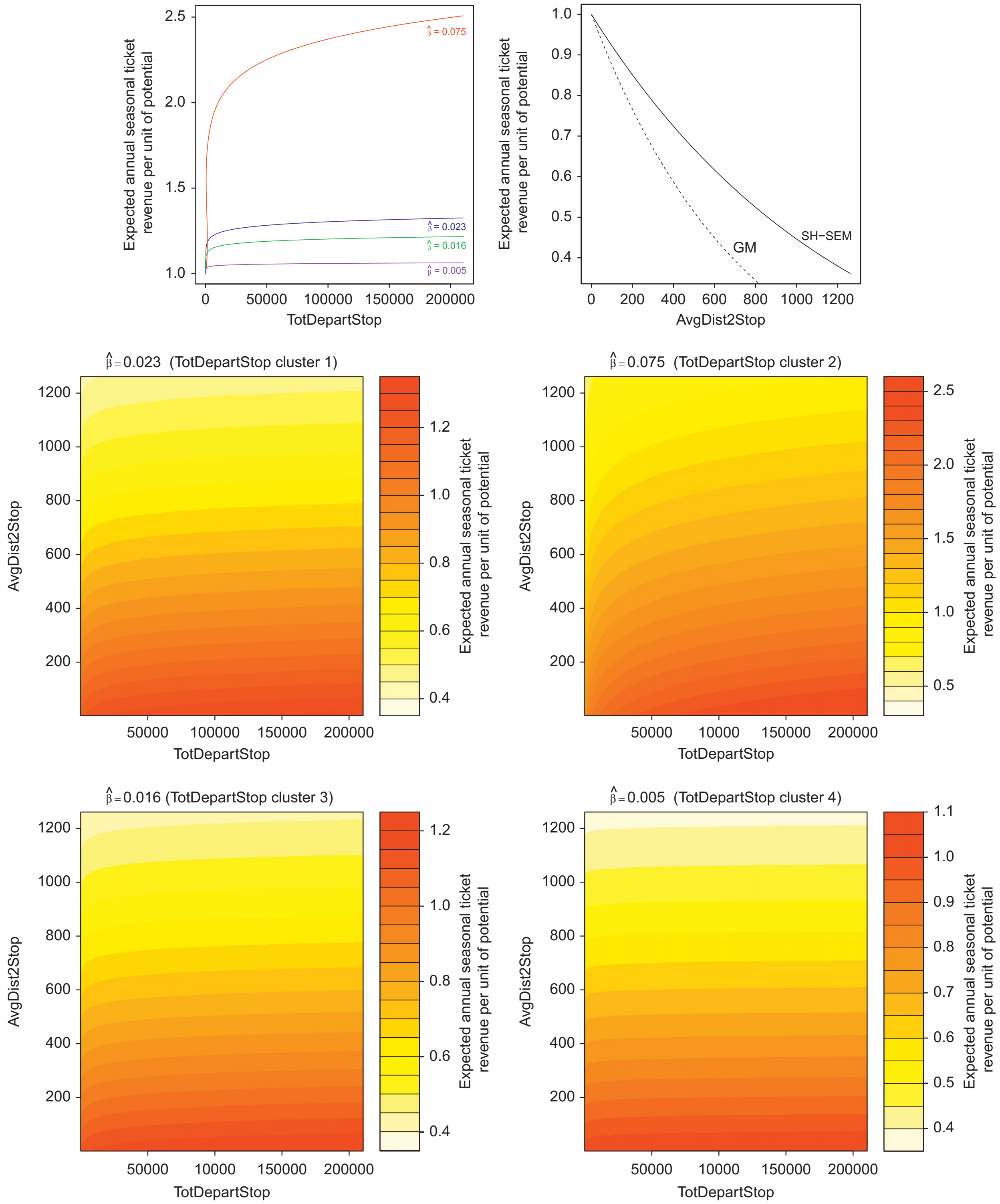


Fig. 3. Revenue response function: The upper two plots display the single effect of both service variables (AvgDist2Stops, TotDepartStop) on revenue, while the four contour plots show the simultaneous effect: for a given realization of TotDepartStop (abscissae) and AvdDist2Stop (ordinates) the color displays the expected annual seasonal ticket revenue response. Consider exemplary the contour plot for cluster 2 of TotDepartStop ($\hat{\beta}=0.075$): For AvgDist2Stop = 200 the expected revenue response to TotDepartStop (0–200k) ranges from light (1.3–1.5) to dark (2.0–2.3). In contrast for AvgDist2Stop = 1200 the expected revenue response is somewhat between 0.5 and 1.0.

$$\text{TotSTRev}_i = e^{RP_i}(\text{TotDepartStop}_i + 1)^{\sum_{l=1}^{k_{17}} \hat{\beta}_{17,l} z_{i,17,l}} e^{\hat{\beta}_{18}\text{AvgDist2Stops}_i} \tag{18}$$

Fig. 3 plots the revenue response function for TotDepartStop and AvgDist2Stops. The corresponding elasticities are

- $\mathcal{E}^{l=1}_{\text{TotSTRev,TotDepartStop}} = 0.023$
- $\mathcal{E}^{l=2}_{\text{TotSTRev,TotDepartStop}} = 0.075$
- $\mathcal{E}^{l=3}_{\text{TotSTRev,TotDepartStop}} = 0.016$
- $\mathcal{E}^{l=4}_{\text{TotSTRev,TotDepartStop}} = 0.005$
- $\mathcal{E}^{\min}_{\text{TotSTRev,AvgDist2Stops}} = -0.123$
- $\mathcal{E}^{\text{mean}}_{\text{TotSTRev,AvgDist2Stops}} = -0.267$
- $\mathcal{E}^{\max}_{\text{TotSTRev,AvgDist2Stops}} = -1.019$

From the revenue response function and the corresponding elasticities we get the following managerial insights

- Annual seasonal ticket revenue seems to be rather inelastic in terms of the service variables *total number of departures from stops per year* (TotDepartStop) and *average distance to the most proximate 3 stops* (AvgDist2Stops). This finding corresponds to studies on transport demand (see for example Goodwin and Williams, 1985 and Holmgren, 2007). However, there are certain regions in the service area, where revenue is less inelastic (see clusters one and two of TotDepartStop in Fig. 5).
- In order to increase revenues it would be efficient to improve the density of stops (AvgDist2Stops) rather to increase the number of departures (TotDepartStop).
- Managers should not rely on the key assumption in the traditional marketing science literature, that the behavior of a consumer is conditionally independent of the behavior of another consumer and that this behavior is spatially homogeneous. That is, using a standard linear global model (GM). As we can see, the results differ remarkably: The single (expected) effect of TotDepartStop of GM corresponds to the blue line in Fig. 3 ($\hat{\beta} = 0.023$). Using GM yields that for all districts of the service area this effect is the same. However, the effect for outskirts districts is expected to be much larger ($\hat{\beta} = 0.075$), while it is assumed to be smaller within the city center ($\hat{\beta} = 0.005$). The same is true concerning AvgDist2Stops.
- Managers are enabled to make spatially differentiated estimates of seasonal ticket revenue if changes in the urban structure occur: Imagine in certain districts there would be a decline in population, population density and a shift toward older people. Then, based on the modified revenue potential, managers are able to estimate the corresponding change in seasonal ticket revenue. Further, they might identify strategies to compensate an expected loss (partially) by an improvement in service.

5. Summary

In marketing science literature there is a growing awareness toward spatial effects which may underly the data (Bronnenberg, 2005; Mittal et al., 2004; Bradlow et al., 2005; and Sridhar et al., 2012 for example). Our analysis clearly shows that researchers and managers must not rely on the key assumption in the traditional marketing science literature, that the behavior of a consumer is conditionally independent of the behavior of another consumer and that this behavior is spatially homogeneous. Rather it is advisable to employ tailored models and methods to handle spatial effects like spatial dependency and spatial drift. In this paper we propose an empirical and data-driven approach that is able to cope with both effects simultaneously (spatially heterogeneous model). Our empirical findings demonstrate the divergence between traditional models and the proposed approach. Moreover, we show that using the spatially heterogeneous model enables managers to design spatially differentiated management strategies based on reliable empirical models. We propose a spatially heterogeneous revenue response function dependent on selected service variables. Since our case study is based on real-world company data, the paper contributes to the literature on seasonal ticket revenue data. This is particularly appealing: in our literature review we only found very few studies dealing with seasonal ticket sales or revenues.

Acknowledgments

We thank an anonymous referee that, through a careful and insightful review, made significant contributions to the paper.

Appendix

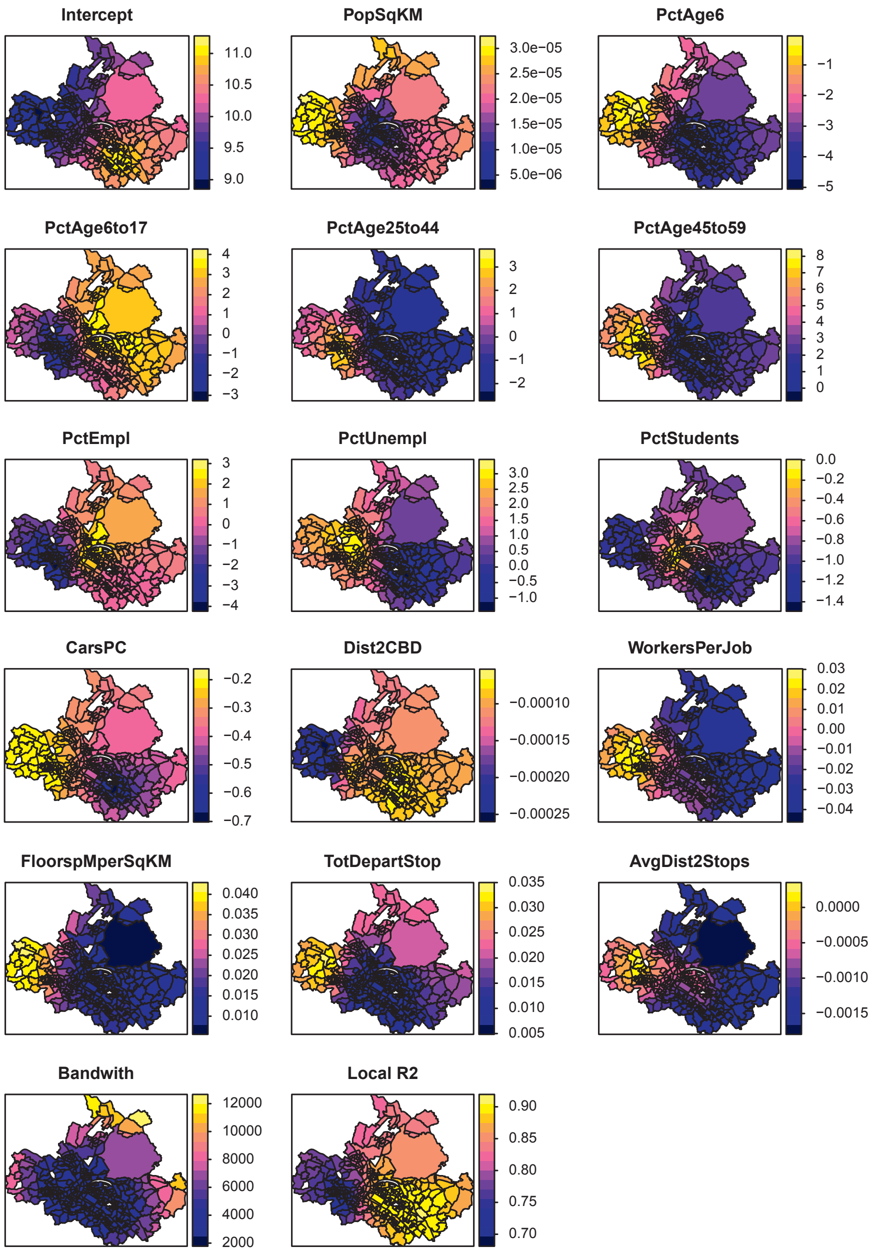


Fig. 4. Spatial distribution of the GWR coefficients exhibiting significant spatial variation according to the $\mathcal{F}_3$-test. The optimal bandwidths of the adaptive kernels of (11) and the local R^2 are displayed as well.

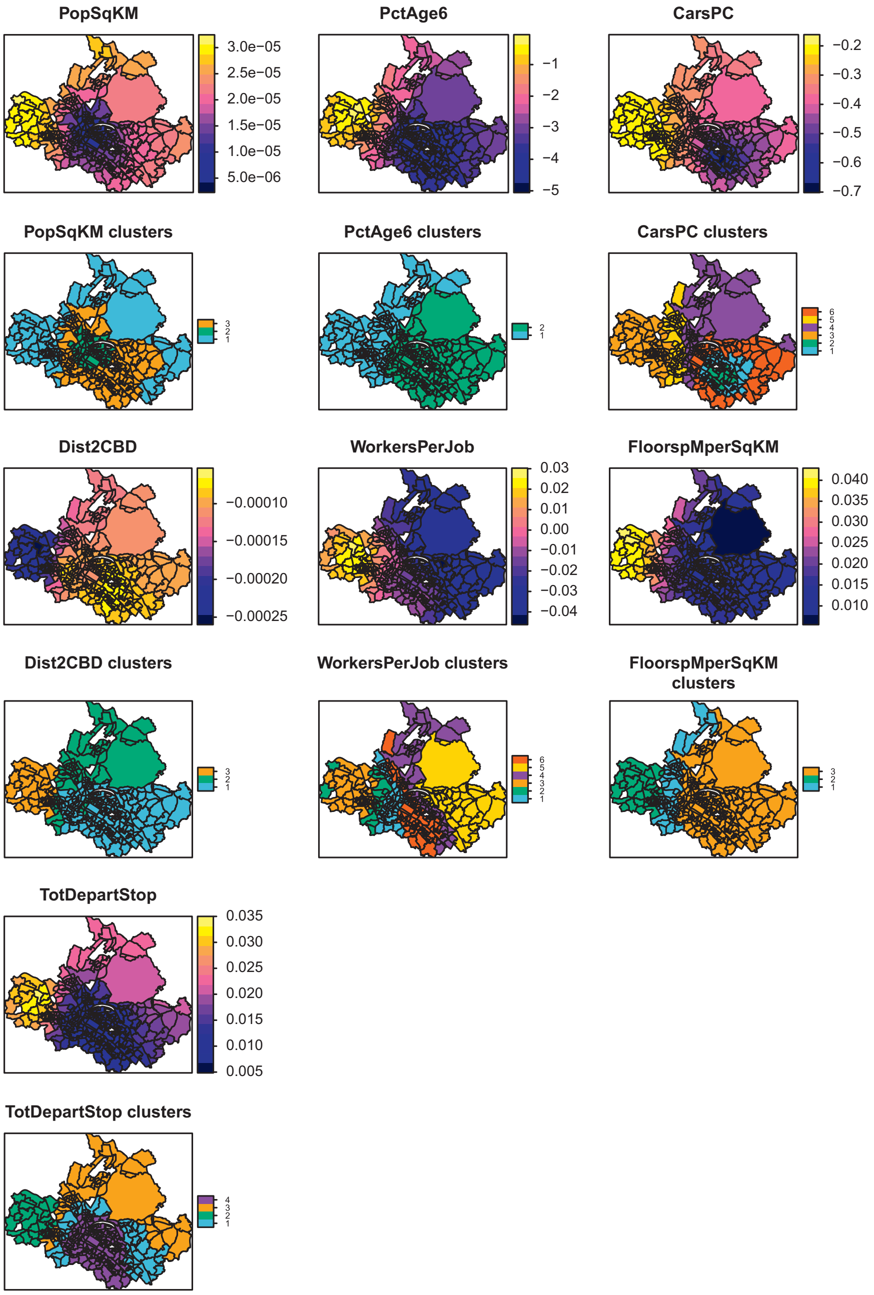


Fig. 5. Spatial clusters based on the GWR coefficient estimates (spatial drift). Only the variables are displayed which are used for the final spatially heterogeneous models.

References

Albers, S., 2011. Optimizable and implementable aggregate response modeling for marketing decision support. International Journal of Research in Marketing 29, 111–122.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Studies in Operational Regional Science, Kluwer.

Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. International Regional Science Review 26, 153–166.

Beer, C., Riedl, A., 2012. Modeling spatial externalities in panel data: the spatial Durbin model revisited. Papers in Regional Science 91, 299–319.

Bitter, C., Mulligan, G., Dall Erba, S., 2007. Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. Journal of Geographical Systems 9, 7–27.

Bivand, R., Pebesma, E., Gómez-Rubio, V., 2008. Applied Spatial Data Analysis with R. Springer, New York.

Bradlow, E., Bronnenberg, B., Russell, G., Arora, N., Bell, D., Duvvuri, S., Hofstede, F., Sismeiro, C., Thomadsen, R., Yang, S., 2005. Spatial models in marketing. Marketing Letters 16, 267–278.

Bronnenberg, B., 2005. Spatial models in marketing research and practice. Applied Stochastic Models in Business and Industry 21, 335–343.

Bronnenberg, B., Mahajan, V., 2001. Unobserved retailer behaviour in multimarket data: joint spatial dependence in market shares and promotion variables. Marketing Science 20, 284–299.
Brown, A., 2002. Do You Come from Manchester? A Postcode Analysis of the Location of Manchester United and Manchester City Season Ticket Holder, 2001. Technical Report, Manchester Metropolitan University.
Cardoso, M., de Carvalho, A., 2009. Quality indices for (practical) clustering evaluation. Intelligent Data Analysis 13, 725–740.
FitzRoy, F., Smith, I., 1999. Season ticktes and the demand for public transport. Kyklos 52, 219–238.
Forrest, D., Simmons, R., Feehan, P., 2002. A spatial cross-sectional analysis of the elasticity of demand for soccer. Scottish Journal of Political Economy 49, 336–355.
Foster, S., Gorr, W., 1986. An adaptive filter for estimating spatially varying parameters: application to modeling police hours spent in response to calls for service. Management Science 32, 878–889.
Fotheringham, A., Brunsdon, C., Charlton, M., 1997. Geographically Weighted Regression: An Analysis of Spatially Varying Relationships. Wiley and Sons, Chichester.
Garber, T., Goldenberg, J., Libai, B., Muller, E., 2004. From density to destiny: using spatial dimension of sales data for early prediction of new product success. Marketing Science 23 (3), 419–428.
Goodwin, P., Williams, H., 1985. Public transport demand models and elasticity measures: an overview of recent British experience. Transportation Research-B 19, 253–259.
Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2001. On clustering validation techniques. Journal of Intelligent Information Systems 17, 107–145.
Harris, R., Moffat, J., Kravtsova, V., 2011. In search of "W". Spatial Economic Analysis 6, 249–270.
Hawkins, D., Roupe, D., Coney, K., 1981. The influence of geographic subcultures in the United States. Advances in Consumer Research 8, 713–717.
Heitland, H., 2007. Alternativen im Verkehr: Abschätzung ihrer Chancen und Risiken durch PC-Simulationsmodelle. Frank & Timme GmbH.
Holmgren, J., 2007. Meta-analysis of public transport demand. Transportation Research-A 41, 1021–1035.
Huang, Y., Leung, Y., 2002. Analysing regional industrialisation in Jiangsu province using geographically weighted regression. Journal of Geographical Systems 4, 233–249.
LeSage, J., Pace, R., 2009. Introduction to Spatial Econometrics. Statistics: Textbooks and Monographs. CRC Press.
Leung, Y., Mei, C., Zhang, W., 2000. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. Environment and Planning A 32, 9–32.
Lloyd, S., 1982. Least square quantization in PCM. IEEE Transactions on Information Theory 28, 129–137.
McDonald, H., 2010. The factors influencing churn rates among season ticket holders: an empirical analysis. Journal of Sports Management 24, 676–701.
Mittal, V., Kamakura, W., Govind, R., 2004. Geographic patterns in customer service and satisfaction: an empirical investigation. Journal of Marketing 68, 48–62.
Parker, P.M., Tavassoli, N.T., 2000. Homeostasis and consumer behavior across cultures. International Journal of Research in Marketing 17, 33–53.
Rust, R.T., Zahorik, A.J., Keiningham, T.L., 1995. Return on quality (roq): making service quality financially accountable. Journal of Marketing 59, 58–70.
Sridhar, K., Bezawada, R., Trivedi, M., 2012. Investigating the drivers of consumer cross-category learning for new products using multiple data sets. Marketing Science 31, 668–688.
Tobler, W., 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46, 234–240.
Tollefson, J., Lessig, V., 1978. Aggregation criteria in normative market segmentation theory. Journal of Marketing Research 15, 346–355.
Wheeler, D., Paez, A., 2010. Geographically weighted regression. In: Fischer, M., Getis, A. (Eds.), Handbook of Applied Spatial Analysis. Springer, pp. 461–486.
Yu, D.-L., 2006. Spatially varying development mechanisms in the greater Beijing area: a geographically weighted regression investigation. Annals of Regional Science 40, 173–190.

Jahrb Reg wiss (2012) 32:113–132
DOI 10.1007/s10037-012-0064-3

ORIGINAL PAPER

Identifying spatial nonstationarity in German regional firm start-up data

Sven Müller

Accepted: 23 January 2012 / Published online: 31 January 2012

Abstract *Background*: We investigate the relationship between the rate at which new firms are established and regional characteristics and whether this relationship is constant over space or not. The characteristics reflect (i) agglomerations that in turn are related to increasing returns to production and (ii) measures which are influenceable immediately by regional decision makers.

Method: In order to account for spatial nonstationarity, we use the geographically weighted regression method for German start-up data on the geographical scale NUTS3. Moreover, we discuss significance test for locally varying regression coefficients.

Results: We are able to verify the global positive relationship between production convexities—measured by population density and growth amongst others—and the start-up rate on a regional level (Kreise). Furthermore, we find the share of industrial real estate to have positive influence on the regional start-up rate. Finally, we find strong empirically evidence that there is spatial nonstationarity in the data and hence the assumed relationship varies locally.

Conclusion: The results give evidence that spatial nonstationarity could not be neglected in the analysis of start-up rates. However, we suggest to develop global models that account—at least partially—for the underlying spatial nonstationarity by exogenous variables.

Keywords Geographically weighted regression · Spatial nonstationarity · Regional firm start-ups · Germany · Entrepreneurship · Economic Geography

S. Müller (✉)
Institute for Transport Economics, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany
e-mail: sven.mueller@wiso.uni-hamburg.de

1 Introduction

The past decades have shown a growing literature both on identifying the determinants of new business start-ups on a regional basis (see Fritsch 1992; Audretsch and Fritsch 1994; Audretsch and Jin 1994; Audretsch and Fritsch 1999 and Audretsch and Dohse 2007 for example) and on identifying spatial nonstationarity in economic structures and processes (see Mittal et al. 2004; Patridge and Rickman 2007; Meurs and Edon 2007; Bitter et al. 2007 and Yu 2006 for example). Spatial nonstationarity appears if the impact of an exogenous variable on the endogenous variable depends on the location. However, it is astounding that there is a lack of literature combining these two strands of research (Breitenecker and Harms 2010). In order to close this gap we have to discuss two questions: First, do start-up rates vary spatially and what is the underlying relationship? And second, is there spatial nonstationarity in this relationship?

1.1 Do start-up rates vary spatially?

Krugman (1991) states that the most striking feature of the geography of economic activity is the concentration of production in space. He focuses on convexities in production arising from spillovers from a pooled labor market; pecuniary externalities enabling the provision of nontraded inputs specific to an industry in a greater variety and at a lower cost; and information or technological spillovers. Here, we link characteristics reflecting these three sources of convexities at a spatial level in Germany to one aspect of the process of the concentration of economic activity: the rate at which new firms are being established (Audretsch and Fritsch 1994).

We assume that it is actually the interaction of increasing returns and uncertainty that bestows advantages to the pooling of labor markets associated with agglomerations. Moreover, we expect that agglomerations are also conducive to a greater provision of non-traded inputs. Such inputs (public transport for example) are provided at both a greater variety and a lower cost. Finally, we presume that technological spillovers are more beneficial to new small businesses than to incumbent large enterprises. These spillovers are expected to be more likely in agglomerations. According to this, we assume a higher rate of start-ups in agglomerated regions. In a given study area we will find agglomerated regions and other regions (rural areas for example). Therefore we assume that the start-up rate varies spatially.

1.2 Is there spatial nonstationarity in start-up data?

Let us say we find a formal link between the rate of start-ups and characteristics reflecting the three sources of convexities mentioned in Sect. 1.1. Then, would it be likely that this relationship is constant over the whole study area? The answer might be "no" since there are reasons why this formal link could be different in different regions. Generally, we could argue that agglomeration affects the establishment of new firms differently in distinct regions. To be precise: Let us consider two regions A and B with nearly the same characteristics reflecting agglomerations. The situation may occur that the effect of the unemployment rate is positive in region A and negative in region B. The negative effect in region B could be due to the fact that most of the businesses are sensitive to slack growth which in turn is indicated by high regional unemployment rates. In contrast the positive effect in region A might occur because firms located here are less sensitive to slack growth but more sensitive to potential labor resources. These resources are higher as unemployment rises. Please note that this effect does not have to be necessarily within the fashion of positive-negative. A spatial variation between less negative (positive) and more negative (positive) is imaginable as well. In general, there exist several explanations why relationships may vary over space and thus generate spatial nonstationarity. For more details on this issue see Fotheringham (1997). Concerning the case of firm birth rates the literature on spatial nonstationarity is scarce. However, there are some studies that shed some light on this issue. Scorsone et al. (2006) raise suspicion that one has to explicitly account for spatial nonstationarity in firm start-up data. They have witnessed spatial heterogeneity for firm start-up data in Kentucky, USA. In order to identify model coefficients that vary across regions they employ a global regression model that is similar to Seemingly Unrelated Regression. Calay et al. (2007) have pointed out that institutional factors like economic policies and social climate may lead to spatial heterogeneity. However, they do not explicitly account for spatial nonstationarity in their principal component analysis. More evidence for the existence of spatial nonstationarity in firm birth-rates give Cattani et al. (2003). Using conditional fixed-effects negative binomial regression models they explored organizational foundings as a function of spatial density, showing that local, more than national, density-dependence processes help explain industry evolution. Based on their results they conclude that local entrepreneurial deeds have repercussions beyond the boundaries of the geographical area in which a given start-up enterprise is residing. However, the overall spectrum of firm start-ups cannot be fully captured simply by looking at the region of the start-up rate itself but to consider neighboring areas as well. Most recently, Breitenecker and Schwarz (2011) are the first who report spatial nonstationarity in predictors of firm start-up activity of Austria using geographically weighted regression. However, they do not provide a theoretical underpinning of their findings.

1.3 How to measure spatial nonstationarity?

Now, if we accept that there might be spatial nonstationarity in firm start-up data, then we have to figure out how to measure this phenomenon. Methods for measuring spatial nonstationarity have been severally proposed in the literature. These methods include the expansion method (Casetti 1972), the method of adaptive filtering (Foster and Gorr 1986), the random coefficients model (Aitken 1996), the multilevel modeling (Goldstein 1987), the moving window approach and geographically weighted regression analysis (Fotheringham et al. 1997). However, geographically weighted regression (GWR) is a relatively simple, but effective technique for exploring spatial nonstationarity. Moreover, it is an important part of the trend towards local analysis and by this it is a truly spatial technique. Roughly speaking, it allows different relationships to exist at different points in space. This in turn is not provided in such an elaborated way by the methods mentioned before. The capabilities of modern software systems like geographic information systems ease the presentation of the results and their further processing and interpretation (Fotheringham 1997). There are well known shortcomings of the GWR as stated by Wheeler and Paez (2010) for example. One point of critique is that it is in essence an ensemble of local geographical regressions where the dependence between regression coefficients at different data locations is not specified in the model. A second issue is related to the repeated use of data to estimate model parameters at different model calibration locations, which causes a multiple comparisons situation. With an increasing number of local models estimated, the probability that some individual tests will appear significant, even if only by chance, will also increase. The problem in this case is related to the tradeoff between the amount of information and confidence, since the usual confidence intervals for regression coefficients are no longer reliable. Another issue with GWR that is directly related to the selection of the kernel bandwidth involves high levels of spatial variation and smoothness of estimated regression coefficients. A natural concern emerges that some variation or smoothness in the pattern of estimated coefficients may be artificially introduced by the technique and may not represent true regression effects. However, we like to use GWR in order to analyze whether there is spatial nonstationarity in German start-up data, because GWR is an accepted method for this issue.

In the next section a discussion of the geographically weighted regression can be found. Section 3 comprehends the description of the data and some issues related to measurement and operationalization. In Sect. 4 we discuss the results of our analysis followed by a short conclusion (Sect. 5).

2 Geographically weighted regression

Geographically weighted regression extends an ordinary linear regression model by allowing variations in rates of changes. The coefficients in GWR are specific to location $i \in I$ ("Kreis" for example) with $I = \{1, 2, \ldots, n\}$ rather than assumed to be constant. GWR is described in detail by Fotheringham et al. (1997a). The brief overview in this section is based on this reference unless stated otherwise. Let us consider

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \epsilon_i \quad i = 1, \ldots, n \tag{1}$$

where y_i is the dependent or endogenous variable and x_{ik} are independent or exogenous variables. ϵ_i are independent normally distributed error terms with zero mean and constant variance. β_k are coefficients to be estimated by (ordinary) least square method (OLS). We consider models that are estimated by OLS as OLS models. The estimators of β_k can be expressed in matrix form as follows:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{Y} \tag{2}$$

with

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1K} \\ 1 & x_{21} & \dots & x_{2K} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & \dots & x_{nK} \end{bmatrix}, \qquad \boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{pmatrix}.$$

It is important to note that the coefficients in (1) are assumed to be the same across the study area. However, this assumption is not always valid because different locations may have different coherences and thus different coefficients. To entertain varying coefficients, GWR extends the OLS model of (1) by allowing the coefficients $\hat{\boldsymbol{\beta}}$ of (2) to be estimated by a weighted least squares procedure. GWR makes the weighting system dependent on the location in geographical space and therefore, allows local rather than global coefficients to be estimated. Since the estimates become specific to location i the GWR model can be written as

$$y_i = \beta_{i0} + \sum_{k=1}^{K} \beta_{ik} x_{ik} + \epsilon_i, \quad i = 1, \dots, n. \tag{3}$$

Where β_{ik} is the value of the kth coefficient at location i.[1] The estimators of β_{ik} are

$$\hat{\boldsymbol{\beta}}_i = \left(\boldsymbol{X}^T \boldsymbol{W}_i \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{W}_i \boldsymbol{Y}. \tag{4}$$

Where the weighting matrix $\boldsymbol{W}_i$ is the n by n matrix whose off-diagonal elements are zero and whose diagonal elements are the weights of each observation, i.e. $\boldsymbol{W}_i = \text{diag}(\boldsymbol{W}_{i1}, \boldsymbol{W}_{i2}, \dots, \boldsymbol{W}_{in})$. It is easy to see that the OLS model of (1) is a special case of the GWR model of (3) with constant coefficients. It is known from the literature that allowing for spatial nonstationarity in the regression coefficients can account for at least some, and possibly a large part, of the autocorrelation in error terms in an OLS model estimated on spatial data. Besides, it appears that a GWR approach to spatial autoregressive modeling provides a relatively easy method of calculating both unconditional and conditional measures of local spatial autocorrelation (Páez et al. 2002).

In summary we see that the GWR model is able to measure spatial variations in relationships between the endogenous variable and the exogenous variables. This leads us to three crucial questions: First, what weighting matrix should be used? Second, how can we calibrate the particular parameters of the weighting function? And third, which model is better, a GWR model or an OLS model?

2.1 Weighting matrix

To estimate coefficients of the GWR model of (3), it is important to choose the criterion to decide on the weighting matrix. The role of the weighting matrix in GWR

[1] If needed, it is possible to obtain estimates at locations that are not data points.

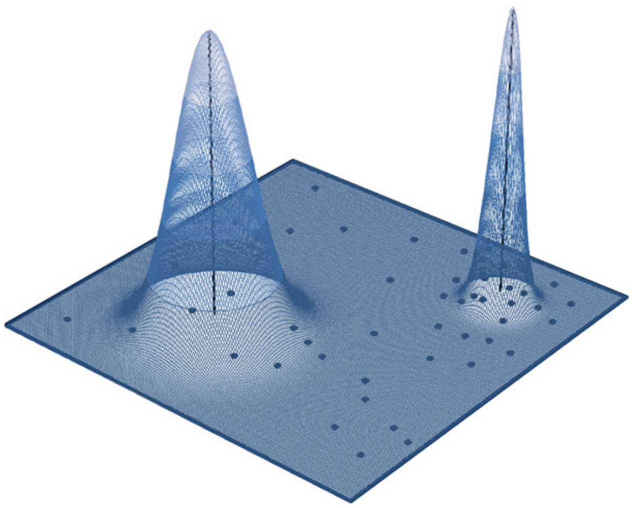


Fig. 1 Adaptive kernels: The kernels (and the respective λ_i) are larger in regions with scarce data and smaller in regions with dense data. Each point on the surface represents one data point

is to represent the importance of individual observation among locations. Based on Tobler's first law (Tobler 1970), in spatial analysis it is commonly assumed that areas close to each other share more common characteristics than areas that are more distant. Hence, when estimating the GWR coefficients at location i, more emphasis should be given to areas close to location i. As stated by Huang and Leung (2002) and Anslin (1988) there exist several ways to represent the weight between two locations i and j, but the weighting function used most commonly in empirical studies is the Gaussian

$$\boldsymbol{W}_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{\lambda}\right)^2\right), \quad i, j = 1, 2, \ldots, n. \tag{5}$$

$\lambda > 0$ is referred to as the bandwidth depicting the way the weights vary with distance. The way we determine λ is described in Sect. 2.2. If i is a point in space at which data are observed, the weighting of other points will decrease according to a Gaussian curve as the distance between i and j increases. For a given d_{ij}, a large λ results in a large weight for the observation at location j. On the other hand, for a given λ, the weight will practically fall to zero for those observations which are far enough from i, effectively excluding these locations in the estimation of coefficients for location i. Hence, if the distance between two locations i and j is sufficiently large, the data of location j have no influence on the start-up rate of location i.

There might be situations when a fixed λ seems to be inappropriate and a spatially varying bandwidth λ_i should be used instead. This results in a spatially varying weighting function as shown in Fig. 1. The rationale for this is twofold: (i) where data points are dense there is more scope for examining changes in relationships over relatively small distances and such changes might be missed with larger λ; and (ii) in regions where data are scarce, the standard errors of the GWR estimates, when a fixed bandwidth is used, will be high because the number of data points used will be small. In essence the problem of fixed λ in regions where data are dense is that the kernel which corresponds to the weighting function is larger than it needs to be and hence the estimates obtained from it are more likely to suffer from bias. Conversely, the problem with fixed λ in regions where data are scarce is one of inefficiency: the kernels are smaller than they need to be to reliably estimate coefficients. So in situations where data density is varying remarkably in space we should choose a varying λ_i.

2.2 Calibration of the weighting function

The coefficients of the GWR model of (3) are estimated using (4) dependent on the value of λ. To find the optimal value of λ we intuitively would select λ such that

$$\min \mathcal{R} = \sum_{i=1}^{N} \left(y_i - \hat{y}_i(\lambda)\right)^2 \tag{6}$$

i.e., minimizing the residual sum of squares. Where $\hat{y}_i(\lambda)$ is the fitted value of y_i at location i with respect to the given parameter λ. In order to find the fitted value $\hat{y}_i(\lambda)$, the $\hat{\beta}_{ik}$ at each location i needs to be estimated using (4). However, as Fotheringham et al. (1997b) note there is a problem when minimizing the residual sum of squares in (6) given a weighting function as (5). When λ is very small, the weighting of all locations except for i itself become negligible. Therefore, while the fitted values at location i, $\hat{y}_i$, will tend to the actual values y_i, the value of $\mathcal{R}$ becomes zero. This suggests that under such an optimizing criterion the value of λ tends to be—or close to—zero. Obviously, this is not expected. To solve this problem Fotheringham et al. (1997b) suggest a cross-validation approach (Bowman 1984 and Cleveland 1979) as

$$\min \mathrm{CV} = \sum_{i=1}^{N} \left(y_i - \hat{y}_{\neq i}(\lambda)\right)^2. \tag{7}$$

$\hat{y}_{\neq i}(\lambda)$ is the fitted value of y_i where the observations for location i are omitted from the calibration process. Thus, when λ becomes very small the model is calibrated only on samples near to i and not at i itself (Brunsdon et al. 1998).

2.3 How to compare a GWR model and an OLS model?

So far we have discussed the functionalities of the GWR model. However, we have to discuss measures that indicate whether the GWR model describes the coherences significantly better than an OLS model. Additionally, we would like to know whether each set of coefficient estimates exhibit significant spatial variation over the study area. Generally, the GWR model will fit a given data set better than an OLS model due to the increase in degrees of freedom. However, from the practical point of view, the simpler the model, the easier it is to be applied and interpreted. In order to test whether the GWR model is more appropriate compared to the OLS model Leung et al. (2000) have constructed several pertinent statistics. Here, we employ two selected test statistics:

1. $\mathcal{F}_1$-test: The null hypothesis of this test statistic is that there is no significant difference between a given OLS model and the GWR model under consideration in describing the coherences of interest. Basically this test consists of the ratio of the

residual sum of squares of the GWR model and the OLS model weighted by the number of parameters to be estimated. The distribution of the test statistic $\mathcal{F}_1$ is approximated by a F-distribution.

2. $\mathcal{F}_3(k)$-test: The null hypothesis of this test statistic is that all GWR coefficients for a given exogenous variable are equal. That is

$$\hat{\beta}_{1k} = \hat{\beta}_{2k} = \cdots = \hat{\beta}_{nk}, \quad \forall k \in K.$$

This test is based on the sample variance of the n estimated values of $\hat{\beta}_{ik}$. Again, the distribution of $\mathcal{F}_3(k)$ is approximated by a F-distribution.

Having at hand these two test statistics, we are able to identify whether the GWR model is statistically significantly better than the OLS model and for any given GWR model we are able to identify parameters which show statistically significant variation over space.

3 Measurement issues

We consider Germany as our study area. The geographical scale is "Kreise" which is roughly European NUTS3 regions. The geographic data can be obtained from (Eurostat 2009). Here we consider the centroids of the Kreise as data points. The year of our analysis is 2004. This is because this is a recent period for which consistent data is mostly available. We mainly use two sources of data. First, the establishment file of the German social infrastructure statistics (Fritsch and Brixy 2004). From this data we excerpt the start-up rate which is the endogenous variable in our case. Therefore, we consider two concepts.

1. *Labor Market Approach* (LMA): This method standardizes the number of new establishments with respect to the size of the workforce. This approach can be seen within in the light of the theory of entrepreneurial choice (Evans and Jovanovic 1989). That is, each new business is started by someone. The labor market approach implicitly assumes that the entrepreneur starting a new business is in the same labor market within which the new establishment operates.
2. *Ecological Approach* (EA): This method standardizes the number of entrants relative to the number of firms in existence. It considers the amount of start-up activity relative to the size of the existing population of businesses.

Second, we employ the "Regionaldatenbank" of the German federal statistical office (Federal Statistical Office and the Statistical Offices of the Länder 2009). Here we use the variables unemployment rate, change in unemployment rate, population density, population growth, gross value added per capita and the share of unskilled and semi-skilled workers in order to reflect the convexities outlined in Sect. 1.1. Since we expect a global positive effect of agglomerations on the firm birth rate, the assumed sign of the coefficients would be positive (negative for share of unskilled and semi-skilled workers). The variable mean establishment size is to control for measurement biases particularly inherent in the ecological approach. Of course, some regions may tend to have more employees per establishment than other regions do. Since ultimately people and not establishments start businesses, such heterogeneity with respect to mean establishment size would result in a measurement bias overstating start-up rates in regions where the mean establishment size is relatively high and understating it in those regions where it is relatively low. Of course, the greater the mean establishment size, the fewer are the number of establishments for any given workforce size. Thus, the predicted start-up rates tend to be systematically higher for those regions where mean establishment size is relatively high compared to what an OLS model without this measure would have predicted. In addition to these measures we would like to introduce the collection rate and the share of industrial real estate to our model. These two measures are appealing and obvious at the same time: On the one hand, regional authorities are—at least to some extent—able to control the collection rate and the amount of industrial real estate. On the other hand, it is very likely to assume that the amount of taxes and the availability of office and production space would influence the location choice of entrepreneurs or at least the decision process whether to become an entrepreneur. Finally, we consider the share of foreigners within a region as a control variable. Literature gives no definite evidence whether the influence of foreigners on the firm birth rate is positive or negative (Brüderl and Mahmood 1996; Boissevain et al. 1990; Lee et al. 2004 and Kontos 2003). We might argue that foreigners have a poor employment outlook. Therefore, they are forced in particular to start a business in order to avoid unemployment. Moreover, foreigners are more likely to serve special demand raising from other foreigners (and natives as well). In contrast, administrative barriers towards self-employment are more severe to foreigners. All variables with corresponding abbreviations, their sources and the expected effect can be found in Table 1.

4 Results

We estimate six different models. Note, that we omit some observations ("Kreise") due to missing data. The first two models (*Model 1 LMA*, *Model 1 EA*) are specified in order to reconcile our findings with the results of Audretsch and Fritsch (1994). The models *Model 2 LMA* and *Model 2 EA* introduce three new variables which have not been considered in the literature so far. The last two models are GWR models (*GWR Model LMA*, *GWR Model EA*) specified such that we are able to verify whether there is spatial nonstationarity in models *Model 2 LMA* and *Model 2 EA*. Table 2 contains all results. In order to avoid confusion with the interpretation of the results of the OLS model and the GWR model, we have to mention that the values given in Table 2 for the GWR models are values of the $\mathcal{F}_3$-test outlined in Sect. 2.3. That is: these values are not coefficients. The coefficients of the GWR models can be found in Figs. 3 and 4. Note that local standard errors, t-statistics and R^2 can be obtained from the GWR estimation procedures as well. However, we omit these measures due to the length of the paper.

Table 1 Variables used in regression analysis and expected global impact on start-up rates. If no measurement is given explicitly, the original value of the respective source is used

Variable	Measurement	Source	Expected sign
Start-up rate (LMA)	1000 × start-ups/(reg. employees + unemployed)	a	
Start-up rate (EA)	100 × start-ups/existing businesses	a	
Unemployment rate		b	+/−
Δ Unemployment rate	(Unempl. rate 2004 − unempl. rate 2000)/unempl. rate 2000	b	+/−
Population density	1000 Persons/km^2	b	+
Population growth	(population 2004 − population 2000)/population 2000	b	+
Share of unskilled & semi-skilled workers		b	−
Gross value added per capita	Gross value added/(1.000.000 × residents)	b	+
Mean establishment size	Employees/firms	a	−
Share of foreigners		b	+/−
Share of industrial real estate	Industrial real estate/total area	b, c	+
Collection rate	Collection rate/100	b	−

[a] Fritsch and Brixy (2004)

[b] Federal Statistical Office and the Statistical Offices of the Länder (2009)

[c] Eurostat (2009)

4.1 Global Estimates

The results of models *Model 1 LMA* and *Model 1 EA* mostly confirm the findings of Audretsch and Fritsch (1994). Merely, the change in unemployment rate and the per capita gross value added show different and not expected signs (although the coefficients cannot be considered as statistically significant). Globally, we can state that the rate of start-ups, and thus new economic activity is higher in regions where production convexities yield the greatest returns to that activity. Unfortunately, the influence of the unemployment rate remains ambiguous. Our findings suggest that, while a high unemployment rate results in a high number of start-ups relative to the number of existing establishments, the propensity of workers to start a new business in a high unemployment region tends to be relatively low. There are two possible interpretations for the negative relationship between the propensity of workers to start a business and the unemployment rate. The first is that the propensity to start a business is lower for unemployed than for employed workers. Thus, as workers shift from being employed to being unemployed, the overall entry rate tends to decline. The alternative explanation is that the propensity to start a business, regardless of employment status, is negatively influenced by higher regional rates of unemployment.

The coefficient for mean establishment size is negative for the labor market approach, but positive for the ecological approach. This discrepancy can be reconciled by the evidence suggesting that the propensity to start a business is greater for workers with experience in a smaller firm than in a large firm. However, the bias inherent

Table 2 Results for global (OLS) and geographically weighted regression (GWR). Coefficients of the GWR models can be found in Figs. 3 and 4. t-Statistics are given in parenthesis

Variables x_{ik}	Estimates $\hat{\beta}_k$				$\mathcal{F}_3$ values	
	Model 1		Model 2		GWR model	
	LMA	EA	LMA	EA	LMA	EA
Intercept	9.347 (30.92)	8.712 (20.88)	9.959 (19.79)	7.826 (11.60)	0.504	1.059
Unemployment rate	−0.028 (−1.99)	0.115 (5.96)	−0.036 (−2.47)	0.106 (5.49)	0.524	0.508
Change in unemployment rate	−0.010 (−0.03)	−0.464 (−1.14)	−0.099 (−0.33)	−0.443 (−1.09)	5.860*	1.599*
Population density	0.077 (1.17)	0.476 (5.26)	−0.075 (−0.78)	−0.060 (−0.46)	0.165	2.814*
Population growth	12.397 (5.66)	19.634 (6.16)	11.408 (4.88)	18.860 (6.02)	0.416	2.379*
Share of unskilled & semi-skilled workers	−7.320 (−5.70)	−5.687 (−3.21)	−8.166 (−6.13)	−7.571 (−4.24)	1.221	1.593*
Gross value added per capita	−0.011 (−1.44)	−0.013 (−1.12)	−0.013 (−1.68)	−0.014 (−1.30)	4.339*	2.231*
Mean establishment size	−0.150 (−12.36)	0.039 (2.36)	−0.160 (−12.42)	0.010 (0.58)	19.304*	4.057*
Share of foreigners			2.151 (1.36)	4.823 (2.28)	1.552*	2.323*
Share of industrial real estate			7.947 (2.46)	13.326 (3.07)	1.938*	0.182
Collection rate			−0.100 (−0.89)	0.414 (2.73)	4.981*	5.882*
N	362	362	359	359	359	359
F	108.3	25.99	78	22.7		
$\mathcal{F}_1$					0.917	0.956
$\mathcal{F}_1$: p-value					0.2115	0.3344
AICc	939.472	1222.133	932.341	1198.595	778.494	1009.798
Moran's I	0.274	0.258	0.265	0.236	0.237	0.201
Breusch-Pagan test	54.35	16.54	72.42	11.81	–	–
Breusch-Pagan critical value	14.07	14.07	18.31	18.31	–	–

* Indicates that the corresponding coefficient varies significantly (at least on the 95%-level) over the study area

under the ecological approach leads to an understatement of start-up activity in regions where the mean establishment size is relatively low, and an overstatement in regions where it is relatively high. This bias more than offsets the differential in the propensity for a worker to become an entrepreneur between large and small firms.

The models *Model 2 LMA* and *Model 2 EA* mostly confirm the findings of Audretsch and Fritsch (1994) as well. However, it seems that the population density does not have a statistically significant influence on the start-up rate. We may suggest, that the population density in the models *Model 1 LMA* and *Model 1 EA* is a proxy for other variables describing the production convexities in agglomerations. A possible explanation might be that regions with a highly densed population are namely agglomerations but possibly not all agglomerations inherit the same degree of production convexities (and thus higher start-up rates). Another explanation why the population density in our models is statistically non-significant compared to Audretsch and Fritsch (1994) could be due to differences in entrepreneurship between the years 1985 and 2004. Moreover, their analysis is carried out on a larger geographical scale ("Arbeitsmarktregionen"). The share of foreigners seems to be a better proxy for the determination of agglomerations providing production convexities in most parts of Germany: If we do not consider this variable in the model the population density becomes statistically significant although both variables do not exhibit perfect multicollinearity (that is the corresponding variance inflation factors are less than five). The existence of a high share of foreigners seems to foster the start-up activity (cf. Breitenecker and Schwarz 2011). Foreigners relatively seem to be more likely to become entrepreneurs. Likewise regions with a high share of foreigners maybe provide an ambience that is positively related to economic activity. As expected, the share of industrial real estate has a positive influence on the location choice of entrepreneurs. This gives at least some evidence that decision makers might influence regional business development by providing adequate offices and production space. Unfortunately, the influence of the collection rate is not clear. The negative influence of the collection rate in *Model 2 LMA* is expected but the coefficient is statistically non-significant. The positive influence in *Model 2 EA* is counter-intuitive but significant. One explanation might be that the collection rate has less influence on new establishments but more on existing firms—particularly in a relocation context. Concerning the goodness-of-fit we would prefer *Model 2 LMA* and *Model 2 EA* over *Model 1 LMA* and *Model 1 EA*.

4.2 Local estimates

For our GWR analysis we use a variable bandwidth λ_i of the spatial kernels in order to account for a varying density of data over the study area. The values of λ_i are determined by the procedure outlined in Sect. 2.2. The bandwidth of each location i is adapted in order to include nearly 50% of the 370 Kreise. Figure 2 shows that the smallest values of λ_i can be found in the center and south-west of Germany. This is due to the small shape of Kreise in this region and the large number of surrounding Kreise. Close to the borders the number of nearby Kreise and thus data points becomes smaller and hence λ_i becomes larger (as depicted exemplary in Sect. 2.1 or rather Fig. 1).

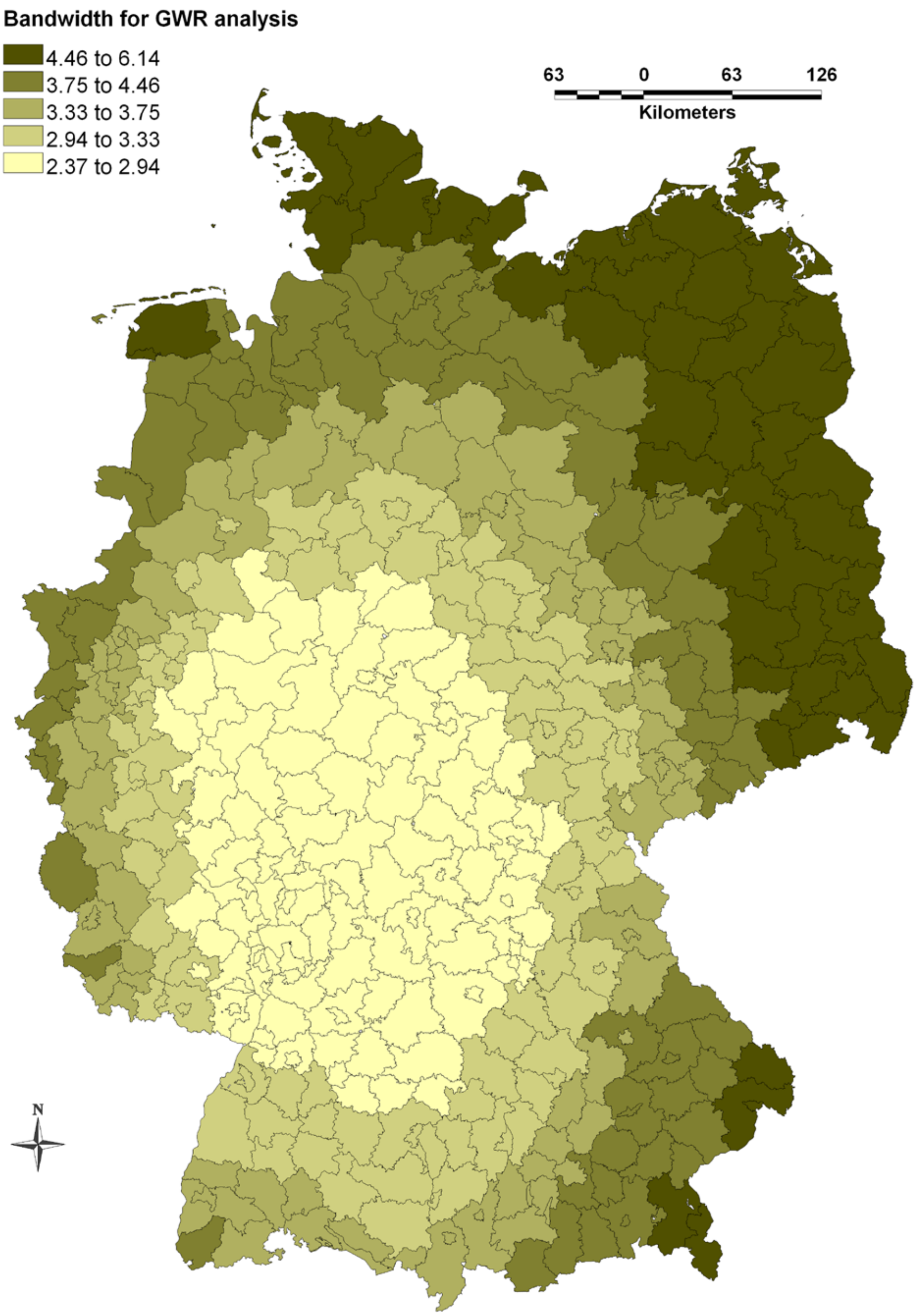


Fig. 2 Bandwidth λ_i used for GWR models. Class intervals are defined by quantiles. Bandwidth numbers are given in hundreds of kilometers

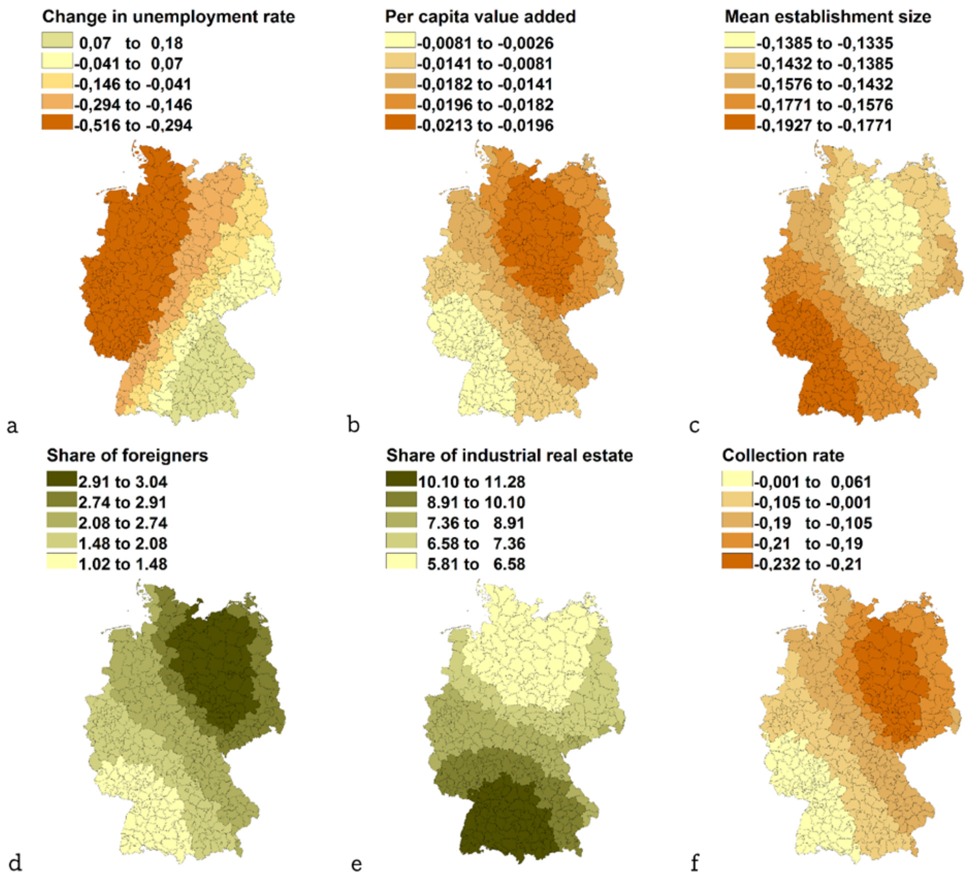


Fig. 3 Coefficients of the GWR model for the labor market approach. Class intervals are defined by quantiles. Note, only coefficients that vary significantly over space are dislayed

Figures 3 and 4 show the estimated local coefficients of the GWR models of Table 2. Although most of the coefficients show significant spatial variation, we should be careful due to the heteroscedasticity apparent in some of the global OLS models. Particularly for the labor market approach heteroscedasticity is remarkable. However, for the ecological approach most of the coefficients do vary significantly over space and the corresponding OLS model does not exhibit heteroscedasticity. Therefore, heteroscedasticity seems not to be the main source of spatial nonstationarity, here.

It is eye-catching that the coefficients that are statistically non-significant in all global models exhibit statistically significant spatial nonstationarity. This is an indicator for the average effect of the global models. Particularly for the labor market approach the global insignificance of the coefficient of the change in unemployment rate seems to be due to spatial nonstationarity. As Fig. 3a depicts there is a change in the sign of the coefficient from south-east Germany to north-west Germany. The same is true for the ecological approach and the variable population density, as we can learn from Fig. 4b. The change in the unemployment rate has a positive or less negative impact in southern Germany and a more negative impact in the north of Germany (Figs. 3a and 4a). The coefficient of the gross value added per capita decreases

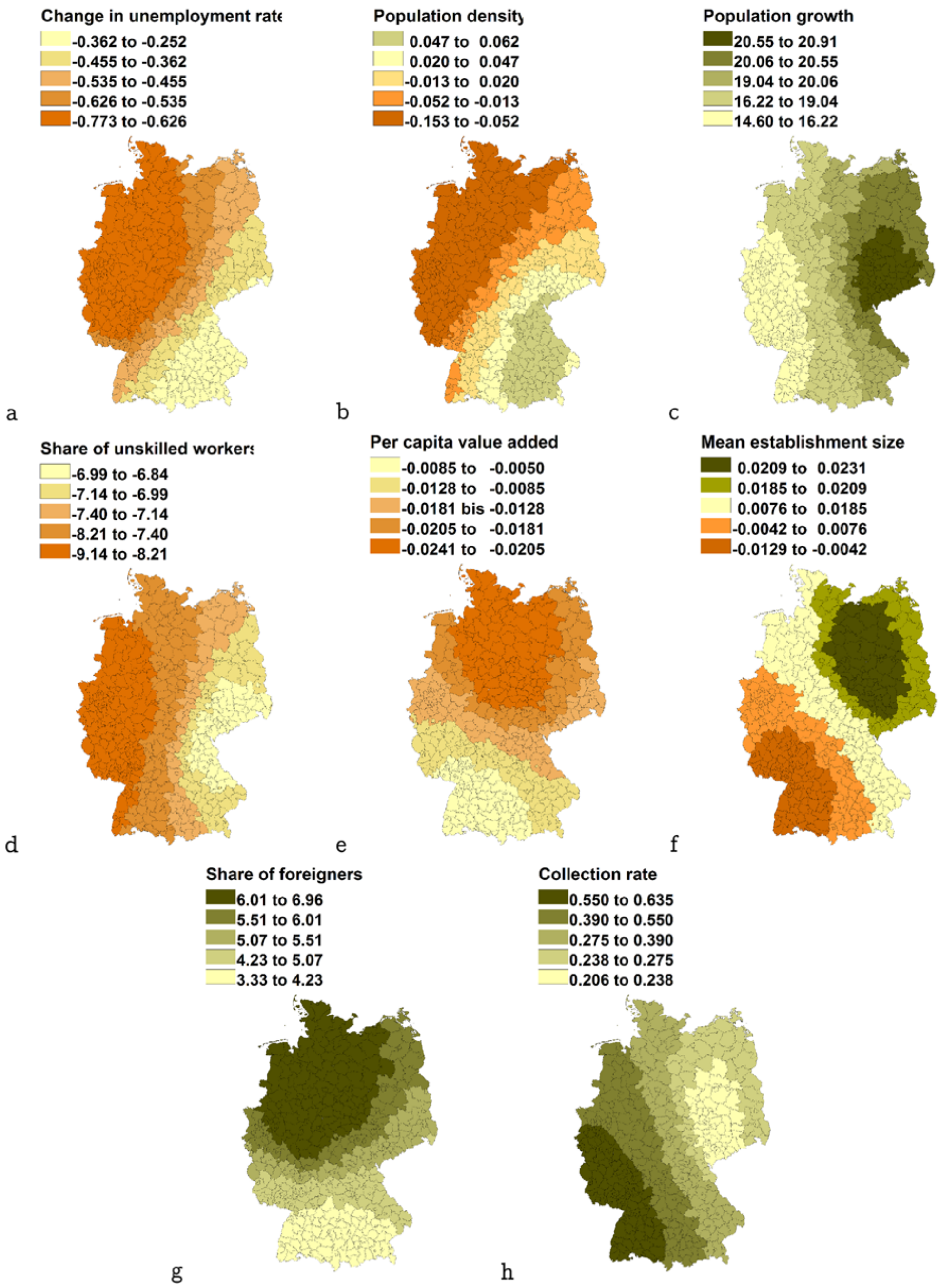


Fig. 4 Coefficients of the GWR model for the ecological approach. Class intervals are defined by quantiles. Note, only coefficients which that significantly over space are dislayed

from south-west to north-east (see Figs. 3b and 4e). It is interesting to see, that—as Fig. 4c depicts—the most positive influence of the population growth can be found in eastern Germany (particularly in Saxony). This might be due to the growth of urban areas in these regions. Contrarily, the rural areas in these regions suffer from severe slack growth and population decline. Especially in the southern "Neue Länder" (south part of the former German Democratic Republic) a high rate of start-ups is accompanied with population growth. Holding all other variables constant this means that if in east-southern regions there is population growth, then the according firm birth rate is higher than compared to regions in western Germany obtaining the same population growth. We might assume that in terms of the relationship between firm birth rate and population growth the core-periphery (or urban-rural) distinction is more pronounced in east-southern Germany than in the other regions.

Now we go back to what Krugman (1991) said about production convexities and the spatial concentration of economic activity (see Sect. 1.1). We see that generally this relationship is strongest in southern and western Germany. That is, besides the actual values of the exogenous variables, their respective local impact shows an expected spatial variation: south-north decline and west-east decline. An example: The coefficient for the share of unskilled workers for the ecological approach is most negative in western Germany and less negative in eastern Germany (see Fig. 4d). We might interpret this in the way, that new firms established in western Germany rely more on high skilled workers than new firms in eastern Germany do. So, if we only consider this variable and hold all others constant the start-up rate differs between a western and an eastern region—i.e., is lower in the western region—although the share of high-skilled workers is the same for both regions.

As stated earlier, the impact of the mean establishment size for the ecological approach is globally positive. However, the corresponding coefficient is statistically non-significant in *Model 2 EA*. If we look at Fig. 4f we see that this insignificance might be due to the "average effect" of the OLS model. The sign of the coefficient changes from positive in eastern Germany to negative in western Germany. Since the OLS model considers all observations in the same way the effect of this variable diminishes in the global perspective. Workers employed in small firms are more likely to start a new business than workers employed in large firms. This relationship is certified by the negative sign of the coefficient of the mean establishment for the labor market approach. This relationship is stronger in south-west Germany than in north-east Germany (see Fig. 3c). The distinctive negative relationship between mean establishment size and the start-up rate of the labor market approach in the south-west might be due to the traditional dominance of small and medium sized businesses in these regions.

The influence of the share of foreigners is most positive in the north and the north-east for both approaches (see Figs. 3d and 4g). If we take into account the change in unemployment rate (the impact is most negative in the north-west), we find strong evidence for our hypothesis, that a bad employment outlook motivates foreigners to become self-employed. The collection rate has the most positive (less negative) influence in the south west for the ecological approach (labor market approach). Hence, we might say that particularly in eastern Germany the collection rate hinders the establishment of new firms (see Fig. 3f and 4h). The coefficient of the share of industrial real estate varies only for the labor market approach (Fig. 3e). So, entrepreneurs in the south are more sensitive to appropriate office or production space.

Now, if we have to decide which model to choose, we face a dilemma. The AICc tells us to prefer the GWR models. This is certified by the significant spatial nonstationarity of most of the exogenous variables in the model. There is evidence in the literature to rely on these two criteria (see Fotheringham et al. 1997a, p. 96; Kam et al. 2005; Clark 2007 and Wheeler et al. 2006 for example). Nevertheless, we would like to rely also on the $\mathcal{F}_1$ test. Due to the values from Table 2 for both GWR models we cannot reject the null hypothesis. So it seems that the GWR model cannot improve the fitness significantly compared to the OLS models (*Model 2 LMA* and *Model 2 EA*). One side-effect that emphasizes the usefulness of GWR is the indicator of spatial autocorrelation Moran's I. Table 2 shows a remarkable decline in Moran's I between the OLS models and the GWR models. We interpret the contrary diagnostics in the way, that we have to search for a global model that heavily accounts for the spatial nonstationarity in the data and that accounts for spatial autocorrelation. This is a formidable task for future research.

5 Conclusion

So far, we have learned that the global empirical underpinnings pointed out by Audretsch and Fritsch (1994) of the theory of Krugman (1991) can be reconciled by our study. Moreover, we find that the share of industrial real estate has a positive significant influence on the regional start-up rate. This is particularly appealing since this measure gives evidence that regional decision makers may influence the rate of firm start-ups (at least partly). However, there are some sources of ambiguity. First, the impact of the unemployment rate could not be elucidated by our study. Second, the globally statistically non-significant coefficients of the population density and the gross value added per capita show negative signs that are not to be expected. Both variables are measures to identify agglomerations. Admittedly, we find that the share of foreigners might be a substitution to the population density in terms of identifying agglomerations in the firm start-up context. Moreover, a high share of foreigners is related to a higher rate at which new businesses are established. The underlying coherence might be the bad employment outlook for foreigners in distinct regions. We find strong evidence that there is spatial nonstationarity in German start-up data. We assume that the reasons for the spatial-nonstationarity are at least to some extent due to spatially different business cultures, economic policies and social climate. In general, we witness a decline in the magnitude of the coefficients from the south-west to the north-east of Germany. This is in line with the general economic capability of the german regions. Hence, firm start-ups in the south-west are more sensitive to the sources of convexities (spillovers from a pooled labor market; pecuniary externalities; and information or technological spillovers) as their counterparts in the north-east. In contrast to this general pattern, regions in eastern Germany that show population growth seem to provide a better environment for firm-start ups than their counterparts in western Germany. This might be due to the positive atmosphere related to the economic revival of these regions after periods of decline in the aftermath of the german reunification. The application of a geographically weighted regression exposes another interesting issue: we see that due to the "average effect" of global or rather ordinary regression models some variables in the global models appear to be statistically non-significant—namely the change in unemployment rate and the population density. However, the results obtained from the geographically weighted regression reveal, that the impact of these variables on the firm birth rate is positive in some regions and negative in other regions and thus this effect cancels out on a global scale. In future research we would like to address the following issues. First, we have to deal with the issue of multicollinearity in exogenous variables and the corresponding effect of correlation among local estimates raised by Wheeler and Tiefelsdorf (2005). Second, we conduct a cluster analysis in order to define distinct areas for which a given relationship of regional characteristics and the start-up rate is relatively constant. Such an analysis might at least shed some light on new variables that are to be introduced to the OLS models in order to account for the spatial non-stationarity. Such an OLS model should explicitly account for spatial autocorrelation and spatial errors. Moreover, such a model would be more reliable due to the GWR shortcomings mentioned in Sect. 1.3. Third, an extension to an industry specific analysis (see Audretsch and Fritsch 1999) and a spatiotemporal bandwidth (see Demsar et al. 2008) should be useful.

Acknowledgements The comments of two anonymous referees on an earlier version of the paper are gratefully acknowledged.

References

Aitken M (1996) A general maximum likelihood analysis of overdispersion in generalized linear models. Stat Comput 6:251–262

Anslin L (1988) Spatial econometrics: methods and models. Studies in Operational Regional Science. Kluwer Academic, Norwel

Audretsch D, Dohse D (2007) Location: a neglected determinant of firm growth. Rev World Econ 143:79–107

Audretsch D, Fritsch M (1994) The geography of firm births in Germany. Reg Stud 28:359–365

Audretsch D, Fritsch M (1999) The industry component of regional new firm formation processes. Rev Ind Organ 15:239–252

Audretsch D, Jin J (1994) A reconciliation of the unemployment—new firm startup paradox. Small Bus Econ 6:381–385

Bitter C, Mulligan G, Dall Erba S (2007) Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. J Geogr Syst 9:7–27

Boissevain J, Blaschke J, Grotenberg H, Joseph I, Light I, Sway M, Waldinger R, Werbner P (1990) Ethnic entrepreneurs and ethnic strategies. In: Waldinger R, Aldrich H, Ward R (eds) Ethnic entrepreneurs: immigrant business in industrial societies. Sage, Newbury Park, pp 131–156

Bowman A (1984) An alternative method of cross-validation for the smoothing of density estimates. Biometrika 71:353–360

Breitenecker R, Harms R (2010) Dealing with spatial heterogeneity in entrepreneurship research. Organ Res Methods 13:176–191

Breitenecker R, Schwarz E (2011) Detecting spatial heterogeneity in predictors of firm start-up activity of Austria with geographically weighted regression. Int J Entrepren Small Bus 12:290–299

Brunsdon C, Fotheringham A, Charlton M (1998) Spatial nonstationarity and autoregressive models. Environ Plan A 30:957–973

Brüderl J, Mahmood T (1996) Small business mortality in Germany: a comparison between regions and sectors. In: Diskussionspapier IV. No. 96–20. Wissenschaftszentrum, Berlin

Calay V, Guyot J, Van Hamme G (2007) Are human capital and culture the key factors in explaining intra-regional differences? Novice entrepreneurship and geo-cultural context in the Walloon region. In: Ulijn J, Drillon D, Lasch F (eds) Entrepreneurship, cooperation and the firm: the emergence and survival of high-technology ventures in Europe. EE Publishing, pp 220–240

Casetti E (1972) Generating models by the expansion method: applications to geographic research. Geogr Anal 4:81–91

Cattani G, Pennings J, Wezel F (2003) Spatial and temporal heterogeneity in founding patterns. Organ Sci 14(6):670–685
Clark S (2007) Estimating local car ownership models. J Transp Geogr 15:184–197
Cleveland W (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829–836
Demsar U, Fotheringham A, Charlton M (2008) Exploring the spatio-temporal dynamics of geographical processes with geographically weighted regression and geo-visual analytics. Inf Visual 7:181–197
Eurostat (2009). Geographic information system of the European commission. http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/introduction
Evans D, Jovanovic B (1989) Estimates of a model of entrepreneurial choice under liquidity constraints. J Polit Econ 95:657–674
Federal Statistical Office and the Statistical Offices of the Länder (2009). Regionaldatenbank. www.regionalstatistik.de
Foster S, Gorr W (1986) An adaptive filter for estimating spatially varying parameters: application to modeling police hours spent in response to calls for service. Manag Sci 32:878–889
Fotheringham A (1997) Trends in quantative methods. i. Stressing the local. Progr Hum Geogr 21:88–96
Fotheringham A, Brunsdon C, Charlton M (1997a) Geographically weighted regression: an analysis of spatially varying relationships. Wiley, Chichester
Fotheringham A, Charlton M, Brunsdon C (1997b) Measuring spatial variations in relationships with geographically weighted regression. In: Fischer M, Getis A (eds) Recent developments in spatial analysis. Springer, Berlin
Fotheringham A, Charlton M, Brunsdon C (1997) Two techniques for exploring nonstationarity in geographical data. Geogr Syst 4:59–82
Fritsch M (1992) Regional differences in new firm formation: evidence from West Germany. Reg Stud 26:233–241
Fritsch M, Brixy U (2004) The establishment file of the German social insurance statistics. J Appl Soc Sci Stud 124:183–190
Goldstein H (1987) Multilevel models in educational and social research. Oxford University Press, Oxford
Huang Y, Leung Y (2002) Analysing regional industrialisation in Jiangsu province using geographically weighted regression. J Geogr Syst 4:233–249
Kam S-P, Hossain M, Bose M, Villano L (2005) Spatial patterns of rural poverty and their relationship with welfare-influencing factors in Bangladesh. Food Policy 30:551–567
Kontos M (2003) Self-employment policies and migrants' entrepreneurship in Germany. Entrepren Reg Devel 15:119–135
Krugman P (1991) Increasing returns and economic geography. J Polit Econ 99:483–499
Lee S, Florida R, Acs Z (2004) Creativity and entrepreneurship: a regional analysis of new firm formation. Reg Stud 38:879–891
Leung Y, Mei C, Zhang W (2000) Statistical tests for spatial nonstationarity based on the geographically weighted regression model. Environ Plan A 32:9–32
Meurs D, Edon C (2007) France: a limited effect of regions on public wage differentials? Manch Sch 75:479–500
Mittal V, Kamakura W, Govind R (2004) Geographic patterns in customer service and satisfaction: an empirical investigation. J Mark 68:48–62
Patridge M, Rickman D (2007) Persistent pockets of extreme American poverty and job growth: is there a place-based policy role? J Agric Resour Econ 32:210–224
Páez A, Uchida T, Miyamoto K (2002) A general framework for estimation and inference of geographically weighted regression models. 2. Spatial association and model specification tests. Environ Plan A 34:883–904
Scorsone E, Fleming R, Somov M, Burke V (2006) The role of socioeconomic factors in entrepreneurship development: a county level analysis. In: Nijkamp P, Moomaw R, Traistaru-Siedschlag I (eds) Entrepreneurship, investment and spatial dynamics: lessons for an enlarged EU. EE Publishing, pp. 141–163
Tobler W (1970) A computer movie simulating urban growth in the Detroit region. J Econ Geogr 46:234–240
Wheeler B, Rigby J, Huriwai T (2006) Pokies and poverty: problem gambling risk factor geography in New Zealand. Health Place 12:86–96

Wheeler B, Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J Geogr Syst 7:161–187

Wheeler D, Paez A (2010) Geographically weighted regression. In: Handbook of applied spatial analysis. Springer, Heidelberg, pp 461–486

Yu D-L (2006) Spatially varying development mechanisms in the Greater Beijing Area: a geographically weighted regression investigation. Ann Reg Sci 40:173–190

Institut für Verkehrswirtschaft
Lehrstuhl für Betriebswirtschaftslehre, insb. Verkehr
Schwerpunkt: Operations & Supply Chain Management

Local Revenue Response to Service Quality: Spatial Effects in Seasonal Ticket Revenue Data

Sven Müller* , Knut Haase*

June 30, 2014

Accepted for publication in European Journal of Marketing

Corresponding author

*Institute for Transport Economics, University of Hamburg, D-20146 Hamburg, Germany, phone: +49 40 42838 9021, fax: +49 40 42838 2924, {sven.mueller, knut.haase }@uni-hamburg.de

Local Revenue Response to Service Quality: Spatial Effects in Seasonal Ticket Revenue Data

Abstract

Purpose - When firms' customers are located in spatially dispersed areas, it can be difficult to manage service quality on a geographically small scale because the relative importance of service quality might vary spatially. Moreover, standard approaches discussed so far in the marketing science literature usually neglect spatial effects, such as spatial dependencies (spatial autocorrelation, for example) and spatial drift (spatial non-stationarity).
Design/methodology/approach - We propose a comprehensive but intelligible approach based on spatial econometric methods that covers spatial dependencies and spatial drift simultaneously. In particular, we incorporate the spatial expansion method (spatial drift) into spatial econometric models (spatial lag model, for example).
Findings - Using real company data on seasonal ticket revenue (dependent variable) and service quality (independent variables) of a regional public transport service provider, we find that the elasticity for the length of the public transport network is between 0.2 and 0.5, while the elasticity for the headway is between -0.2 and 0.6, for example. We control for several socio-economic, socio-demographic, and land-use variables.
Practical implications - Based on the empirical findings, we show that addressing spatial effects of service data can improve management's ability to implement programs aimed at enhancing seasonal ticket revenue. Therefore, we derive a spatial revenue response function that enables managers to identify small scale areas that are most efficient in terms of increasing revenue by service improvement.
Originality/value - The paper addresses the need to account for spatial effects in revenue response functions of public transport companies.

Keywords: Spatial econometrics; spatial expansion method; spatial dependencies; spatial drift; revenue response function; public transport seasonal ticket; revenue management; service quality

Paper type Research paper

1 Introduction

Revenue management is essential to the profitability and the success of a firm. Hence, reliable information about the influencing factors that drive revenue is needed. The coherence between these factors and revenue might be mapped by a revenue response function (Tollefson and Lessig, 1978; Albers, 2011). The revenue response of customers are expressed as a function of price, service quality and/or advertising activities, for example. Managers might use such a function in order to predict the increase of revenue due to improvements in service or the expenditures in advertising and promotion. However, especially at firms that provide spatially varying service (-quality), managers face several challenges in implementing revenue management strategies (Groenroos 1984 and Rust *et al.* 1995). For geographically dispersed service areas, revenue and the importance placed on service quality will vary over locations. The firm's ability to provide sufficient service may also vary spatially. To account for the spatial dimension, we consider two types of spatial effects: (i) *spatial dependencies*: locations proximate to one another usually share resources, history, and socio-demographic and economic make-up. Therefore, consumer culture, lifestyle, values, attitudes, benefits, and consumption tend to be spatially associated as well. Empirical support for such local similarities in cultural, attitudinal, and behavioral patterns can be found in several studies (e.g., Foster and Gorr 1986, Garber *et al.* 2004, Bronnenberg and Mahajan 2001, Bronnenberg 2005, and Anselin 2003). As a typical marketing related application we might imagine a consumer, whose decision to adopt a new telephone service is affected by interactions with other consumers who live or work in the same district. This kind of spatial dependency is called spatial autocorrelation. (ii) *spatial drift*: Let us say we find a formal link between revenue and service quality. Then, would it be likely that this relationship is constant over the whole study area? The answer might be "no" since there are reasons why this formal link could be different at different locations. Mittal *et al.* (2004) argue that geography dictates the parameters of a satisfaction rating regression model due to differences in lifestyle and climate. This kind of spatial effect can be regarded as a representation of unobserved spatial heterogeneity (sometimes called spatial non-stationarity) in which the parameters (as opposed to the dependent variable per se) follow a spatial process: spatial drift. For the incidence of spatial drift in spatial data analysis strong evidence exists in the literature in general (e.g., Casetti 2010, Liao and Wei 2012, Huang and Leung 2002, Bitter *et al.* 2007, and Helbich *et al.* 2012) and in the marketing science literature in particular (Mittal *et al.* 2004 and Du and Kamakura 2011). In contrast, the (implicit) key assumption in the traditional marketing science literature is that the behavior of a consumer is conditionally independent of the behavior of another consumer and that this behavior is spatially homogeneous (Bradlow *et al.*, 2005). However, remarkable contributions exist in the marketing science literature which account for dependencies between consumers (see Norton and Bass 1987, Bass 2004, Müller and Rode 2013, and Bollinger and Gillingham 2012, for example).

In this study, we employ revenue and service data of the year 2007 of a monopolistic public transport company (Münchener Verkehrsverbund MVV) located in the city of Munich, Germany, and the surrounding region. Revenue and service data of public transport companies are particularly appealing, because public transport companies operate in a geographically dispersed service area and they provide spatially varying service quality. Nevertheless, the ticket price is the determinant of revenue that first comes into mind. Numerous publications on revenue management and pricing already exists (Talluri and Van Ryzin, 2005). The impact of ticket pricing on revenue has also been well studied in general (Lowengart *et al.* 2003, Eckard and Smith 2012 and Courty and Pagliero 2012) and for non-for-profit organizations (Olson *et al.* 2005 and Hume *et al.* 2006) as well as for (monopolistic) public transport service providers (Cervero 1990 and Dana 2001) in particular. However, in our case study prices are constant (we consider a cross-sectional study). Therefore, we are interested in other factors that trigger revenue for a public transport service provider. In particular, we study the impact of service quality on seasonal ticket revenue. Interestingly, there is very sparse literature on cross-sectional data of seasonal ticket revenue in general (Brown 2002, Forrest *et al.* 2002, McDonald 2010) and on public transport seasonal tickets in particular (FitzRoy and Smith, 1999).

In our analysis we find that the accessibility of public transportation (distance to access points, in the broadest sense) and the frequency of departures have significant positive impact on seasonal ticket revenue. The results provide some evidence that the response of revenue to service quality is inelastic. Moreover, total population and socio-economic determinants (buying power and employment status) influence the spatial variation of seasonal ticket revenue. Most importantly, we find statistically significant spatial dependencies (correlated error terms) and spatial drift (spatially varying coefficients for service quality) indicating that standard approaches might suffer from biased results. Therefore, this paper contributes to the literature in two ways. First, we propose a novel but intelligible modeling approach in order to deal with spatial dependencies and spatial drift simultaneously, i.e. accounting for both in one single model (section 2). Second, we offer empirical findings about the coherence between seasonal ticket revenue and service variables while we control for several socio-economic variables and land-use variables (section 4.1). Our study employs real company data (section 3). Based on this, we specify a revenue response function of selected service variables that yields valuable managerial insights (section 4.2).

2 Spatial Econometric Models

The discussion in this section is fairly brief. For a more detailed discussion of spatial models we refer to Diggle (1983), Upton and Fingleton (1985), Anselin (1988), Cressie (1992), and LeSage and Pace (2009). For a specific marketing science view we refer to Bradlow *et al.* (2005) and Bronnenberg (2005).

Roughly speaking, spatial econometric models assume that individuals (or, more generally, units of analysis, such as postal codes) can be located in a space. Typically, responses by individuals are assumed to be correlated in such a manner that individuals near one another in space generate similar outcomes[1] - as stated in Tobler's First Law of Geography (Tobler, 1970, p. 236): "Everything is related to everything else, but near things are more related than distant things". The methodology can integrate complex spatial correlations between entities into a model in a parsimonious and flexible manner. Although these models are able to cope with spatial dependencies, they still imply a global relationship between the dependent variable (here: revenue) and the independent variables (service quality, for example). Therefore, we call such models, *global spatial models*.

There are several methods established to account for spatial drift. These methods include the expansion method (Casetti, 1972), the method of adaptive filtering (Foster and Gorr, 1986), the random coefficients model (Aitken, 1996), the multilevel modelling (Goldstein, 1987), the moving window approach and geographically weighted regression analysis (Fotheringham *et al.*, 1997). In the following brief discussion of the pros and cons we focus on the spatial expansion method and the geographically weighted regression. The former one is appealing due to its intelligibility, the latter one is a more recent and general method. Both methods are based on the idea, that the model parameters are a function of the observation's location in space. Hence, both methods are able to unmask spatial drift. Geographically weighted regression can be seen as a series of local regression models using a spatial kernel density as a weighting function. Unfortunately, geographically weighted regression (i) is complex in terms of interpretation and it might be difficult to deduce a coherent management strategy from the results of a geographically weighted regression (Müller *et al.*, 2013). (ii) geographically weighted regression models are known to deal only accidentally with spatial autocorrelation. Following Wheeler and Paez (2010) geographically weighted regression does not propose a base model for the source of the spatial drift and is thus more appropriately seen as a heuristic approach. The geographically weighted regression yields local estimates for every observation (location) of every independent variable. While the high number of parameters increases the amount of information about relationships, it also makes the results hard to interpret (particularly in terms of management strategies) as stated by Mittal *et al.* (2004). They propose to search manually for regions of homogeneous model parameters to implement management strategies. The spatial expansion method simply generates estimates as a function of longitude and latitude coordinates of the observations. As we aim to account for both, spatial dependency and spatial drift, in an intelligible way, we consider the spatial expansion method in order to account for spatial drift (section 2.2). For a more comprehensive comparison of geographically weighted regression and spatial expansion method see Paez (2005) and Fotheringham and Brunsdon (1999).

[1] In a competitive context, individuals might generate dissimilar (negatively correlated) outcomes.

2.1 Global Spatial Models

As pointed out by Anselin (1988), spatial dependence can be either caused by model misspecification, measurement problems like spill-over effects or result from the spatial organization and structure of the phenomena. Since the premise of independence of observations cannot be held in the presence of spatial dependencies (spatial autocorrelation), estimates can be inefficient, biased and/or inconsistent. In contrast to temporal autocorrelation, spatial autocorrelation can potentially go in any direction in space, increasing the complexity of this influence. The basis for global spatial models is the spatial relationship between observations defined in a spatial weights matrix $\boldsymbol{W}$. To get $\boldsymbol{W}$, we start with a binary matrix $\mathbf{B}$ indicating the neighborhood of locations i and j by setting the matrix element:

$$b_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are neighbors and } i \neq j \\ 0 & \text{else} \end{cases} \tag{1}$$

The definition of the neighborhood of locations can be achieved in a lot of different ways. In this paper two locations are set as neighbors when they have contiguous boundaries, meaning that they share at least two pairs of coordinates. An overview of standard and alternative methods of constructing $\boldsymbol{W}$ is given by Harris *et al.* (2011). We use the most common method of row-standardization, where every matrix element b_{ij} is divided by the respective sum of its row:

$$w_{ij} = \frac{b_{ij}}{\sum_{j'=1}^{n} b_{ij'}} \tag{2}$$

with w_{ij} being the elements of $\boldsymbol{W}$ and n being the number of observations (locations).[2] Having built the spatial weights matrix $\boldsymbol{W}$, we are now able to use spatial models that can deal with spatial dependencies.[3]

Spatial Lag Model

The spatial lag model (SLM) incorporates spatial dependencies by using a spatially lagged dependent variable as an independent variable (spatial autoregressive form). It is based on a global linear regression model (LM) of the form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3}$$

For n observations and p parameters, $\mathbf{y}$ is an n by 1 vector of dependent observations (here: revenue), $\mathbf{X}$ is a n times p matrix of independent explanatory variables (here: service variables) with the elements of the first column set to 1, and $\boldsymbol{\beta}$ is a p by 1 vector of respective coefficients. $\boldsymbol{\epsilon}$ is an n by 1 vector of independent normally distributed error terms with zero mean and constant variance σ^2 (i.i.d. normal). Using the spatial

[2] If i has no neighbor, then $w_{ij} = 0$.

[3] We have tested several other specifications of $\boldsymbol{W}$ in our empirical study in section 4. Since the specification of (2) and (1) yields best results we consider this specific definition only.

weights matrix $\boldsymbol{W}$, the spatial lag for y at location i is then built as:

$$[\boldsymbol{W}\mathbf{y}]_i = \sum_{j=1}^{n} w_{ij} \cdot y_j \tag{4}$$

and added to the model. The spatial lag model (SLM) is then expressed as:

$$\mathbf{y} = \rho\boldsymbol{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{5}$$

where ρ represents a spatial autoregressive coefficient. The SLM might be an appropriate model for our data, because a high revenue (i.e., high demand for transit services) in location A might foster an information spillover (transit service is reliable, for example) to a neighboring location B. Hence, customers located in B evaluate transit services as reliable yielding high revenue scores in location B.

Spatial Error Model

The spatial error model (SEM) is also based on the global regression model of equation (3), but here the spatial dependence is included in the error term:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{6}$$

$$\boldsymbol{\epsilon} = (\mathbf{I} - \lambda\boldsymbol{W})^{-1}\boldsymbol{\upsilon} \tag{7}$$

where λ is the coefficient of the spatially lagged autoregressive errors and $\mathbf{I}$ is an n by n identity matrix (Bivand *et al.*, 2008, pp. 289-296). $\boldsymbol{\upsilon}$ are assumed to be i.i.d. normal. It is likely that the SEM is the most appropriate model, since we expect omitted explanatory variables to have impact on the revenue as well. However, the omitted variables might be correlated over neighboring locations yielding correlated error terms. Consider travel-times, for example. Of course, the duration of a transit trip (compared to car travel times) has impact on transit demand and hence revenues. Let us consider two neighboring outskirt districts. The customers located in these districts are likely to face similar travel-times, because most of their (commuting) trips are expected to terminate in the city center or central business district (i.e., have equal length). Therefore, the travel-times of the two district might be correlated. Now assume that we neglect travel-times in our analysis. Then, the error terms of our models are expected to be correlated as well.

Spatial Durbin Model

The spatial Durbin model is basically a spatial lag model with an additional set of spatially lagged independent variables. The term $\boldsymbol{W}\mathbf{X}$ adds average neighboring observation values of the independent variables to the equation. γ is a $(p-1)$ by 1 vector (the intercept is not lagged) measuring the marginal impact of the independent variables from neighboring observations on $\mathbf{y}$. The spatial Durbin model (SDM) can be written as (Beer and Riedl, 2012):

$$\mathbf{y} = \rho\boldsymbol{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{W}\mathbf{X}\gamma + \boldsymbol{\epsilon} \tag{8}$$

It is easy to imagine that explanatory variables from neighboring locations have impact on the revenue of the actual location. Assume there are many bus stops located in district A close to the border to district B. We might further assume only a few (or none) bus stops are located in B. However, customers located in B might use the bus stops located in A yielding high revenue scores in B.
As Anselin (1988, p. 85) states, estimating the parameters of the SLM, SEM, and SDM with an ordinary least squares (OLS) approach leads to biased and inconsistent results, so maximum likelihood estimation is used to avoid these problems.

2.2 Spatial Expansion Method

In order to account for spatial drift we use the latitude-longitude coordinates to create multiplicative interaction variables in the model under concern. Hence, this model allows independent (explanatory) variables to have different impact on the dependent variable based on the point in space from which the sample data are collected. To illustrate this method we replace in (3), (5), (7) and (8), the matrix of independent variables $\mathbf{X}$ by

$$\begin{pmatrix} x_{10} & x_{10}LO_1 & x_{10}LA_1 & x_{11} & x_{11}LO_1 & x_{11}LA_1 & \dots & x_{1p} & x_{1p}LO_1 & x_{1p}LA_1 \\ x_{20} & x_{20}LO_2 & x_{20}LA_2 & x_{21} & x_{21}LO_2 & x_{21}LA_2 & \dots & x_{2p} & x_{2p}LO_2 & x_{2p}LA_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{i0} & x_{in0}LO_i & x_{i0}LA_i & x_{i1} & x_{i1}LO_i & x_{i1}LA_i & \dots & x_{ip} & x_{ip}LO_i & x_{ip}LA_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{n0} & x_{n0}LO_n & x_{n0}LA_n & x_{n1} & x_{n1}LO_n & x_{n1}LA_n & \dots & x_{np} & x_{np}LO_n & x_{np}LA_n \end{pmatrix} \tag{9}$$

with LO_i as the longitude or the X-coordinate and LA_i as the latitude or the Y-coordinate of the location of observation i. Further we replace the vector of coefficients $\boldsymbol{\beta}$ by

$$\begin{pmatrix} \beta_0 \\ \beta_0^{LO} \\ \beta_0^{LA} \\ \beta_1 \\ \beta_1^{LO} \\ \beta_1^{LA} \\ \vdots \\ \beta_p \\ \beta_p^{LO} \\ \beta_p^{LA} \end{pmatrix} \tag{10}$$

Note, for the spatial Durbin model of (8) we have to replace $\boldsymbol{\gamma}$ correspondingly. Consider exemplarily an observation i. The influence of independent variable m on y_i is

$$\left(\beta_m + \beta_m^{LO} LO_i + \beta_m^{LA} LA_i\right) x_{im}.$$

That is, for each explanatory variable we obtain two additional estimates describing the impact of location on revenues (β_m^{LO} and β_m^{LA}). The functional form is simply linear, what makes the estimation process straight forward. For example, the linear model of (3) can be written as

$$y_i = \underbrace{\beta_0 + \beta_0^{LO}LO_i + \beta_0^{LA}LA_i}_{\text{local intercept}} + \underbrace{\left(\beta_1 + \beta_1^{LO}LO_i + \beta_1^{LA}LA_i\right)}_{\text{local parameter of variable } x_{i1}} x_{i1} + \cdots$$
$$+ \underbrace{\left(\beta_p + \beta_p^{LO}LO_i + \beta_p^{LA}LA_i\right)}_{\text{local parameter of variable } x_{ip}} x_{ip}.$$

We see that each parameter is a linear function of geographical coordinates. Using the spatial expansion method, we may miss certain patterns of the assumed spatial drift. However, using a more general form in order to account for a (possible) richer pattern yields a more sophisticated model (simultaneous consideration of spatial drift and spatial dependencies). We assume that this linear form is a good compromise between neglecting spatial drift (as done so far in large parts of the marketing science literature) and accounting for a rich pattern of spatial drift via a general but rather sophisticated model.

It is known that if we employ (9) and (10) in (3) then (10) can be estimated by OLS. We consider such a model as a local linear model (LLM). If we consider (9) and (10) in the spatial lag model of (5), the spatial error model of (7) and the spatial Durbin model of (8) then we are able to account for spatial dependencies and spatial drift simultaneously. We call such a model *local spatial model* (LSM). Obviously, we do not need to consider spatial drift for all independent variables. The LSM can be estimated by maximum likelihood methods.

3 Research Setting, Data, and Model Building

We employ the annual seasonal ticket revenue of a public transport service provider of the region of Munich, Germany (Münchner Verkehrsverbund (MVV)). We consider a specific segment of the public transport revenue: seasonal ticket revenue.[4] A seasonal ticket of MVV is valid for a month and costs between 40 Euro and 60 Euro in 2007 - depending on the number of fare zones the ticket is valid for. The seasonal ticket enables the customer to use all public transport services. The service area of the MVV is displayed in figure 1a. Since annual seasonal tickets are personalized, the locations of the customers are known. We obtain aggregated revenue data on the scale of the 244 postal zip code areas in 2007 (see figure 1b). Public transport seasonal ticket sales - and thus revenue - is triggered by the demand for public transport services. This demand in turn is mainly influenced by the fare (or the price) a customer has to pay and the service quality (travel time, for example). Now, in our cross-sectional data on public transport seasonal ticket revenue, the price for a given seasonal ticket is constant. Hence, for our revenue response function only service quality variables are

[4]Other segments of revenue are single tickets and student passes, for example.

of interest. In the service quality literature higher quality is assumed to lead to customer satisfaction, which leads to customer loyalty and this drives firms' profitability (Storbacka *et al.* 1994 and Chang and Chen 1998). The firm's profitability increase may be derived from revenue expansion, cost reduction, or both simultaneously. Rust *et al.* (2002) suggest that firms that adopt primarily a revenue expansion emphasis perform better than firms that try to emphasize cost reduction and better than firms that try to emphasize both revenue expansion and cost reduction simultaneously. In empirical marketing research there exist strong evidence that service quality is a major antecedent of firms' revenue (Prag and Casavant 1994, Rust *et al.* 1995, Zeithaml *et al.* 1996, and Babakus *et al.* 2004). The MVV is a public, non-for-profit organization. Therefore, we are interested whether marketing literature reveals service quality for non-for-profit organizations as well (Dolnicar and Lazarevski 2009 and Macedo and Pinho 2006). For example, Woodside *et al.* (1989) stress the importance of service quality of hospitals in terms of willingness-to-recommend. Accessibility is identified as a core service variable in health care by several studies (Carman 1990, Vandamme and Leunis 1993, Lam 1997, and Sohail 2003). Most interestingly, Paulley *et al.* (2006) identified in a meta analysis that service quality is an important factor that triggers demand (and hence, revenue of) for public transportation services. In particular, accessibility of the public transport service and the frequency of departures are found to be very important measures of service quality (Nash 1978, Eboli and Mazzulla 2008, and Hensher and Stanley 2003). For a public transport service provider in Madrid, Spain, Matas (2004) reveals positive relationships between service quality, demand and revenue.

Against this background, we consider three selected service variables in our empirical study: (i) Besides the ticket price, the travel time is the most important service variable in public transportation. Since we lack information about the individual trips of the customers (and the related travel times) we rely on the headway instead of travel times. Total travel time includes waiting time.[5] Waiting time decreases with increasing headway (frequency of departures). We assume that a higher frequency of departures from public transport stops per period improves the attractiveness of public transport services, leading to a positive effect on (seasonal) ticket revenue. Therefore, we consider the maximum number of departures per hour of all stops located in a zip code area ("HEADWAY"). (ii) we assume the more kilometers of the public transport network within a zip code area is ("NETLENGTH"), the higher is the accessibility to public transport services. (iii) we might consider accessibility as the euclidean distance between an address and the closest rapid transit stop within a zip code area. Since we use aggregate data we consider the average of this measure over all addresses located in a zip code area ("AVGDIST2STOP"). This measure is particularly appealing if we consider figures 1 (a) and (b), because zip code areas of high revenue spatially follow the rapid transit lines.

[5]The demand for public transport is more elastic to waiting time than in-vehicle time.

Besides service quality, it is evidenced that consumer characteristics influence firms' revenue (Hawkins *et al.* 1981, Parker and Tavassoli 2000, and Maegi 2003). For example, the findings of Heo and Lee (2011) provide an opportunity for managers in the hotel industry to identify customers' particular characteristics that affect customers' perceptions of the fairness of revenue management practices. Of course, consumer characteristics influence the demand for public transportation and thus the revenue of public transport service providers as well (see Currie 2004 and Cascetta 2009, for example). Based on these empirical findings, we assume that several socio-demographic and socio-economic as well as land-use variables have an impact on the seasonal ticket revenue.

Socio-demographic variables: Population and population density are assumed to have positive influence on the seasonal ticket revenue because the more people live in a zip code area the higher the potential for public transport demand. In areas with high population density the relative attractiveness of public transport services might increase due to a higher propensity of traffic congestion and lack of parking space.

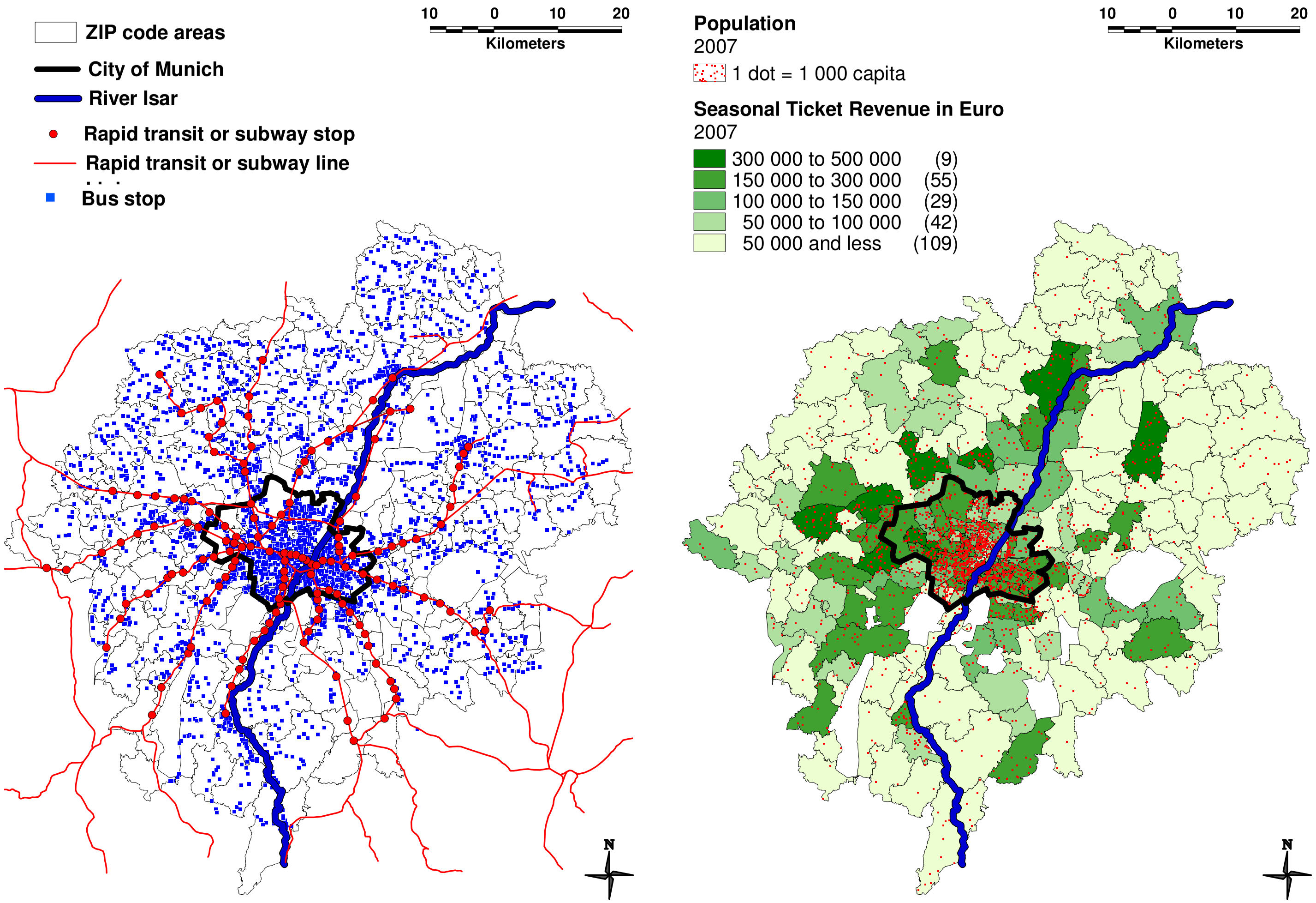
Socio-economic variables: We expect the employment status to have an impact on the seasonal ticket revenue as well. On the one hand, unemployed persons are less able to afford a car and thus they are more dependent on public transport service. On the other hand, these persons might not be able to afford an annual seasonal ticket. In contrast, employed persons are expected to be in a position to pay for both, a car and an annual seasonal ticket. Altogether, we do not make a distinct assumption about the expected sign. Further, we expect the higher the number of cars per capita within a zip code area the more likely it is for the persons located in this zip code area to own a car. Car ownership is assumed to reduce the demand for public transport and hence we assume a negative impact on revenue.

Land-use variables: Because the demand for public transportation is a derived demand, we assume that land-use patterns influence revenue. We presume that activities within a zip code area are accessible by non-motorized transport modes. Due to its high concentration of jobs, we deem the central business district (CBD) to be a good indicator for the influence of work on transport demand. The CBD is located in the city center of Munich. With very short or very long distances to the CBD, public transport is either unnecessary or inefficient. Otherwise, on medium distances, we assume that public transport is in fact an efficient transport mode. We therefore expect mixed, location dependent influences of the distance to the CBD on revenue. We further assume that a lower number of firms in a zip code area increases the propensity of its residents to travel to other places in order to work. If we consider the ratio of the number of employed persons located in a zip code area and the number of firms located in the same zip code area, we assume the higher this ratio is, the higher is the demand for transport and revenue. Like work, consumption drives the transport demand of people. Therefore, we suppose that a high density of retail firms located in a zip code area reduces the need to travel elsewhere to fulfill consumer needs. We assume a high retail firm density to reduce public transport seasonal ticket revenue in that zip code area. All variables and the corresponding summary statistics can be found in table 1.

There are two observations with PCTEMPL of nearly 100%. These are zip code areas with low population numbers. In conjunction with the economical prosperity of the region of Munich this phenomenon is reasonable (the minimum value of PCTUNEMPL corresponds to these observations). For one observation there are as many cars as capita. This is reasonable, because a large car rental agency is located there (the cars are registered with this zip-code area). The euclidean distance to the zip-code area of the CBD is zero for the zip-code area of the CBD itself, yielding a minimum value of DIST2CBD of zero. There are some observations with no public transport service. Therefore, the minimum value of NETLENGTH and HEADWAY is zero. We avoid potential scaling issues due to the variables BUYPOW, DIST2CBD, and AVGDIST2STOP by non-linear transformations of these variables (see model specification below).

Abbreviation	Description	min	median	mean	SD	max
REV	Total annual revenue of seasonal tickets of customers located in zip code area	462	60,570	96,650	95,166	461,300
POP	Total population in 10 000	0.04	0.89	1.09	0.85	4.17
POPDENS	POP per sq km	0.05	0.35	2.52	4.30	28.54
PCTEMPL	Employed persons / persons aged 20 and older	0.05	0.36	0.36	0.13	0.99
PCTUNEMPL	Unemployed persons / persons aged 20 and older	0.00	0.03	0.03	0.02	0.12
BUYPOW	Average annual disposable income	6,071	23,312	26,562	14,372	176,281
CARSPCAP	Number of cars per capita	0.33	0.60	0.61	0.12	0.99
RETAILDENS	Number of retail firms per sq km	0.05	2.28	18.82	51.17	567.20
FIRMSPCAP	Number of firms per (PCTEMPL $\times$ POP)	0.01	0.06	0.07	0.06	0.65
DIST2CBD	Euclidean distance in meters to the CBD (center of Munich)	0	20,345	21,040	13,754	53,188
AVGDIST2STOP	Average euclidean distances in meters between addresses and closest rapid transit stop	222.20	1,007.70	3,435.90	4,685.28	21,947.60
NETLENGTH	Public transport network length in km	0.00	1.12	2.60	3.54	22.81
HEADWAY	Maximum number of departures per hour over all stops	0.00	1.07	2.52	5.31	30.00

Table 1: Summary statistics. All values are given for 2007. The observations (n=244) are the zip code areas of the service area of the MVV (see figure 1)



(a) Transport network of the MVV

(b) Annual seasonal ticket revenue 2007

Figure 1: Service area of the MVV and annual seasonal ticket revenue per zip code area.

Based on the variables shown in table 1 and some previously performed model estimations we propose the following local linear model (LLM):[6]

$$
\begin{aligned}
\log(\mathrm{REV}) = \; & \beta_1 + \\
& \beta_2 \cdot \mathrm{POP} + \\
& \beta_3 \cdot \mathrm{POPDENS} + \\
& \beta_4 \cdot \log(1 + \mathrm{PCTEMPL}) + \\
& \beta_5 \cdot \mathrm{PCTUNEMPL} + \\
& \beta_6 \cdot \sqrt{\mathrm{BUYPOW}} + \\
& \beta_7 \cdot \mathrm{CARSPCAP\_mod} + \\
& \beta_8 \cdot (\mathrm{CARSPCAP\_mod})^2 + \\
& \beta_9 \cdot (\mathrm{CARSPCAP\_mod})^3 + \\
& \beta_{10} \cdot \mathrm{RETAILDENS} + \\
& \beta_{11} \cdot \mathrm{FIRMSPCAP} + \\
& \beta_{12} \cdot \log(1 + \mathrm{AVGDIST2STOP} \times \mathrm{DIST2CBD}) + \\
& \beta_{13} \cdot \log(1 + \mathrm{AVGDIST2STOP} \times \mathrm{DIST2CBD}) \times \mathrm{LO} + \\
& \beta_{14} \cdot \log(1 + \mathrm{AVGDIST2STOP} \times \mathrm{DIST2CBD}) \times \mathrm{LA} + \\
& \beta_{15} \cdot \log(1 + \mathrm{NETLENGTH}) + \\
& \beta_{16} \cdot \log(1 + \mathrm{NETLENGTH}) \times \mathrm{LO} + \\
& \beta_{17} \cdot \log(1 + \mathrm{NETLENGTH}) \times \mathrm{LA} + \\
& \beta_{18} \cdot \log(1 + \mathrm{HEADWAY}) + \\
& \beta_{19} \cdot \log(1 + \mathrm{HEADWAY}) \times \mathrm{LO} + \\
& \beta_{20} \cdot \log(1 + \mathrm{HEADWAY}) \times \mathrm{LA} + \\
& \epsilon
\end{aligned}
$$

Since the minimum values of DIST2CBD, NETLENGTH, and HEADWAY are zero, we add a quantity of one to these values in order to enable the log-transformation. We found a better fit for models with the interaction of AVGDIST2STOP and DIST2CBD than for models which consider both variables separately. Due to the cubic transformation of the CARSPCAP variable we avoid multicollinearity issues by using a modified car density variable (Dunlap and Kemery, 1987)

$$\mathrm{CARSPCAP\_mod} = \frac{\mathrm{CARSPCAP} - \mathrm{mean}(\mathrm{CARSPCAP})}{\mathrm{sd}(\mathrm{CARSPCAP})}.$$

Table 2 holds the estimation results for the LLM and a global linear model (LM) of the form of (3), as well as a local spatial model (LSM) of the form of (7) using partly (9) and (10) as specified at the beginning of section 4. In the following we consider the 95% confidence interval for the discrimination between statistically significant and statistically insignificant influence of independent variables.

[6] We omit the subscript i for convenience reasons.

4 Results

In this section we discuss empirical findings in section 4.1 and managerial insights including the specification of the revenue response function in section 4.2. We have used the software R (www.r-project.org) for model estimation.

A comparison between the spatial lag model, spatial error model and the spatial Durbin model points to the spatial error model to be the best model to account for spatial dependencies (see table 3 in the appendix). Only the spatial Durbin model and the spatial error model yield residuals that are spatially uncorrelated (statistically insignificant Moran's $\mathcal{I}$ statistic). This offers some evidence that the spatial correlation stems from omitted variables. Finally, the AIC of the spatial error model (425.49) is remarkably lower than the AIC of the spatial Durbin model (440.54). Moreover, the spatial error model is less sophisticated in terms of estimation, degrees of freedom, and interpretation than the spatial Durbin model (see section 2.1). Therefore, we consider the spatial error model to be the best model to deal with spatial dependencies. If we compare the three models LM, LLM and LSM we see that our assumptions about the sign of the coefficients are confirmed except the statistically insignificant land use variables RETAILDENS and FIRMSPCAP. There are no severe multicollinearity issues (max(VIF) less than 10) or heteroscedasticity issues (non-significant Breusch-Pagan-Test statistic). However, we see that the non-spatial models (LM and LLM) suffer from spatial autocorrelation (significant Moran's $\mathcal{I}$ statistic). In contrast, the LSM (here a spatial error model (SEM)) is able to account accurately for spatial dependencies (spatially correlated error terms). The positive and statistically significant estimate of λ indicates spatial correlation of the error terms. Moreover, we find the LSM offers the best (lowest) AIC. Therefore, LSM is our preferred model.

Nr.	log(REV) Variable	LM $\hat{\beta}$	LM SE	LLM $\hat{\beta}$	LLM SE	SEM (LSM) $\hat{\beta}$	SEM (LSM) SE
1	Intercept	10.76867	0.83383 ***	11.30178	0.79451 ***	11.13861	0.72446 ***
2	POP	0.79632	0.06742 ***	0.78339	0.06395 ***	0.80846	0.06030 ***
3	POPDENS	0.03289	0.02268	0.03727	0.02172 ·	0.03336	0.02127
4	log(PCTEMPL)	0.92328	0.30826 **	0.88041	0.29590 **	0.82956	0.27858 **
5	PCTUNEMPL	-3.45738	2.34395	-8.81727	2.36656 ***	-9.85228	2.36527 ***
6	sqrt(BUYPOW)	0.01224	0.00328 ***	0.00905	0.00315 **	0.00878	0.00296 **
7	CARSPCAP_mod	-0.21989	0.08693 *	-0.18643	0.08160 *	-0.12218	0.07530
8	$(\text{CARSPCAP\_mod})^2$	-0.07666	0.03906 ·	-0.06783	0.03654 ·	-0.04126	0.03407
9	$(\text{CARSPCAP\_mod})^3$	0.03234	0.01599 *	0.02748	0.01490 ·	0.01607	0.01373
10	RETAILDENS	-0.00196	0.00191	-0.00123	0.00184	-0.00134	0.00166
11	FIRMSPCAP	0.42246	1.42893	-0.16777	1.39509	0.05102	1.32380
12	log(1 + AVGDIST2STOP × DIST2CBD)	-0.01447	0.00442 **	-0.01324	0.00417 **	-0.01242	0.00386 **
13	log(1 + AVGDIST2STOP × DIST2CBD) × LO × 0.001			-0.00106	0.00020 ***	-0.00105	0.00023 ***
14	log(1 + AVGDIST2STOP × DIST2CBD) × LA × 0.001			0.00003	0.00020	0.00009	0.00024
15	log(1 + NETLENGTH)	0.34532	0.05802 ***	0.31860	0.05814 ***	0.32620	0.05588 ***
16	log(1 + NETLENGTH) × LO × 0.001			0.00050	0.00262	0.00084	0.00243
17	log(1 + NETLENGTH) × LA × 0.001			0.00314	0.00318	0.00328	0.00304
18	log(1 + HEADWAY)	0.25965	0.05392 ***	0.23266	0.05255 ***	0.21058	0.05222 ***
19	log(1 + HEADWAY) × LO × 0.001			0.00951	0.00374 *	0.00716	0.00375 ·
20	log(1 + HEADWAY) × LA × 0.001			-0.00530	0.00448	-0.00703	0.00449
21	λ					0.38885	0.08211 ***
	AIC	470.33		440.13		425.49	
	max(VIF)	8.04		8.60		8.61	
	Breusch-Pagan-Test	15.30		22.32		20.20	
	(p-value)	0.288		0.268		0.383	
	Moran's $\mathcal{I}$	0.25		0.16		-0.005	
	(p-value)	0.000		0.000		0.507	

Table 2: Results. Number of observations 244 for all three models. LSM is the spatial error model. Significance codes: 0 ***, 0.001 **, 0.01 *, 0.05 ·

4.1 Empirical Insights

In the following we focus on statistically significant variables. As expected the more persons are located in a zip code area, the higher the annual seasonal ticket revenue. Ceteris paribus we expect the annual seasonal ticket revenue to increase by 8.1% if the population increases by 100 persons. Concerning the socio-economic variables the results confirm our theory that seasonal tickets are only efficient for and affordable to persons who make many trips per year (commuters, for example). Employed persons have to make many more trips due to commuting than unemployed persons. Accordingly, we find a negative estimate for PCTUNEMPL. More specifically, we expect the revenue to decline by 0.005% if the percentage of unemployed persons increases by 1%. If the percentage of employed persons (PCTEMPL) increases by 1% then we expect the revenue to decrease by 0.83%. The positive sign of sqrt(BUYPOW) is in line with this finding: Since annual seasonal tickets are expensive, the expected revenue increases with the buying power: For an increase of the annual disposable income by 150 Euros we expect an 0.9% increase in seasonal ticket revenue. Concerning the service variables AVGDIST2STOP, NETLENGTH and HEADWAY we first define the parameters

$$\alpha_i = \hat{\beta}_{12}^{LSM} + \frac{\hat{\beta}_{13}^{LSM} LO_i + \hat{\beta}_{14}^{LSM} LA_i}{1000} \tag{11}$$

$$\delta_i = \hat{\beta}_{15}^{LSM} + \frac{\hat{\beta}_{16}^{LSM} LO_i + \hat{\beta}_{17}^{LSM} LA_i}{1000} \tag{12}$$

$$\omega_i = \hat{\beta}_{18}^{LSM} + \frac{\hat{\beta}_{19}^{LSM} LO_i + \hat{\beta}_{20}^{LSM} LA_i}{1000}. \tag{13}$$

α_i is the local impact of log(1 + AVGDIST2STOP $\times$ DIST2CBD) on log(REV), δ_i is the local impact of log(1 + NETLENGTH) on log(REV), and ω_i is the local impact of log(1 + HEADWAY) on log(REV). For example, for a given location i, ω_i represents the impact of HEADWAY (log(1 + HEADWAY), to be precise) at location i on the revenue at location i (log(REV)). Since we have 244 observations, we have 244 values of ω_i (and α_i and δ_i, as well).[7] The parameter values are summarized in empirical density plots (see figure 2). Figure 2 also displays the correlation between α_i, δ_i, and ω_i. All coefficients are approximately normal distributed and we do not witness serious correlation between the coefficients (i.e., there is no perfect linear relationship). The negative slope of the lines fitted to the scatter plots (linear regression and locally weighted polynomial curves) indicate a negative relationship between all parameters. For example, high values of δ_i (0.4 to 0.5) correspond to low values of ω_i (-0.1 to 0.1). This means that for locations i that exhibit the most positive impact of NETLENGTH on REV, the impact of HEADWAY on REV is least positive at the same time.

Concerning accessibility (log(1 + AVGDIST2STOP $\times$ DIST2CBD)) we find our assumptions to be confirmed: The more distant the rapid transit stops to the locations of customers the less the expected revenue. This effect becomes stronger the farther a customer is located away from the CBD. We witness a significant spatial drift of this

[7]Note, that we are able to compute parameter values for any pair of coordinates.

relationship. The impact of HEADWAY on revenue varies over space as well. In contrast, the assumed positive impact of NETLENGTH on revenue seems to be constant over space. The spatial variation of α_i, δ_i, and ω_i are displayed in figure 3 a to c. The most positive impact of HEADWAY on revenue is expected to be in the eastern parts of the service area. The negative values of ω_i in the north-western area are artifacts due to the specification (estimates are constant factors of coordinates). The impact of AVGDIST2STOP $\times$ DIST2CBD on revenue is most negative in the eastern parts of the service area. Altogether, we might assume that customers located in the (south-) eastern part of the service area are more sensitive to service improvements than customers located in other areas.

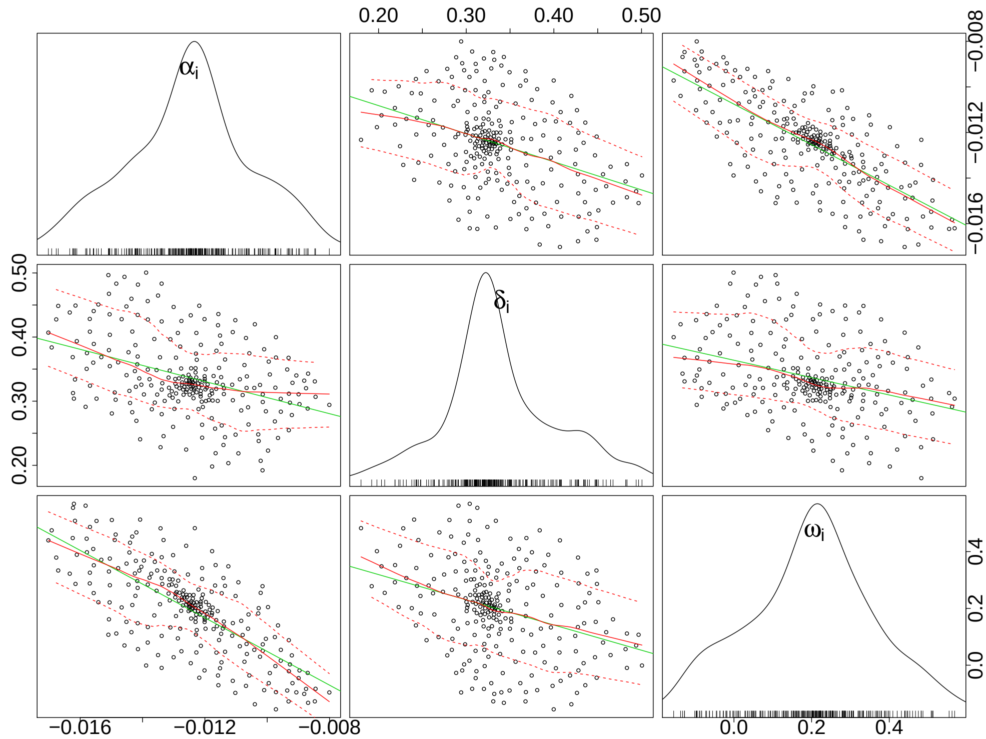


Figure 2: Empirical density plots and scatter plots of the spatially varying parameters α_i, δ_i, and ω_i. Axes hold the corresponding parameter values: α_i: bottom of left column, right to top row; δ_i: top of middle column, left to middle row; ω_i: bottom of right column, right to bottom row. Off-diagonal scatterplots show the correlations between α_i, δ_i, and ω_i. The relationship between the parameter values are described by regression lines (green) and locally weighted polynomial curves (red) and the corresponding 95-% confidence region (dotted, red). On the diagonal we see the empirical densities. The ticks immediately above the abscissae represent the observations (i.e., one tick corresponds to one observation). Consider ω_i, for example. The parameter values range between -0.1 and 0.5 (values are given at the bottom of the right column). Obviously, most observations exhibit values of ω_i around 0.2, yielding the highest density approximately at $\omega_i = 0.2$.

4.2 Managerial Insights

From a managerial point of view we are interested (i) in a revenue response function based on the empirical findings and (ii) how this revenue response function might be used in order to derive management strategies for service improvement. Revenue response functions are well established in marketing research and practice. Usually, concave response functions are used in applications (Skiera and Albers, 1998; Haase and Müller, 2014). In a large empirical comparison Hruschka (2006) finds merits of the concave function as well. Applications of revenue response functions can be found in sales force deployment (Drexl and Haase, 1999), revenue management (Talluri and Van Ryzin, 2005), asset management (Berger *et al.*, 2002), and many more (Rust and Zahorik 1993, Dekimpe and Hanssens 1999, Villanueva *et al.* 2008, and Hansses *et al.* 2001).

In order to specify a revenue response function based on our empirical results we first define an expected base profitability. The base profitability is not immediately influenceable by the management (population density, for example). Considering the proposed model of section 3 (LSM of table 2), here the base profitability of a zip code area i is defined as

$$\mathrm{BP}_i = \sum_{m=1}^{11} \hat{\beta}_m^{\mathrm{LSM}} x_{im}. \tag{14}$$

This is the sum of the socio-economic variables, socio-demographic variables, and land-use variables (including the intercept) weighted by their corresponding coefficient estimate (variables and estimates 1 to 11 of table 2). The values of BP_i are mapped in figure 4.

Then, we may write our model as

$$\begin{aligned}
\log(\mathrm{REV}_i) = \; & \mathrm{BP}_i \\
+ \; & \alpha_i \log(1 + \mathrm{AVGDIST2STOP}_i \times \mathrm{DIST2CBD}_i) \\
+ \; & \delta_i \log(1 + \mathrm{NETLENGTH}_i) \\
+ \; & \omega_i \log(1 + \mathrm{HEADWAY}_i) \\
+ \; & \epsilon_i
\end{aligned}$$

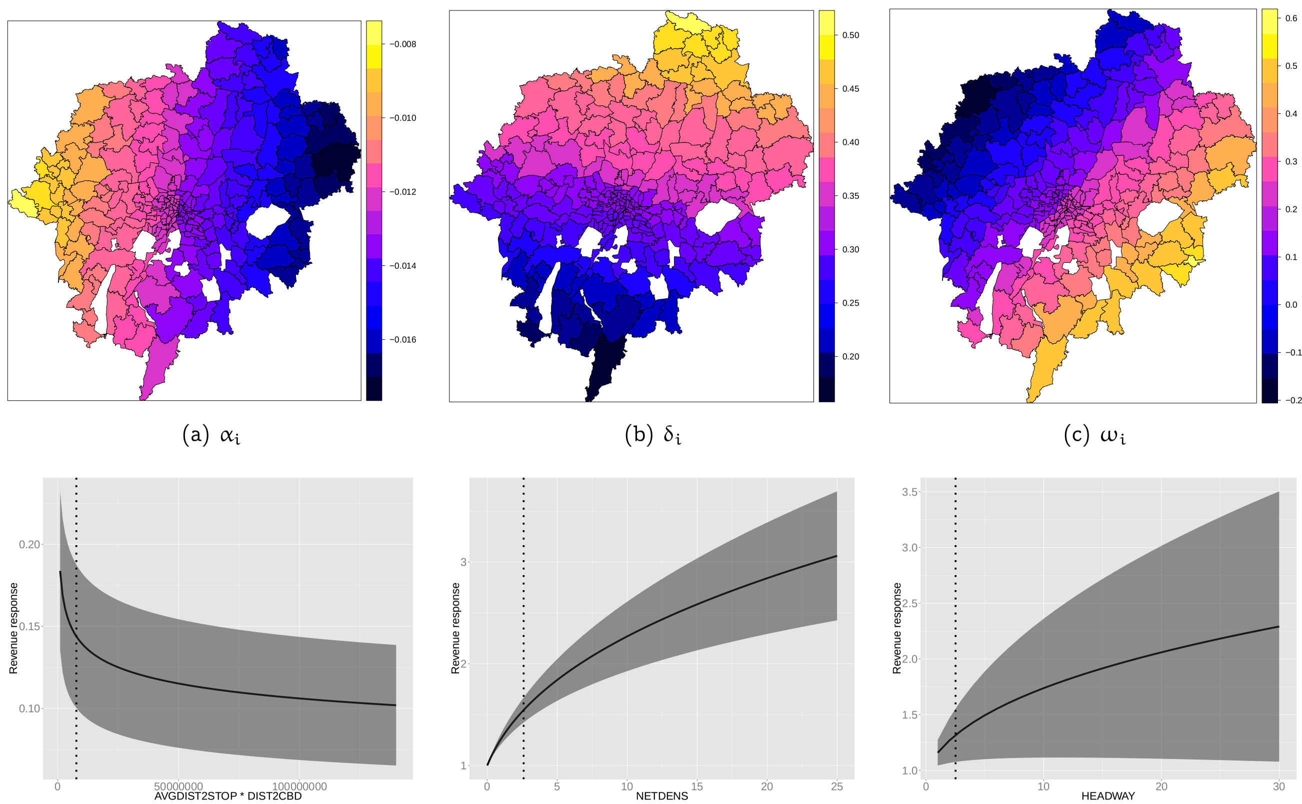
With ϵ_i defined as in (7) using $\hat{\lambda}$ (parameter 21 in table 2). Based on this we define our revenue response function as

$$\begin{aligned}
\mathrm{REV}_i = \; & e^{\mathrm{BP}_i} \\
\times \; & (1 + \mathrm{AVGDIST2STOP}_i \times \mathrm{DIST2CBD}_i)^{\alpha_i} \\
\times \; & (1 + \mathrm{NETLENGTH}_i)^{\delta_i} \\
\times \; & (1 + \mathrm{HEADWAY}_i)^{\omega_i} \\
\times \; & e^{\epsilon_i}
\end{aligned} \tag{15}$$

The service variables are factors that determine to what extend the expected base profitability (BP_i) can be skimmed. Since the corresponding weights (α_i, δ_i, and ω_i) vary spatially, the marginal extend of the skimming of the expected base profitability depends on the location. This enables managers to make local decisions. The partial functions of the revenue response function are plotted in figure 3 d to f.

Since the coefficients α_i, δ_i, and ω_i from (15) delineate the elasticities we see that the expected annual seasonal ticket revenue is inelastic (see figures 2 and 3). This is in line with the majority of empirical research on public transport demand (Cascetta, 2009) as well as empirical findings in other fields of research (see Fort 2004, for example). The empirical findings of section 4.1 show that only the impact of AVGDIST2STOP $\times$ DIST2CBD (α_i) and HEADWAY (ω_i) varies significantly over space. This is confirmed by figures 3 (d) to (f). There is only small variance in the responsiveness to NETLENGTH so that managers might consider the impact on NETLENGTH as spatially constant (figure 3 (e)). In contrast, the variance of the impact of HEADWAY on REV increases with increasing values of HEADWAY. Hence, managers have to be aware of the local impact, since the increase of HEADWAY may have large impact in some regions and only a small impact in other regions (see figure 3 (f)).

Now, in order to derive a spatially adjusted revenue management strategy, managers first identify regions with high expected base profitability from figure 4. Then, they employ figures 3 (a) to (c) to verify which service variable promises the most efficient impact on revenue for the selected regions. Neglecting the spatial drift of the impact of HEADWAY for example (i.e., using a global estimate), yields false predictions of the expected revenues and hence might foster wrong managerial decisions. Using a global estimate, managers might consider to increase HEADWAY in eastern parts of the study region because of the high expected base profitability. However, we know from figure 3c that the impact of HEADWAY is low in eastern districts and therefore the expected revenues would be smaller than assumed (using the global estimate). Further, if we take into account the negative relationship between α_i, δ_i, and ω_i (see figure 2), then it becomes obvious that an improvement of two service variables at one location might not be the best choice. For example, there are some locations where revenue response to HEADWAY is large while at the same time the revenue response to NETLENGTH is small (the most southern zip code area in figure 3 (b) and (c)). As a consequence, only an increase in HEADWAY seems to be promising in this case. In project management, decision makers might use the revenue response function in order to evaluate different local projects in terms of their expected outcome on the local level. For example, managers might choose between two projects on network expansion - one located in the north and one located in the south. An evaluation of these projects based on our revenue response functions might point managers to the "northern" project, because of the higher expected revenue response to NETLENGTH.



(a) α_i (b) δ_i (c) ω_i

(d) $(1 + \text{AVGDIST2STOP} \times \text{DIST2CBD})^{\alpha_i}$ (e) $(1 + \text{NETLENGTH})^{\delta_i}$ (f) $(1 + \text{HEADWAY})^{\omega_i}$

Figure 3: Maps of spatially varying parameters ((a) to (c)) and corresponding revenue response function plots ((d) to (f)). In maps (a) to (c) the color code represents the local parameter value. For example, ω_i takes the largest values (0.5 - 0.6) in the (south-) eastern regions (yellow), while ω_i takes negative values (-0.2 - 0) in the north-west (dark blue). Plots (d) to (f) show the revenue response to service variables (AVGDIST2STOP $\times$ DIST2CBD, NETLENGTH, and HEADWAY) holding all others constant. Consider HEADWAY, for example. A HEADWAY of 10 yields on average (solid line) a revenue response per $\exp(BP_i)$ of 1.75 (all others constant). Since we have 244 locations i and hence 244 different elasticities ω_i, we have 244 different revenue responses for a given variable value (here, HEADWAY = 10). The black solid line represents the average over the 244 revenue response values and the shaded area covers the standard deviation from the average. The dotted lines indicate the mean value of the service variables (see table 1).

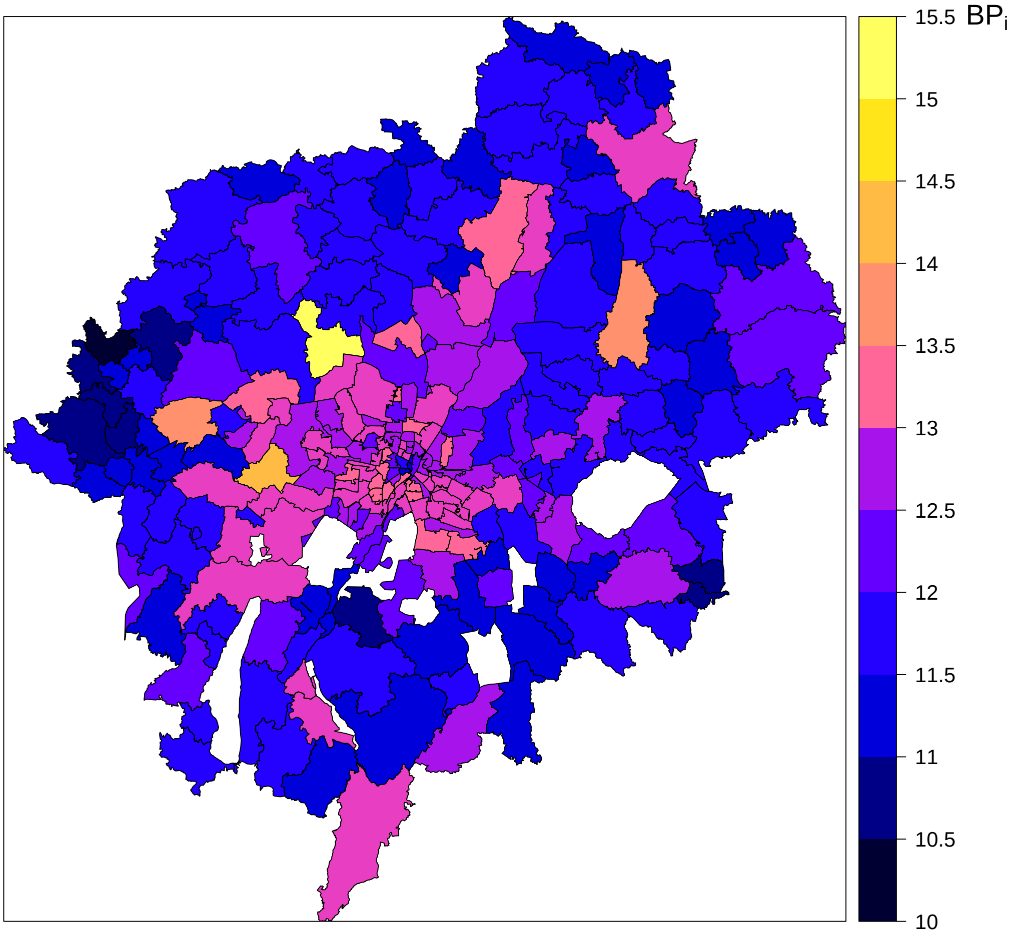


Figure 4: Expected base profitability $\mathrm{BP_i}$ as defined in (14). The color coded legend represents the values of $\mathrm{BP_i}$.

5 Conclusion

We set out to develop an approach that enables managers to simultaneously account for spatial dependencies and spatial drift while analyzing the antecedents of revenue. In particular, we focus on service quality as a major factor influencing demand and hence firms' revenue. Our approach unites spatial econometric methods and the spatial expansion method such that the estimates of our model are expected to be unbiased by spatial effects. In our empirical analysis we employ unique data (zip code area level) of a non-for-profit public transport service provider in Munich, Germany. We use the accessibility of the public transport service and the frequency of departures as proxies for service quality. We find statistically significant while inelastic impact of service quality on revenues. Most interestingly, our study provides some evidence that the coherence between service quality and revenue varies spatially (spatial drift).

In general, we learn from the study that the evaluation of service quality by the customers is likely to depend on the location.

The results show that researchers and managers must not rely on the key assumption in the traditional marketing science literature, that the behavior of a consumer is conditionally independent of the behavior of another consumer and that this behavior is spatially homogeneous. Therefore, they should employ a spatially heterogeneous revenue response function of service quality. In contrast to revenue response functions used so far, our function depends on the location (i.e., we obtain a function for every location). Managers (of public transport companies) should develop localized management strategies instead of global strategies in order to enhance seasonal ticket revenue by service variables. Having said this, managers are expected to make better decisions while implementing localized strategies to improve service quality in order to enhance revenue. Of course, our approach is valuable for researchers, too. Using our approach they are able to identify unbiased, small scale, local structures and processes which would remain undiscovered when traditional approaches are used.

Nevertheless, our research has limitations. At the same time, these limitations illustrate possible future research directions. First, the spatial expansion method used here, limits the local estimates to be a linear function function of the geographic coordinates of the observations. This means that spatial drift is strictly unidirectional. As a consequence, richer patterns of spatial drift remain unmasked. To overcome these issues more sophisticated methods should be employed (geographically weighted regression, for example). Second, the ticket price is expected to have a remarkable impact on expected revenues as well. Therefore, the analysis of revenues of several periods (panel data) would be necessary.

References

Aitken, M. (1996), "A general maximum likelihood analysis of overdispersion in generalized linear models", *Statistics and Computing*, Vol. 6, pp. 251–262.

Albers, S. (2011), "Optimizable and implementable aggregate response modeling for marketing decision support", *International Journal of Research in Marketing*, Vol. 29, pp. 111–122.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Studies in Operational Regional Science, Kluwer.

Anselin, L. (2003), "Spatial externalities, spatial multipliers, and spatial econometrics", *International Regional Science Review*, Vol. 26, pp. 153–166.

Babakus, E., Bienstock, C.C. and Van Scotter, J.R. (2004), "Linking perceived quality and customer satisfaction to store traffic and revenue growth", *Decision Sciences*, Vol. 35, No. 4, pp. 713–737.

Bass, F.M. (2004), "Comments on: A new product growth for model consumer durables the bass model", *Management Science*, Vol. 50, No. 12, pp. 1833–1840.

Beer, C. and Riedl, A. (2012), "Modeling spatial externalities in panel data: the spatial Durbin model revisited", *Papers in Regional Science*, Vol. 91, pp. 299–319.

Berger, P.D., Bolton, R.N., Bowman, D., Briggs, E., Kumar, V., Parasuraman, A. and Terry, C. (2002), "Marketing actions and the value of customer assets: A framework for customer asset management", *Journal of Service Research*, Vol. 5, No. 1, pp. 39–54.

Bitter, C., Mulligan, G. and Dall Erba, S. (2007), "Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method", *Journal of Geographical Systems*, Vol. 9, pp. 7–27.

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

Bollinger, B. and Gillingham, K. (2012), "Peer effects in the diffusion of solar photovoltaic panels", *Marketing Science*, Vol. 31, No. 6, pp. 900–912.

Bradlow, E., Bronnenberg, B., Russell, G., Arora, N., Bell, D., Duvvuri, S., Hofstede, F., Sismeiro, C., Thomadsen, R. and Yang, S. (2005), "Spatial models in marketing", *Marketing Letters*, Vol. 16, pp. 267–278.

Bronnenberg, B. (2005), "Spatial models in marketing research and practice", *Applied Stochastic Models in Business and Industry*, Vol. 21, pp. 335–343.

Bronnenberg, B. and Mahajan, V. (2001), "Unobserved retailer behaviour in multimarket data: Joint spatial dependence in market shares and promotion variables", *Marketing Science*, Vol. 20, pp. 284–299.

Brown, A. (2002), "Do you come from Manchester?: A postcode analysis of the location of Manchester United and Manchester City season ticket holder, 2001", Tech. rep., Manchester Metropolitan University.

Carman, J.M. (1990), "Consumer perceptions of service quality: An assessment of the servqual dimensions", *Journal of Retailing*, Vol. 66, No. 1, pp. 33–55.

Cascetta, E. (2009), *Transportation System Analysis*, Vol. 29 of *Optimization and its Applications*, Springer, 2 edn.

Casetti, E. (1972), "Generating models by the expansion method: Applications to geographic research", *Geographical Analysis*, Vol. 4, pp. 81–91.

Casetti, E. (2010), "Expansion method, dependency, and multimodeling", in *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* (M. Fischer and A. Getis, eds.), Springer, chap. 6, pp. 487 – 505.

Cervero, R. (1990), "Transit pricing research", Vol. 17, No. 2, pp. 117–139–.

Chang, T.Z. and Chen, S.J. (1998), "Market orientation, service quality and business profitability: a conceptual model and empirical evidence", *Journal of Services Marketing*, Vol. 12, No. 4, pp. 246–264.

Courty, P. and Pagliero, M. (2012), "The impact of price discrimination on revenue: Evidence from the concert industry", *Review of Economics and Statistics*, Vol. 94, No. 1, pp. 359–369.

Cressie, N. (1992), "Statistics for spatial data", *Terra Nova*, Vol. 4, No. 5, pp. 613–617.

Currie, G. (2004), "Gap analysis of public transport needs:measuring spatial distribution of public transport needs and identifying gaps in the quality of public transport provision", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1895, No. -1, pp. 137–146.

Dana, J.D.J. (2001), "Monopoly price dispersion under demand uncertainty", *International Economic Review*, Vol. 42, No. 3, pp. 649–670.

Dekimpe, M.G. and Hanssens, D.M. (1999), "Sustained spending and persistent response: A new look at long-term marketing profitability", *Journal of Marketing Research*, Vol. 36, No. 4, pp. 397–412.

Diggle, P. (1983), *Statistical analysis of spatial point patterns*, Arnold.

Dolnicar, S. and Lazarevski, K. (2009), "Marketing in non-profit organizations: an international perspective", *International Marketing Review*, Vol. 26, No. 3, pp. 275–291.

Drexl, A. and Haase, K. (1999), "Fast approximation methods for sales force deployment", *Management Science*, Vol. 45, No. 10, pp. 1307–1323.

Du, R.Y. and Kamakura, W.A. (2011), "Measuring contagion in the diffusion of consumer packaged goods", *Journal of Marketing Research*, Vol. 48, No. 1, pp. 28–47.

Dunlap, W.P. and Kemery, E.R. (1987), "Failure to detect moderating effects: Is multicollinearity the problem?", *Psychological Bulletin*, Vol. 102418-420, No. 3.

Eboli, L. and Mazzulla, G. (2008), "A stated preference experiment for measuring service quality in public transport", *Transportation Planning and Technology*, Vol. 31, No. 5, pp. 509–523.

Eckard, E.W. and Smith, M.A. (2012), "The revenue gains from multi-tier ticket pricing: Evidence from pop music concerts", *Manage. Decis. Econ.*, Vol. 33, No. 7-8, pp. 463–473.
URL: *http://dx.doi.org/10.1002/mde.2549*

FitzRoy, F. and Smith, I. (1999), "Season ticktes and the demand for public transport", *Kyklos*, Vol. 52, pp. 219–238.

Forrest, D., Simmons, R. and Feehan, P. (2002), "A spatial cross-sectional analysis of the elasticity of demand for soccer", *Scottish Journal of Political Economy*, Vol. 49, pp. 336–355.

Fort, R. (2004), "Inelastic sports pricing", *Managerial and Decision Economics*, Vol. 25, No. 2, pp. 87–94.

Foster, S. and Gorr, W. (1986), "An adaptive filter for estimating spatially varying parameters: Application to modeling police hours spent in response to calls for service", *Management Science*, Vol. 32, pp. 878–889.

Fotheringham, A., Charlton, M. and Brunsdon, C. (1997), "Two techniques for exploring nonstationarity in geographical data", *Geographical Systems*, Vol. 4, pp. 59–82.

Fotheringham, A.S. and Brunsdon, C. (1999), "Local forms of spatial analysis", *Geographical Analysis*, Vol. 31, No. 4, pp. 340–358.

Garber, T., Goldenberg, J., Libai, B. and Muller, E. (2004), "From density to destiny: Using spatial dimension of sales data for early prediction of new product success", *Marketing Science*, Vol. 23, No. 3, pp. 419–428.

Goldstein, H. (1987), *Multilevel Models in Educational and Social Research*, Oxford University Press.

Groenroos, C. (1984), "A service quality model and its marketing implications", *European Journal of Marketing*, Vol. 18, No. 4, pp. 36–44.

Haase, K. and Müller, S. (2014), "Upper and lower bounds for the sales force deployment problem with explicit contiguity constraints", *European Journal of Operational Research*, Vol. 237, No. 2, pp. 677–689.
URL: *http://www.sciencedirect.com/science/article/pii/S037722171400100*

Hansses, D.M., Parsons, L.J. and Schultz, R.L. (2001), *Market Response Models*, Kluwer.

Harris, R., Moffat, J. and Kravtsova, V. (2011), "In search of "W"", *Spatial Economic Analysis*, Vol. 6, pp. 249–270.

Hawkins, D., Roupe, D. and Coney, K. (1981), "The influence of geographic subcultures in the United States", *Advances in Consumer Research*, Vol. 8, pp. 713–717.

Helbich, M., Brunauer, W., Hagenauer, J. and Leitner, M. (2012), "Data-driven regionalization of housing markets", *Annals of the Association of American Geographers*, Vol. 103, No. 4, pp. 871–889.

Hensher, D.A. and Stanley, J. (2003), "Performance-based quality contracts in bus service provision", *Transportation Research Part A: Policy and Practice*, Vol. 37, No. 6, pp. 519–538.

Heo, C.Y. and Lee, S. (2011), "Influences of consumer characteristics on fairness perceptions of revenue management pricing in the hotel industry", *International Journal of Hospitality Management*, Vol. 30, No. 2, pp. 243–251.

Hruschka, H. (2006), "Relevance of functional flexibility for heterogeneous sales response models: A comparison of parametric and semi-nonparametric models", *European Journal of Operational Research*, Vol. 174, No. 2, pp. 1009–1020.

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

Hume, M., Sullivan Mort, G., Liesch, P.W. and Winzar, H. (2006), "Understanding service experience in non-profit performing arts: Implications for operations and service management", *Journal of Operations Management*, Vol. 24, No. 4, pp. 304–324.

Lam, S.S.K. (1997), "Servqual: A tool for measuring patients' opinions of hospital service quality in hong kong", *Total Quality Management*, Vol. 8, No. 4, pp. 145–152.

LeSage, J. and Pace, R. (2009), *Introduction to Spatial Econometrics*, Statistics: Textbooks and Monographs, CRC Press.

Liao, F. and Wei, Y. (2012), "Modeling determinants of urban growth in Dongguan, China: a spatial logistic approach", *Stochastic Environmental Research and Risk Assessment*, Vol. 1, pp. 1–16.

Lowengart, O., Mizrahi, S. and Yosef, R. (2003), "Effect of consumer characteristics on optimal reference price", *Journal of Revenue and Pricing Management*, Vol. 2, No. 3, pp. 201–215.

Macedo, I.M. and Pinho, J.C. (2006), "The relationship between resource dependence and market orientation: The specific case of non-profit organisations", *European Journal of Marketing*, Vol. 40, No. 5/6, pp. 533–553.

Maegi, A.W. (2003), "Share of wallet in retailing: the effects of customer satisfaction, loyalty cards and shopper characteristics", *Journal of Retailing*, Vol. 79, No. 2, pp. 97–106.

Matas, A. (2004), "Demand and revenue implications of an integrated public transport policy: The case of madrid", *Transport Reviews*, Vol. 24, No. 2, pp. 195–217.

McDonald, H. (2010), "The factors influencing churn rates among season ticket holders: An empirical analysis", *Journal of Sport Management*, Vol. 24, pp. 676–701.

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

Müller, S. and Rode, J. (2013), "The adoption of photovoltaic systems in wiesbaden, germany", *Economics of Innovation and New Technology*, Vol. 22, No. 5, pp. 519–535.
URL: *http://dx.doi.org/10.1080/10438599.2013.804333*

Müller, S., Wilhelm, P. and Haase, K. (2013), "Spatial dependencies and spatial drift in public transport seasonal ticket revenue data", *Journal of Retailing and Consumer Services*, Vol. 20, No. 3, pp. 334–348.

Nash, C.A. (1978), "Management objectives, fares and service levels in bus transport", *Journal of Transport Economics and Policy*, Vol. 12, No. 1, pp. 70–85.

Norton, J.A. and Bass, F.M. (1987), "A diffusion theory model of adoption and substitution for successive generations of high-technology products", *Management Science*, Vol. 33, No. 9, pp. 1069–1086.

Olson, J.R., Belohlav, J.A. and Boyer, K.K. (2005), "Operational, economic and mission elements in not-for-profit organizations: the case of the chicago symphony orchestra", *Journal of Operations Management*, Vol. 23, No. 2, pp. 125–142.

Paez, A. (2005), "Local analysis of spatial relationships: A comparison of GWR and the expansion method", in *Computational Science and Its Applications ICCSA 2005* (O. Gervasi, M. Gavrilova, V. Kumar, A. Lagana , H. Lee, Y. Mun, D. Taniar and C. Tan, eds.), Springer Berlin, Vol. 3482 of *Lecture Notes in Computer Science*, pp. 631–637.

Parker, P.M. and Tavassoli, N.T. (2000), "Homeostasis and consumer behavior across cultures", *International Journal of Research in Marketing*, Vol. 17, pp. 33–53.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J. and White, P. (2006), "The demand for public transport: The effects of fares, quality of service, income and car ownership", *Transport Policy*, Vol. 13, No. 4, pp. 295–306.

Prag, J. and Casavant, J. (1994), "An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry", Vol. 18, No. 3, pp. 217 – 235.

Rust, R.T., Moorman, C. and Dickson, P.R. (2002), "Getting return on quality: Revenue expansion, cost reduction, or both?", *Journal of Marketing*, Vol. 66, No. 4, pp. 7–24.

Rust, R.T. and Zahorik, A.J. (1993), "Customer satisfaction, customer retention, and market share", *Journal of Retailing*, Vol. 69, No. 2, pp. 193–215.

Rust, R.T., Zahorik, A.J. and Keiningham, T.L. (1995), "Return on quality (roq): Making service quality financially accountable", *Journal of Marketing*, Vol. 59, pp. 58–70.

Skiera, B. and Albers, S. (1998), "COSTA: Contribution Optimizing Sales Territory Alignment", *Marketing Science*, Vol. 17, No. 3, pp. 196–213.

Sohail, M.S. (2003), "Service quality in hospitals: more favourable than you might think", *Managing Service Quality*, Vol. 13, No. 3, pp. 197 – 206.

Storbacka, K., Strandvik, T. and Groenroos, C. (1994), "Managing customer relationships for profit: The dynamics of relationship quality", *International Journal of Service Industry Management*, Vol. 5, No. 5, pp. 21–38.

Talluri, K. and Van Ryzin, G. (2005), *The Theory and Practice of Revenue Management*, Springer Verlag.

Tobler, W. (1970), "A computer movie simulating urban growth in the Detroit region", *Economic Geography*, Vol. 46, pp. 234–240.

Tollefson, J. and Lessig, V. (1978), "Aggregation criteria in normative market segmentation theory", *Journal of Marketing Research*, Vol. 15, pp. 346–355.

Upton, G. and Fingleton, B. (1985), *Spatial data analysis by example. Volume 1: Point pattern and quantitative data*, John Wiley and Sons Ltd.

Vandamme, R. and Leunis, J. (1993), "Development of a multiple-item scale for measuring hospital service quality", *International Journal of Service Industry Management*, Vol. 4, No. 3, pp. 30–49.

Villanueva, J., Yoo, S. and Hanssens, D.M. (2008), "The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth", *Journal of Marketing Research*, Vol. 45, No. 1, pp. 48–59.

Wheeler, D. and Paez, A. (2010), "Geographically weighted regression", in *Handbook of Applied Spatial Analysis* (M. Fischer and A. Getis, eds.), Springer, chap. Geographically Weighted Regression, pp. 461–486.

Woodside, A.G., Frey, L.L. and Daly, R.T. (1989), "Linking service quality, customer satisfaction, and behavioral intention.", *Journal of Health Care Marketing*, Vol. 9, No. 4, pp. 5–17.

Zeithaml, V.A., Berry, L.L. and Parasuraman, A. (1996), "The behavioral consequences of service quality.", *Journal of Marketing*, Vol. 60, No. 2, pp. 31–46.

6 Appendix

	log(REV)	SLM		SEM		SDM	
Nr.	Variable	$\hat{\beta}$	SE	$\hat{\beta}$	SE	$\hat{\beta}$	SE
1	Intercept	9.22621	1.03296 ***	11.13861	0.72446 ***	8.73901	2.28889 ***
2	POP	0.75936	0.06141 ***	0.80846	0.06030 ***	0.80197	0.06234 ***
3	POPDENS	0.03067	0.02054	0.03336	0.02127	-0.02003	0.03066
4	log(PCTEMPL)	0.71188	0.28443 *	0.82956	0.27858 **	0.66665	0.29252 *
5	PCTUNEMPL	-8.70816	2.22928 ***	-9.85228	2.36527 ***	-11.43351	2.47839 ***
6	sqrt(BUYPOW)	0.00783	0.00300 **	0.00878	0.00296 **	0.00859	0.00305 **
7	CARSPCAP_mod	-0.17896	0.07679 *	-0.12218	0.07530	-0.12481	0.07786
8	$(\text{CARSPCAP\_mod})^2$	-0.06666	0.03440 ·	-0.04126	0.03407	-0.03054	0.03480
9	$(\text{CARSPCAP\_mod})^3$	0.02839	0.01406 *	0.01607	0.01373	0.01855	0.01439
10	RETAILDENS	-0.00106	0.00173	-0.00134	0.00166	-0.00057	0.00178
11	FIRMSPCAP	0.06406	1.31394	0.05102	1.32380	-1.02905	1.67090
12	log(AVGDIST2STOP × DIST2CBD)	-0.01018	0.00406 *	-0.01242	0.00386 **	-0.00877	0.00424 *
13	log(AVGDIST2STOP × DIST2CBD) × LO × 0.001	-0.00094	0.00019 ***	-0.00105	0.00023 ***	-0.00223	0.00136 ·
14	log(AVGDIST2STOP × DIST2CBD) × LA × 0.001	0.00012	0.00019	0.00009	0.00024	0.00225	0.00161
15	log(NETLENGTH)	0.33164	0.05487 ***	0.32620	0.05588 ***	0.35364	0.05674 ***
16	log(NETLENGTH) × LO × 0.001	0.00038	0.00247	0.00084	0.00243	0.00114	0.00241
17	log(NETLENGTH) × LA × 0.001	0.00291	0.00299	0.00328	0.00304	0.00142	0.00320
18	log(HEADWAY)	0.19927	0.05065 ***	0.21058	0.05222 ***	0.17473	0.05545 **
19	log(HEADWAY) × LO × 0.001	0.00973	0.00353 **	0.00716	0.00375 ·	0.00646	0.00383 ·
20	log(HEADWAY) × LA × 0.001	-0.00634	0.00422	-0.00703	0.00449	-0.00731	0.00495
lagged	variables					$\hat{\gamma}$	SE
21	POP					-0.42949	0.13910 **
22	POPDENS					0.02200	0.04876
23	log(PCTEMPL)					-0.15373	0.69200
24	PCTUNEMPL					5.67207	4.39738
25	sqrt(BUYPOW)					-0.00251	0.00747
26	CARSPCAP_mod					-0.33117	0.18142 ·
27	$(\text{CARSPCAP\_mod})^2$					-0.12253	0.06879 ·
28	$(\text{CARSPCAP\_mod})^3$					0.06380	0.03144 *
29	RETAILDENS					0.00407	0.00491
30	FIRMSPCAP					-2.67766	3.83587
31	log(1 + AVGDIST2STOP × DIST2CBD)					-0.04520	0.09298
32	log(1 + AVGDIST2STOP × DIST2CBD) × LO × 0.001					0.00142	0.00144
33	log(1 + AVGDIST2STOP × DIST2CBD) × LA × 0.001					-0.00242	0.00170
34	log(1 + NETLENGTH)					-0.18389	0.11527
35	log(1 + NETLENGTH) × LO × 0.001					0.00230	0.00578
36	log(1 + NETLENGTH) × LA × 0.001					-0.00232	0.00643
37	log(1 + HEADWAY)					0.09508	0.11103
38	log(1 + HEADWAY) × LO × 0.001					0.00167	0.00667
39	log(1 + HEADWAY) × LA × 0.001					0.01097	0.00913
40	λ			0.38885	0.08211 ***		
41	ρ	0.15164	0.05279 ***			0.31702	0.085188 ***
	AIC	434.22		425.49		440.54	
	Breusch-Pagan-Test	24.6845		20.2008		34.0965	
	(p-value)	0.1712		0.3826		0.6506	
	Moran's $\mathcal{I}$	0.095		-0.005		0.003	
	(p-value)	0.006		0.507		0.435	
	Lagrange multiplier	8.17		16.24		16.82	
	(p-value)	0.004		0.000		0.000	

Table 3: Comparison of local spatial models: spatial lag model (SLM), spatial error model (SEM) and spatial Durbin model (SDM). Number of observations: 244 for all three models. Significance codes: 0 ***, 0.001 **, 0.01 *, 0.05 ·

2.1.2 Disaggregate Data

1. Müller, S., Haase, K., Seidel, F. (2012): **Exposing unobserved spatial similarity: Evidence from German school choice data**. *Geographical Analysis* **44**(1), 65-86.

2. Müller, S., Tscharaktschiew, S., Haase, K. (2008): **Travel-to-school mode choice modelling and patterns of school choice in urban areas**. *Journal of Transport Geography* **16**(5), 342-357.

3. Müller, S., Rode, J. (2013): **The adoption of photovoltaic systems in Wiesbaden, Germany**. *Economics of Innovation and New Technology* **22**(5), 519-535.

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Transport Geography 16 (2008) 342–357
JOURNAL OF TRANSPORT GEOGRAPHY
www.elsevier.com/locate/jtrangeo

Travel-to-school mode choice modelling and patterns of school choice in urban areas

Sven Müller *, Stefan Tscharaktschiew, Knut Haase

Institute for Transport and Economics, Dresden University of Technology, Andreas-Schubert-Strasse 23, 01062 Dresden, Germany

Abstract

Because of declining enrollment and school closures in some German regions students have to choose a certain school location from a reduced set of schools. For the analysis of adverse effects of school closures on transport mode choice the patterns of school choice are specified first. It seems that proximity and the profile offered (languages as a core for example) are adequate factors. Second, the travel-to-school mode choice are modelled using a multinomial logit approach, since students might switch from low cost transport modes (cycling for instance) to modes with remarkably higher costs (public transport for instance). Here, the most influencing factors are distance, car availability and weather. Furthermore, these findings are incorporated into a case study to quantify the effects of a modal-shift (switch from one transport mode to another). For this analysis a comprehensive survey was undertaken and a method of data disaggregation and geocoding is presented.

Keywords: Multinomial logit model; Data disaggregation; School choice; Travel-to-school mode choice

1. Introduction

More and more German regions are confronted with declining enrollment numbers caused by decreasing population and negative net migration. This in turn implies the necessity to close some school locations. Students have to choose a certain school location from a reduced set of remaining schools and may face a longer way to school. Since distance strongly influences the travel-to-school mode choice, students switch from modes appropriate for short distances like cycling to modes appropriate for longer distances like public transport (modal-shift). Latest studies on travel-to-school mode choice stress the establishment of neighborhood schools and thus the preponderance of activity-related travel-modes like walking or biking due to short travel-to-school distances (Ewing et al., 2005; de Boer, 2005). Inter alia, these modes are beneficial for students' health (McDonald, 2005; McMillan, 2003). Our focus is on the economic benefit of neighborhood schools and short distances: modes like car or public transport are related to considerably higher costs in contrast to walking and biking. Moreover, neighborhood schools are desirable, because any policy which forces people to use motorized transport modes might not be appropriate within the context of climate change and peaking of global oil production. The closure of schools leads to savings for authorities in infrastructural and personnel costs, but there could be an increase in transport costs, which yields increased total costs. For an estimation of the additional costs within the framework of dynamic school network planning one has to analyze the process and the most influencing factors of school choice and travel-to-school mode choice first. This is a more complex task in urban than in rural areas. Recent studies explain school choice (in Germany) by proximity and tuition fees among others but do not cover the school's profiles – i.e. special courses (Speiser, 1993; Mahr-George, 1999; Hoxby, 2003; Schneider, 2004;

* Corresponding author. Tel.: +49 351/463 3 6815; fax: +49 351/463 3 7758.
E-mail addresses: sven.mueller@tu-dresden.de (S. Müller), knut.haase@tu-dresden.de (K. Haase).

0966-6923/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtrangeo.2007.12.004

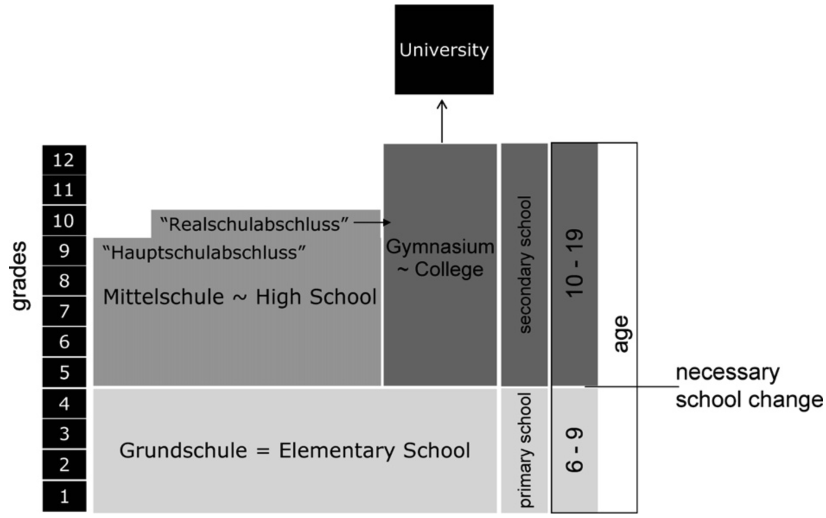


Fig. 1. Main aspects of the educational system of Saxony.

Hastings et al., 2005). We expect that students choose the school closest to their home and those who do not, choose a school with a different profile than the closest one. In this paper we analyze the consequences of a school closure in the City of Dresden, Saxony and present the results of a large empirical study ($n = 4700$).

The remainder of this article is organized as follows. In Section 2 we describe the data used and present a method of data disaggregation. This is followed by the examination of the school choice behavior (Section 3) and the modelling of travel-to-school mode choice (Section 4). In Section 5 we present an example of school closure and modal-shift for the City of Dresden. Some final remarks can be found in Section 6.

2. Data and disaggregation

In this section we depict how the survey was accomplished, what data are available and how these data are disaggregated using a commercial Geographic Information System (GIS). The data are analyzed in detail in Sections 3 and 4.

2.1. Data

This study is focused on secondary schools, particularly colleges (German = "Gymnasium"). College students are aged between 10 and 19 years (see Fig. 1). In Dresden around 45% of all secondary school students are college students (City Council of Dresden (=Landeshauptstadt Dresden, 2003). The possibility to enroll on a college or high school depends on the elementary school report (overall average grade). Our data set includes administrative areas (spatial units), the school locations, the street network, the bus and tram stops and the routes of the public transportation system of Dresden. As administrative areas we consider districts and blocks[1]. A block is bordered by streets (see Fig. 2). Note, each district consists of a unique set of blocks. Using a shortest path algorithm we have determined the street network distances between all blocks within Dresden. These distances have to be interpreted as walking distances in the absence of information about accessibility for cars around one-way street systems for instance. As this paper just considers the commute to school, the car and motorcycle do not play an important role (see also Section 4). Population data cover the age groups 10–19 years at block level for the years 2004 and 2008 (forecast). These data are needed to compute the absolute effects of modal-shift due to a school closure in 2008 compared to the situation in 2004.

In 2004, a survey was carried out covering nearly 4700 of 14000 college students at 12 of the 23 colleges in Dresden lasting from January to November including a pre-test. A short form questionnaire (two pages) was used very similar to that used by the German Federal Ministry of Transport (Federal Ministry of Transport, Building and Urban Affairs, 2002). Information was obtained of each student's home district, the school attended, age, sex, car availability and whether the student owns a driver's license as well as travel-to-school mode choice and total travel-time. The total travel-time is related to the most preferred transport mode from home to school in the summer term. Students were asked to state their preferred transport mode which is usually chosen for the way to school and back home both in winter and summer term. Fair weather was assumed to be synonymous with the summer term and bad weather with the winter term, respectively (see Fig. 3). Furthermore, the students were asked how often they use a certain mode while commuting to school within a representative week.

[1] There are 64 districts and more than 6400 blocks in Dresden.

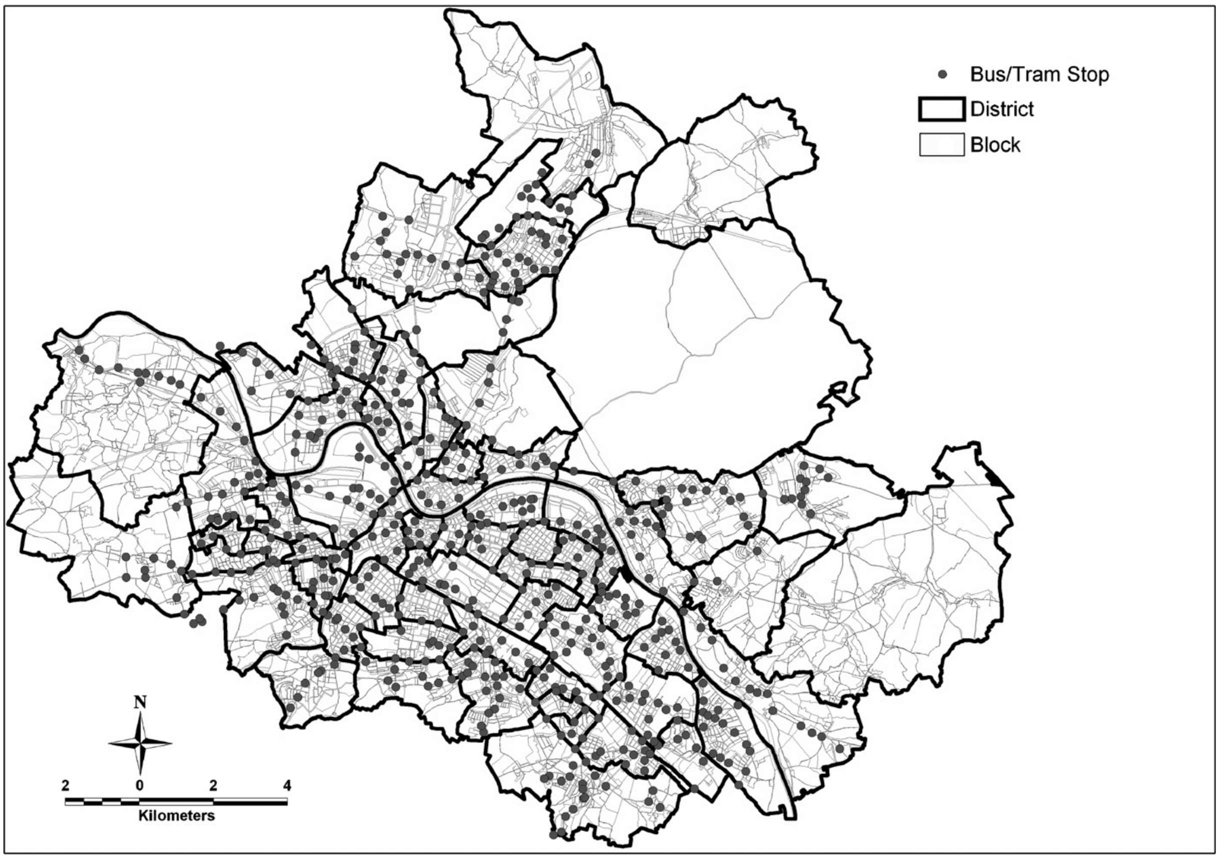


Fig. 2. The City of Dresden – Administrative areas and public transport access.

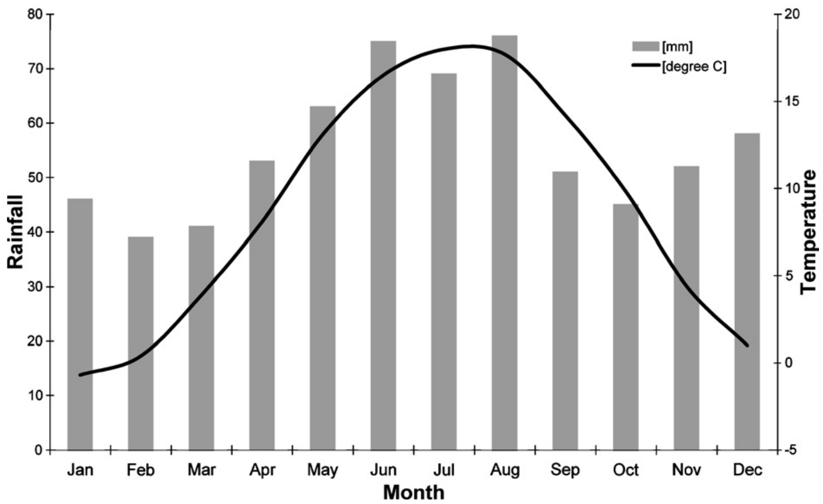


Fig. 3. Climate diagram for Dresden, Saxony.

Again, this information is available for the summer and the winter term. In case of the usage of public transportation, there is information about bus routes and stops (origin, destination and change). Moreover, the students were asked to state their waiting times (departure station, change) and access as well as egress times, which are the walking time from home to the departure station and the walking time from the destination station to school. The questionnaire ends with questions, among others, on the ticket used and the satisfaction with the level of service.

2.2. Disaggregation

Due to administrative restrictions which prohibit inquiring about detailed student addresses, a method was devised for small scale (blocks) geocoding of the survey data using a GIS. The data were collected on the scale of districts. Since distance is an important variable discriminating between most of the transport modes, data as disaggregated as possible are needed in order to obtain a good approximation of exact distances for each student. Several authors stress the use of disaggregated data for distance related analysis (Goodchild, 1979; Bach, 1981; Fotheringham et al., 1995; Longley et al., 2001). There are only a few methods that deal with data disaggregation for transport surveys, but some work has been done in other fields of research (Gimona et al., 2000; Spiekermann and Wegener, 2000; Van der Horst, 2002; Greaves et al., 2004; Oosterhaven, 2005).

Most of the students use public transportation on their way to school (50–60%, see Section 4). Thus, the departure bus or tram stop used and the time needed to get there from home are known. Now, let us assume a student is located in district A (see Fig. 4). Taking into account an average walking speed of 4 km/h, one can determine a student specific isochrone around the stated departure bus or tram stop. So, just a few blocks possibly contain the home of the student. Blocks without population are eliminated. The number of possible blocks could be reduced by considering the bus or tram route chosen by the student. This is based on the assumption that most of the students use the bus stop of the chosen line which is closest to their home. However, the situation arose that more than one possible block has to be taken into account for allocating the specific student although using all information available. Students with comparable properties (travel-time, home district) are allocated to the considered blocks relative to the population of the specific age-group.

Regarding students who never commute to school by public transportation this detailed information is not available. In this case the following procedure has to be used: Imagine another student living in district A and the school attended is located in district B (see Fig. 5). Again, the information of the commuting mode is available from our survey data as well as the total travel-time. We assume a transport mode specific average speed for walking of 4 km/h and for cycling of 12 km/h (Federal Environment Agency Germany, 2007). The speed limit for cars and motorcycles is usually 50 km/h. Due to traffic lights and congestion we suppose an average speed of 30 km/h for cars and motorcycles in (German Aerospace Center, 2007). We expect these average speeds to be sufficient for the geocoding process. Using the average speed and the stated travel-times, we are able to determine a student specific

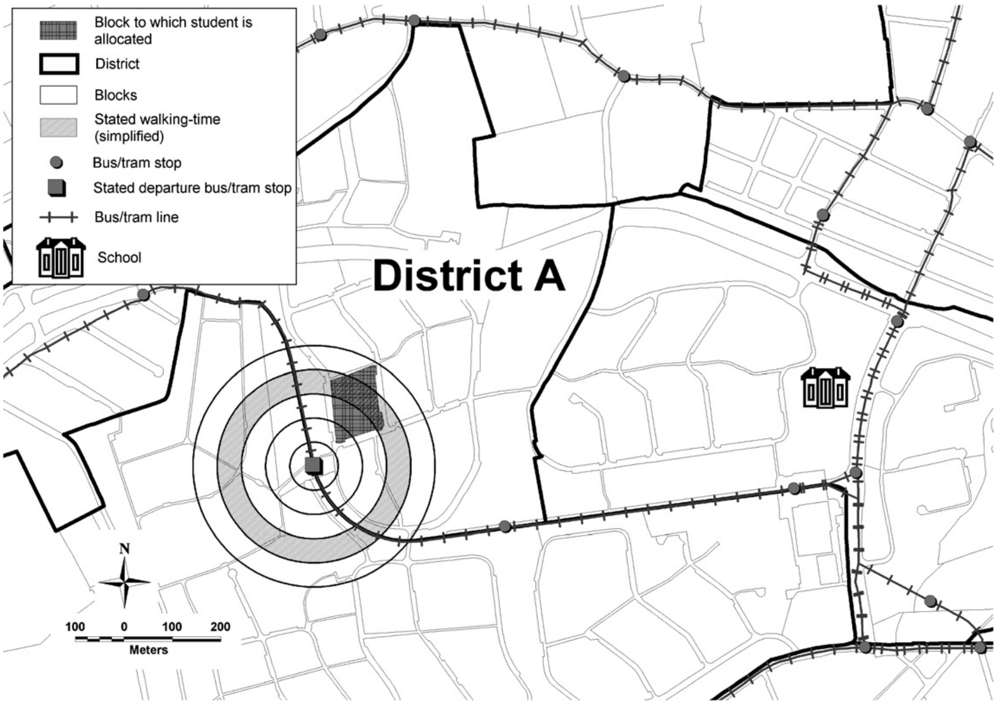


Fig. 4. Allocation of students using public transportation.

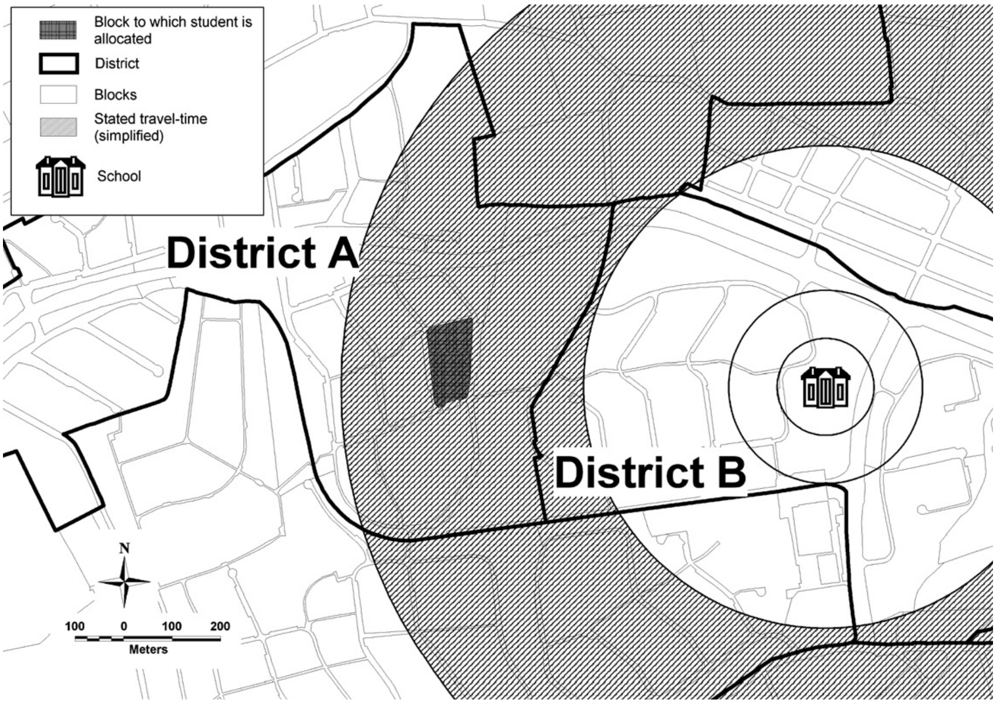


Fig. 5. Allocation of students not using public transportation.

isochrone around the school attended. For the modes biking and in particular car/motorcycling these isochrones are larger than those around bus stops (see above). According to this, there is more uncertainty about the correctness of the allocation of students to blocks in this case. However, there is just a very small percentage (6–10%) of students who commute to school by car or motorcycle (see Section 4). But we expect that possible errors will be limited due to the extent of the sample.

3. Patterns of school choice

In Saxony, no regulations exist restricting the choice of schools. So, there are no intrinsic school-districts and students are free to choose a certain school location. Several surveys yield proximity and the authority responsible (private or public school) as two very important factors of school choice. Others are the reputation of schools and tuition fees, for example Speiser (1993), Mahr-George (1999), Hoxby (2003), Schneider (2004) and Hastings et al. (2005). We expect that the school's profiles could have influence on school choice as well. In this study we will focus on distance, the school profile and the authority responsible to determine the school location choice, since most of the other influencing factors stated in the literature cannot be applied here due to the lack of data or unimportance (i.e. average household income and tuition fees). With regard to profile we differentiate between schools with a common profile and schools with an unique profile. A common profile is offered by several colleges. So these schools are substitutable by others (mathematics/science for example). A unique profile[2] – i.e. advanced-level/core languages – is only offered by one specific school. For an overview of school locations and profiles offered, see Fig. 6.

3.1. School catchment area and proximity

We have to determine the surrounding catchment area of each school first. Therefore, the nearest school location has been verified for each block. Because students will not always realize this strictly drawn border, we have added two zones with virtually reduced distances (zone 2: −1000 m and zone 3: −2000 m). Consequently, the distances of blocks within zones 2 and 3 are minimal to the specific school location (see Fig. 7). Table 1 shows the percentage of students within the corresponding zones for all schools of our sample. In example, 84.8% of all students attending Klotzsche college are located in zone 1 of this college. The surrounding catchment area of each school consists of three zones as defined above. We believe that within this area students recognize the specific school as the closest one. Two main patterns are evident:

- Students attending schools with a common profile mostly are located in the surrounding catchment area. Thus, one could assume that proximity is an important factor for school choice.

[2] Which is comparable to magnet schools.

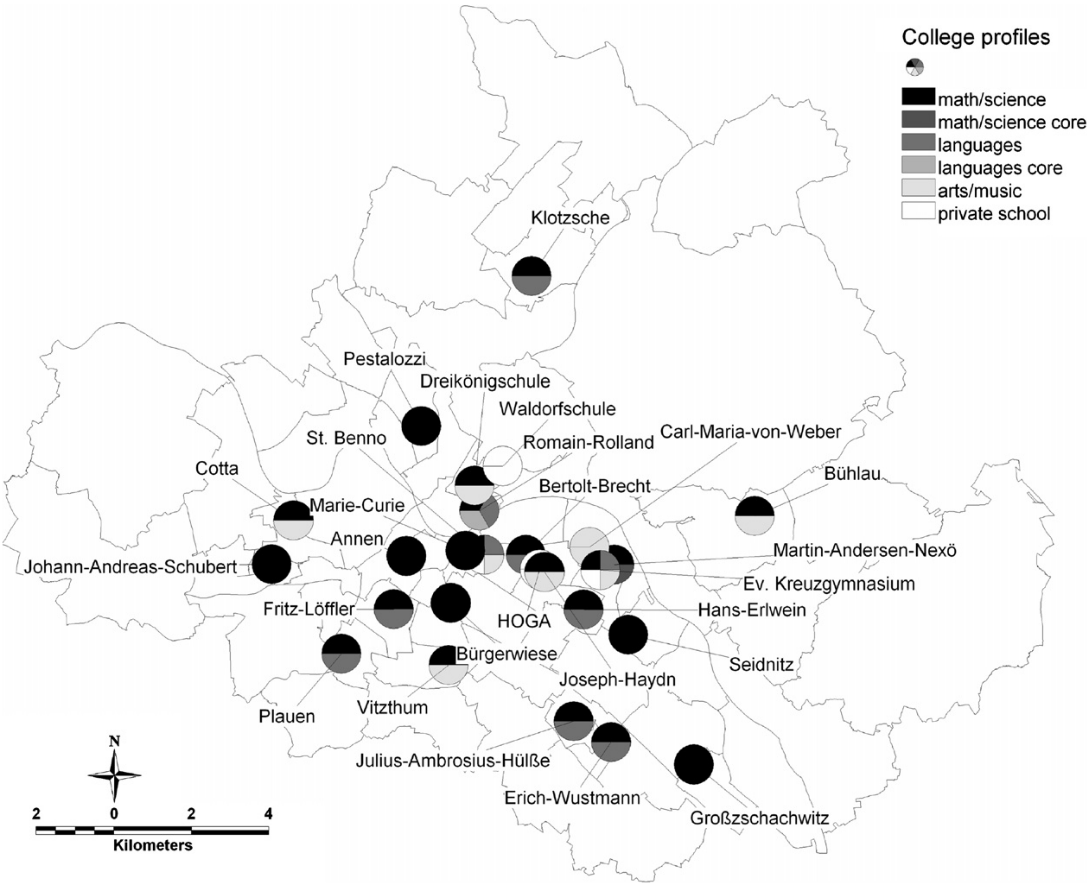


Fig. 6. Colleges in Dresden.

- For schools with a unique profile and for private schools this does not hold true. It seems to be that a unique profile, or a private school, reduces the importance of proximity.

Outliers in both groups – Marie-Curie College and Martin-Andersen-Nexö College, for example – are due to the topology of the school network (see Fig. 6). Schools located in an area with a high density of school locations (spatial cluster) obtain smaller surrounding catchment areas and thus fewer students within them. At the outskirts these catchment areas are larger and possibly contain more students. For students located in blocks close to a spatial cluster it is not always obvious which school location is the closest one. Within a cluster there are many choice alternatives available within remarkable proximity (Bertold-Brecht College for instance, see Fig. 6). It is reasonable to assume that if the closest school location does not match the preferences of students for a combination of profiles, etc., a school within the cluster does so. For students located in a spatial cluster, proximity is less important than for those students located at the outskirts. At the same time other properties like profile, are more important for the decision which school to enroll at.

3.2. School profile and school choice

For a deeper investigation of the influence of profile and the authority responsible we consider Table 2. It shows the distribution of those students who attend schools which are not the closest one. Over all most of the students (80%) choose schools with a different profile and/or a different authority responsible. Let us take Klotzsche College as an example: for 100 students Klotzsche College is the closest one, but they actually choose a different school (sum 1–5). Eighty-eight (0.88, see last column) of them choose colleges with a different profile offered and/ or a different authority responsible (sum 2–5). Twenty of these 88 students choose colleges with an alternative profile (column 3). Over all colleges nearly 70% of the students who choose a different school than the closest one, choose a school with a different profile. Therefore, we assume that profile and the authority responsible are two factors which influence the choice of a certain school. Those 20% of students who attend a

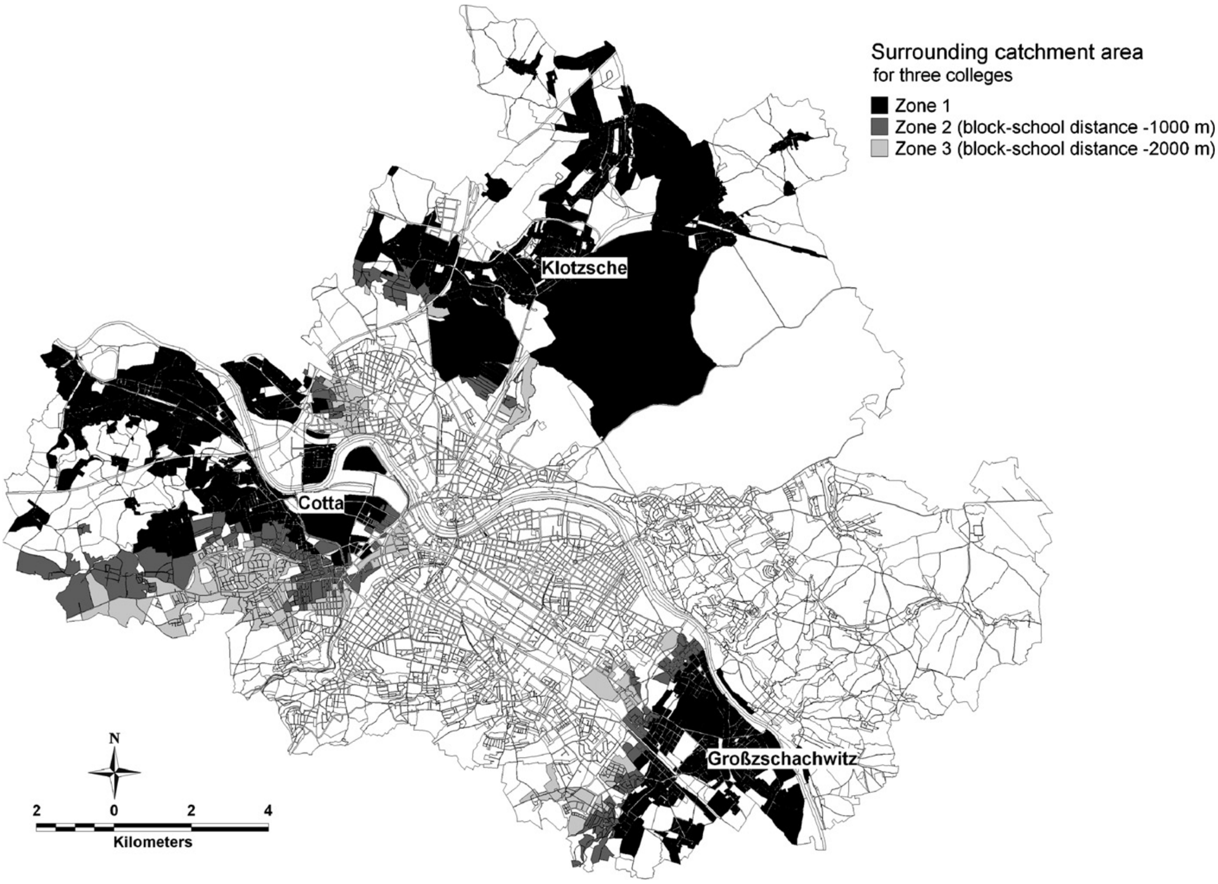


Fig. 7. Surrounding catchment areas: blocks to which the three exemplary colleges are the closest one.

Table 1
School-specific catchment areas

College	Zone 1	Zone 2	Zone 3	Other	Sum zones 1–3
Klotzsche	84.8	0.5	0	14.7	85.3
Großzschachwitz	68	12.9	1.7	17.4	82.6
Cotta	50	23.3	17.5	9.3	90.7
Julius-Ambrosius-Hülße	32.6	28.7	12.4	26.4	73.6
Fritz-Löffler	35.8	11.2	23	30	70
Plauen	46	9.4	8.9	35.7	64.3
Vitzthum	48.6	5.7	17.7	28	72
Romain-Rolland	9.4	18.8	3.6	68.2	31.8
Marie-Curie	8.3	14.8	11.6	65.3	34.7
Martin-Andersen-Nexö	15.1	26.1	17.6	41.2	58.8
Joseph-Haydn	7.3	17	32.1	43.6	56.4
St. Benno[a]	0.4	4.1	10.2	85.3	14.7

Values are given in percentage.
Only schools which are covered in the survey by at least 150 students are considered.
Colleges written bold are magnet schools (unique profile).
[a] Private school.

school with the same properties as their closest one, may be attracted by factors not considered here (extracurricular program for example). Another possible reason may be inner-city student migration.

Table 2
Conditional school choice of students choosing a different school than the closest one

# of students in zone 1 of college	Attending a college with					Sum 2–5	Sum 1–5	Sum 2–5/sum 1–5
	1	2	3	4	5			
Klotzsche	12	29	20	0	39	88	100	.88
Großzschachwitz	49	0	169	0	30	199	248	.80
Cotta	23	0	20	0	17	37	60	.62
Julius-Ambrosius-Hüßle	37	0	31	0	32	63	100	.63
Fritz-Löffler	0	8	81	0	11	100	100	1
Plauen	40	4	67	0	25	96	136	.71
Vitzthum	1	50	97	0	20	167	168	.99
Romain-Rolland	0	3	12	0	13	28	28	1
Marie-Curie	1	0	21	0	6	27	28	.96
Martin-Andersen-Nexö	47	5	19	0	33	57	104	.55
Joseph-Haydn	23	0	20	0	16	36	59	.61
St. Benno[a]	0	0	0	24	0	24	24	1
Average (total numbers)	19.4	8.3	46.4	2	20.2	76.8	96.3	.80
Average (relative numbers, %)	20	9	48	2	21	80		

1: Identical profile and authority responsible; 2: additional feature (additional language i.e.); 3: alternative profile; 4: alternative authority responsible; 5: alternative profile and authority responsible.
Colleges written bold are magnet schools (unique profile).
[a] Private school.

4. Travel-to-school mode choice modelling

Regarding the travel-to-school mode choice, the mode is a categorical variable. We suggest a student chooses the transport mode with the highest utility. So we revert to multinomial logistic regression since this is based on utility theory and appropriate for categorical data analysis. The logit approach has been widely used in fields of transport modelling. The modeler assumes the utility U_{ij} of a transport mode i (walking, cycling, public transport and car/motorcycle) to a student j, and includes a deterministic component V_{ij} and an additive random component ε_{ij}

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{1}$$

Here, the deterministic component of the utility function is linear in parameters. Assuming that the random component, which represents errors in the modeler's ability to represent all the elements that influence the utility of a transport mode to an individual, is independently and identically Gumbel-distributed across individuals and transport modes, the multinomial logit model (MNL) is as follows:

$$P_{ij} = \frac{\exp V_{ij}}{\sum_{i'=1}^{I} \exp V_{i'j}} \tag{2}$$

where P_{ij} is the probability that transport mode i is chosen by student j and I is the set of different transport modes. The closed form of the MNL makes it straightforward to estimate (maximum likelihood estimation procedure),

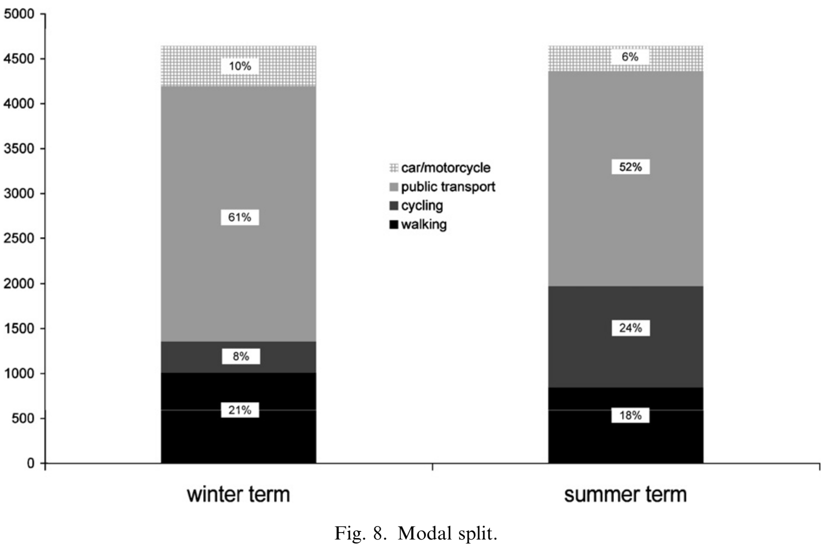


Fig. 8. Modal split.

interpret and use. Detailed work on theory, shortcomings and some applications can be found in the literature (McFadden, 1973; Ben-Akiva and Lerman, 1985; Bhat, 1997; Koppelmann and Sethi, 2000; Greene, 2003). Recent studies concerning travel-to-school mode choice utilizing MNL have been focused on urban form, built environment and distance (McMillan, 2003; Black et al., 2004) as well as travel-time (Woodside et al., 2002; Ewing et al., 2004; McDonald, 2005). Ewing et al. (2005) and de Boer (2005) focused on the relationships between travel-to-school mode choice and school location, safety, and vehicle emission.

We like to analyze the influence of the variables distance, car availability, season or weather, respectively, on commuting mode choice. Age is considered as an explanatory variable as well, admittedly it turned out to be not significant for public transport. Distance is a continuous variable measured in kilometers. Car availability (all time/not all time) and weather (fair/bad) are dummy variables. Car availability means, whether the student has the possibility of travelling to school by car. This includes the possibility of the student being passenger while the mother for instance drives the car. Car availability equals one, if the student has the possibility of commuting to school by car every day. We just consider a few variables for forecasting purposes and for an easy interpretation of the relationships.

Table 3 displays an aggregated overview of the survey data set. It is remarkable that the average distance of public transport and car/motorcycle increases in summer while the absolute number of students decreases. We suggest that this is because in summer (or fair weather) only those students who are not able to switch to walking or cycling due to too long distances use the bus or car. In winter (or bad weather) there are some students taking the bus/car for reasons of convenience – i.e. avoid walking in the rain – although the distance to school would be acceptable for cycling or walking. With regard to cycling the slight increase in average distance in summer is related to the strong increase in the number of students choosing to cycle. Some of these additional students who are cycling in the summer term show longer distances (using public transportation or car in winter).

We expect that the slight decrease in average distance for walking in winter is conditional on students who switch from cycling in summer (due to distance) to walking in winter due to weather conditions. For example, they avoid taking a risk going by bike in case of snowfall.

Table 3
Mean distances for each transport mode and absolute car availability

Variable	Mean	Standard deviation	*n* cases
Distance, km	3.941	3.566	4644
Distance – walking (summer)	0.710	0.408	845
Distance – walking (winter)	0.859	0.526	1010
Distance – cycling (summer)	2.364	1.307	1130
Distance – cycling (winter)	2.022	0.988	349
Distance – public transportation (summer)	5.390	3.111	2390
Distance – public transportation (winter)	4.943	3.034	2838
Distance – car/motorcycle (summer)	7.693	6.910	279
Distance – car/motorcycle (winter)	6.036	6.104	447
		Winter *n* cases	Summer *n* cases
Car availability – all time		391	502
Car availability – not all time		4253	4142
Season/weather		4644	4644

4.1. Model results and interpretation

The results of the estimation are shown in Table 4. There are 4650 college students within our data set. The sample size for estimation is 9300 because we regard each student as twofold: once for summer and once for winter.

Table 4 shows that on average 81% of all cases are correctly predicted by our model. A logistic analogy to R^2 in ordinary least squares (OLS) regression is the McFadden R^2. In general, the McFadden R^2 greater than 0.4 can be interpreted as a very good goodness of fit (Backhaus et al., 2003). With reference to these aspects, the model appears to have good explanatory qualities.

Table 4
Regression parameters

	Logistic regression		Correctly predicated (per cent)
	Coefficient	Wald	
Walking			88
Absolute term	10.774	2119.891	
Distance	−4.376	1369.57	
Winter season/bad weather	−0.591	13.024	
Car availability (all time)	−5.279	489.696	
Cycling			42
Absolute term	6.57	1196.853	
Distance	−0.904	748.843	
Winter season/bad weather	−2.081	200.51	
Car availability (all time)	−4.772	675.716	
Public transport			89
Absolute term	4.477	686.946	
Distance	−0.052	8.796	
Winter season/bad weather	−0.489	1510.892	
Car availability (all time)	−0.553	13.111	
Car/motorcycle			81
Average			81
Number of observations			9300
2 log likelihood			−9584
McFaddenR^2			0.63

All variables are significant at the 1% level. The Transport mode car/motorcycle is defined as reference category and parameters are set to zero. This means that all the other regression coefficients have to be interpreted in relationship to this category.

Compared to the mode car/motorcycle all other transport modes have a higher utility. If we ignore other influencing factors, walking is the most preferred mode (10.774). Taking into account the other variables, it is obvious that the utility of car/motorcycle will increase in relation to the three other modes. Although some information is provided by the coefficients themselves, the interpretation of the choice probabilities is more revealing. Fig. 9 shows the transport mode choice probabilities. Walking is the most important transport mode for short distances (up to 1 km) regardless of car availability and weather. Concerning cycling, weather and distance have a strong influence on associated probability. Students with car availability switch from bike to car at shorter distances than those with no car available who switch from bike to public transportation. To discriminate between the modes public transport and car/motorcycle the stated car availability is the most important factor. The gap between summer and winter in both motorized transport modes within the range of 1–3 km is related to the reduced probability of travel-to-school by bike in winter. Mostly, distance influences the decision to go by bike or walk on the one hand and to use public transportation or car/motorcycle on the other.

For several reasons it is recommendable to avoid a high proportion of students choosing transport modes other than walking or cycling. Obviously, there are higher costs related to transport modes like car/motorcycle than this is the case for walking and cycling. Moreover, walking and cycling are more activity related and thus better for students' health than motorized transport modes. A large percentage of students using public transport or car/motorcycle yields a negative impact on the environment due to emission (noise/pollution). In the following section these issues will be discussed in more depth.

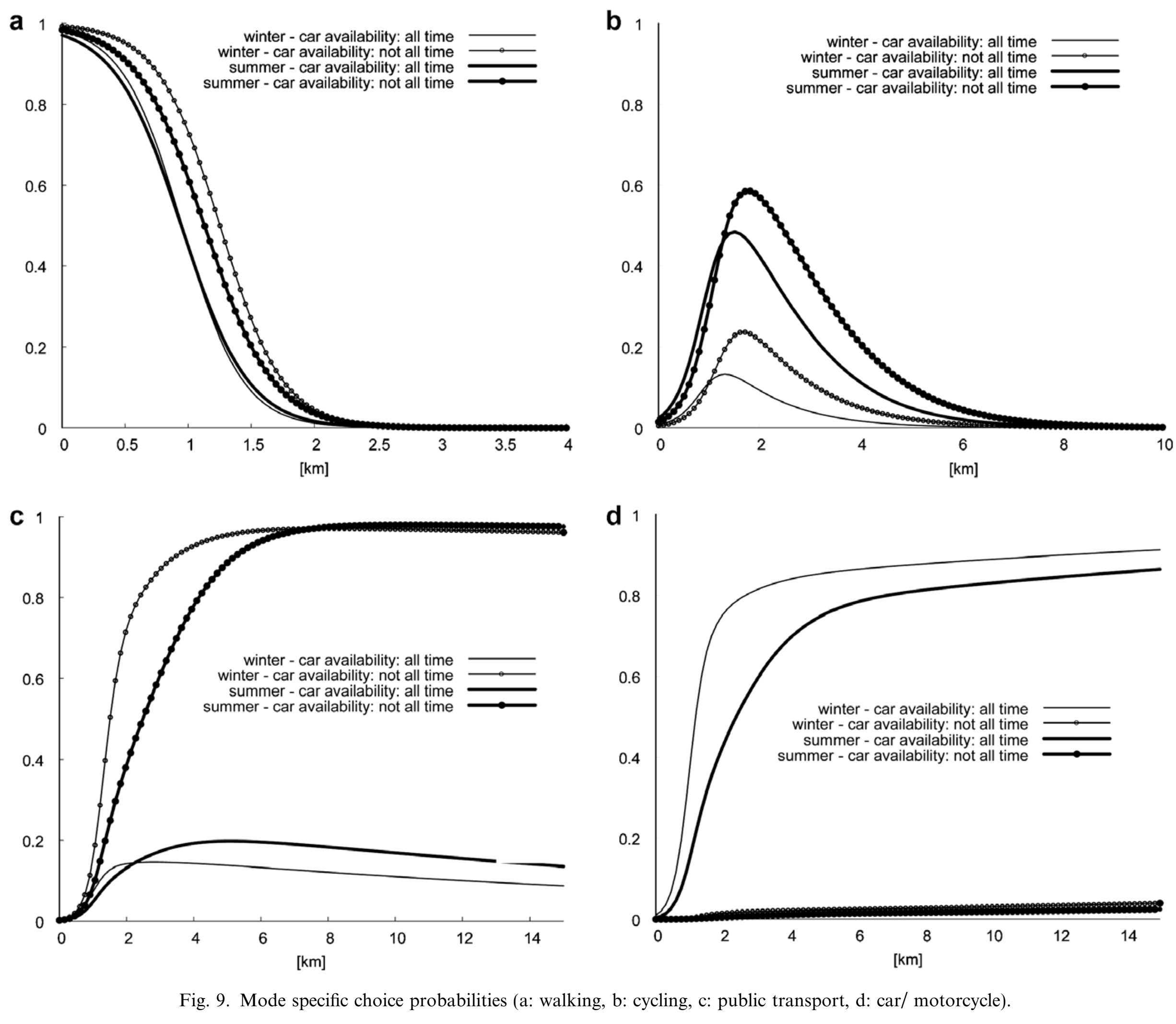


Fig. 9. Mode specific choice probabilities (a: walking, b: cycling, c: public transport, d: car/ motorcycle).

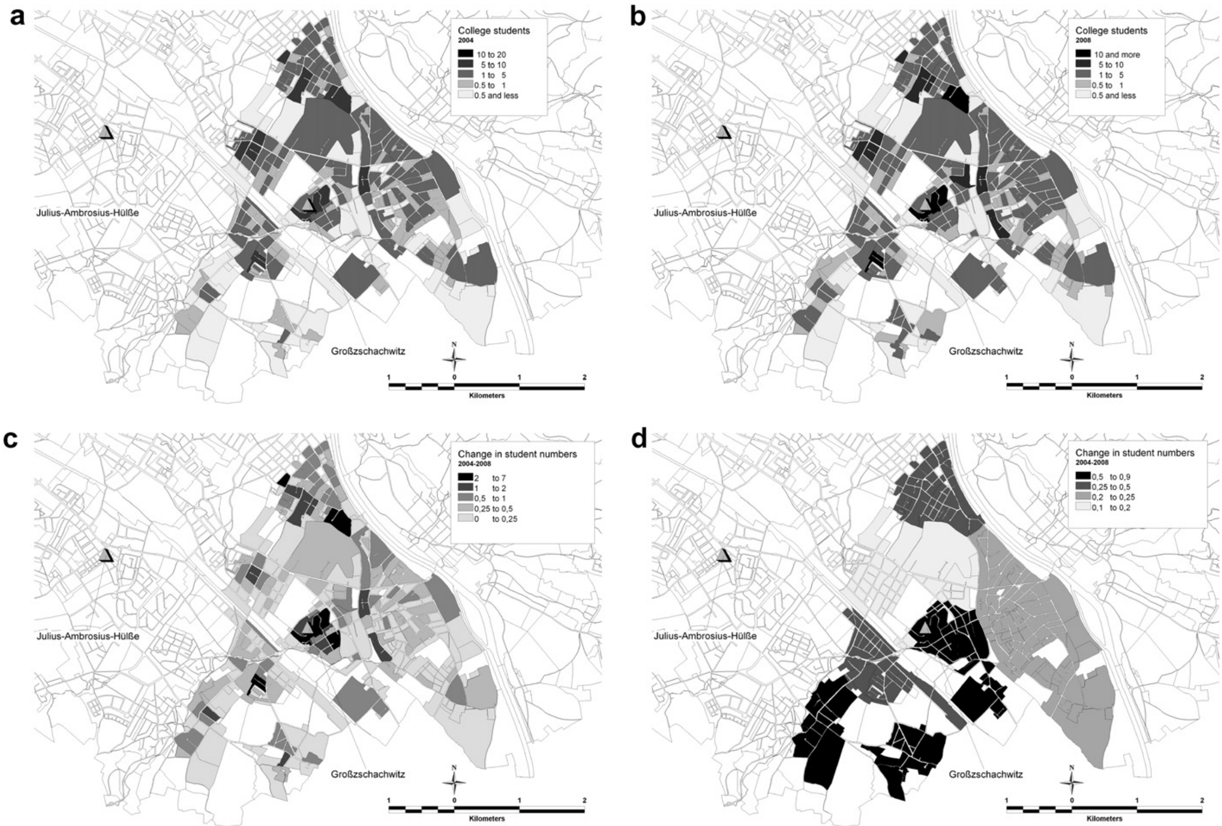


Fig. 10. Zone 1 of Großzschachwitz College: Absolute (c) and relative (d) change in student numbers from 2004 (a) to 2008 (b).

5. Modal-shift and school closure – an example

Under-utilization usually forces authorities to close schools. This is often justified for economic reasons. In this section, we like to analyze whether there is economic or social/ecological evidence emerging from a modal-shift, which could justify keeping open an under-used school.

5.1. Differences in mode choice due to school closure

In the year 2000, the school authorities in Saxony decided to close several school locations in Dresden due to declining enrollment in the 1990s. One of them is Großzschachwitz College which will be closed in summer 2008. According to this, the students affected have to attend different schools which are available. Here, we analyze the shift in transport mode choice and the related consequences. The example is based on the year 2004 and covers a student number forecast for 2008. The forecast shows that student numbers and hence enrollment will increase again (see Fig. 10). This phenomenon is typical for recently prospering cities in Eastern Germany. After years of dramatic decline, the population increases again.

According to Table 1, there are 68% of the students located in zone 1 attending Großzschachwitz College. In 2004, there are overall 467 students enrolled at Großzschachwitz College. Hence, 318 students of Großzschachwitz College are located in zone 1. The total of college students in zone 1 of Großzschachwitz College is 403 in 2004. Thus, 79% of all college students located in a block of zone 1 of Großzschachwitz College attend this college in 2004. Based on this, we assume that 79% of the students located in zone 1 enroll at the closest college available (see Section 3). We apply the MNL specified in Section 4 and yield the number of students[3] choosing a given transport mode for the years 2004 and 2008 (see Figs. 11 and 12). In both cases we just consider those 79% of the students located in zone 1 who attend the closest college, which is Großzschachwitz College in 2004 and Julius-Ambrosius-Hülße College in 2008. Three main patterns are evident:

1. Usually there is no possibility for most of the students to travel-to-school by car (see Fig. 8 and Table 3). So, in both scenarios there is only a small number of students commuting to school by car or motorcycle.

[3] We have computed the average utility due to summer and winter term.

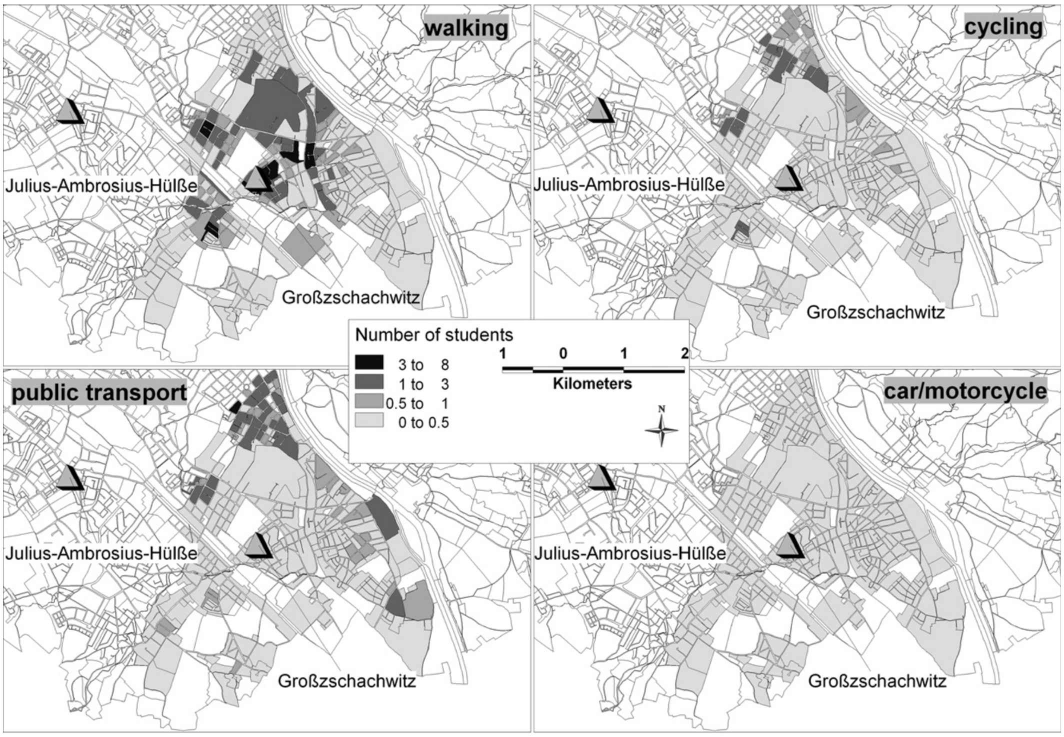


Fig. 11. Number of students walking, cycling, using public transportation or car/ motorcycle in 2004 (before the closure of Großzschachwitz College).

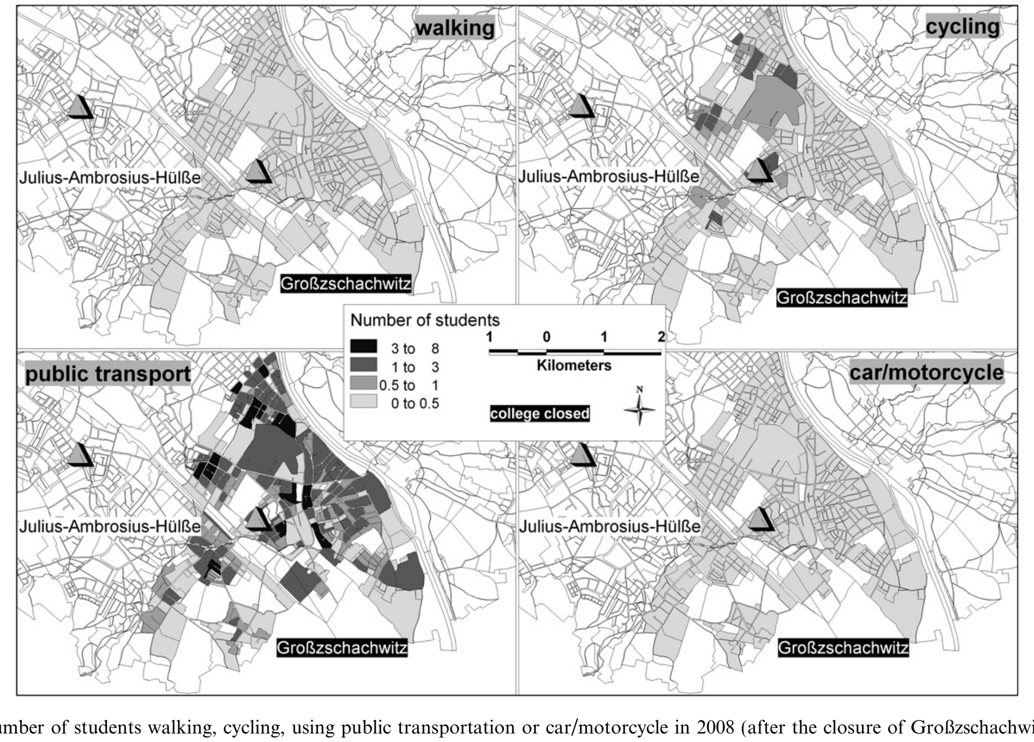


Fig. 12. Number of students walking, cycling, using public transportation or car/motorcycle in 2008 (after the closure of Großzschachwitz College).

2. Although no strong difference in the absolute number of students cycling can be observed, one can identify a difference in the spatial pattern: the spatial center of gravity of students who commute to school by bike shifts toward the location of Großzschachwitz College.
3. Most obviously there is a strong increase in the use of public transport while remarkably fewer students walk to school in 2008.

5.2. Quantification of modal-shift

Here, we try to quantify the consequences of the modal-shift due to a school closure. Since we focus on the transport sector we ignore costs related to the school location like maintenance and rent as well as external location costs like those of the loss of local neighborhood community (i.e. shops and services that depend on local schools are forced to close). We are aware of the difficulties associated with quantifying the modal-shift by costs since these costs are not always easy to determine – particularly external diseconomies. For convenience we do not discuss the different cost figures stated in the literature (see Infras/IWW, 2004; Planco Consulting GmbH, 1993 and Bickel and Friedrich, 1995 for example). Here, we use the cost figures

Table 5
Cost figures

Name	Costs in €	Unit	Source
Plain costs			
Cycling	.005	Student km	Assumed
Bus/tram (fare)	1	Student choosing public transport	Verkehrsverbund Oberelbe (2007)
Car/motorcycle	.165	Student km	FGSV (2002)
Value of travel-time			
Walking	.03	Student min	Baum et al. (1998)
Cycling	.035	Student min	Assumed
Public transport	.04	Student min	Axhausen et al. (2001)
Car/motorcycle	.065	Student min	Axhausen et al. (2001) Baum et al. (1998)
Accident			
Public transport	.28	Student km	
Car/motorcycle	1.64	Student km	
Noise			Baum et al. (1998)
Public transport	.00525	Student km	
Car/motorcycle	.00645	Student km	
Pollution			Baum et al. (1998)
Public transport	.00745	Student km	
Car/motorcycle	.01455	Student km	

Note that all figures are costs per trip.

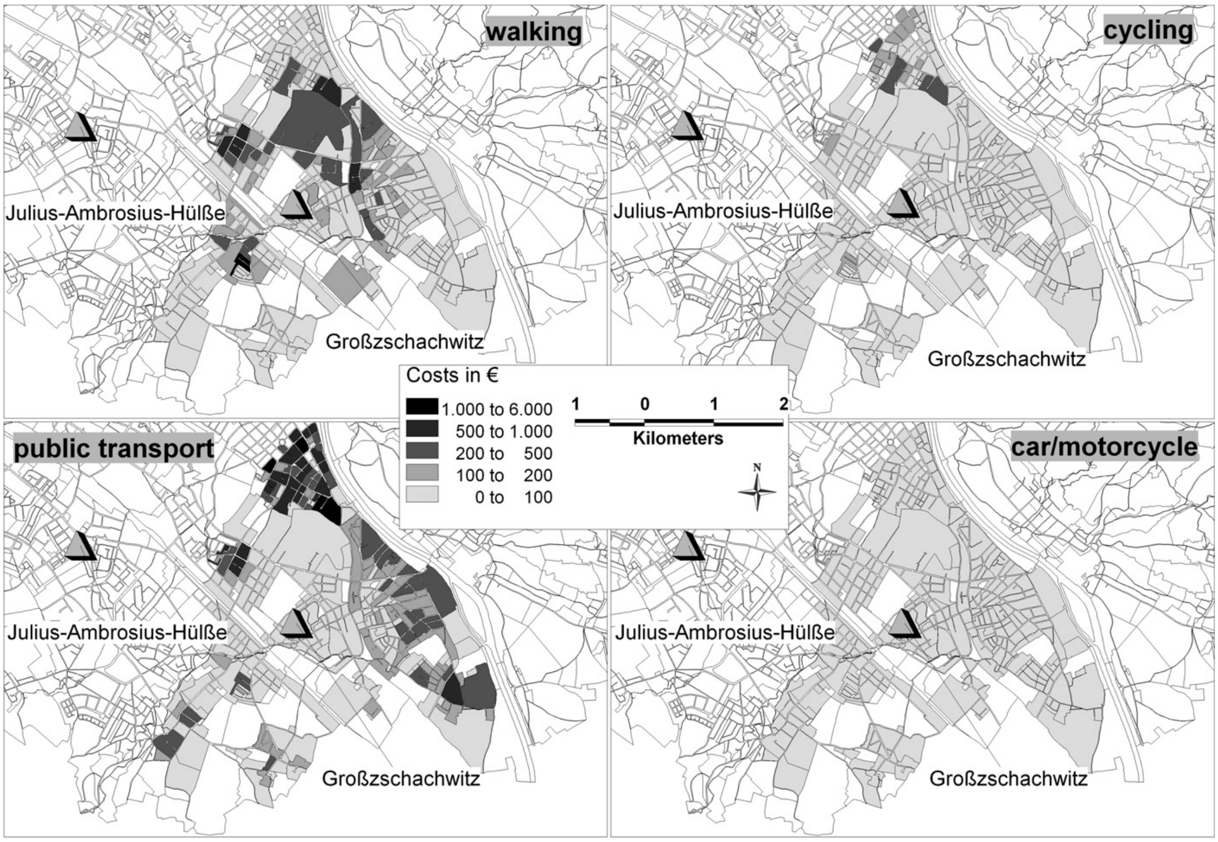


Fig. 13. Transport costs per mode in 2004.

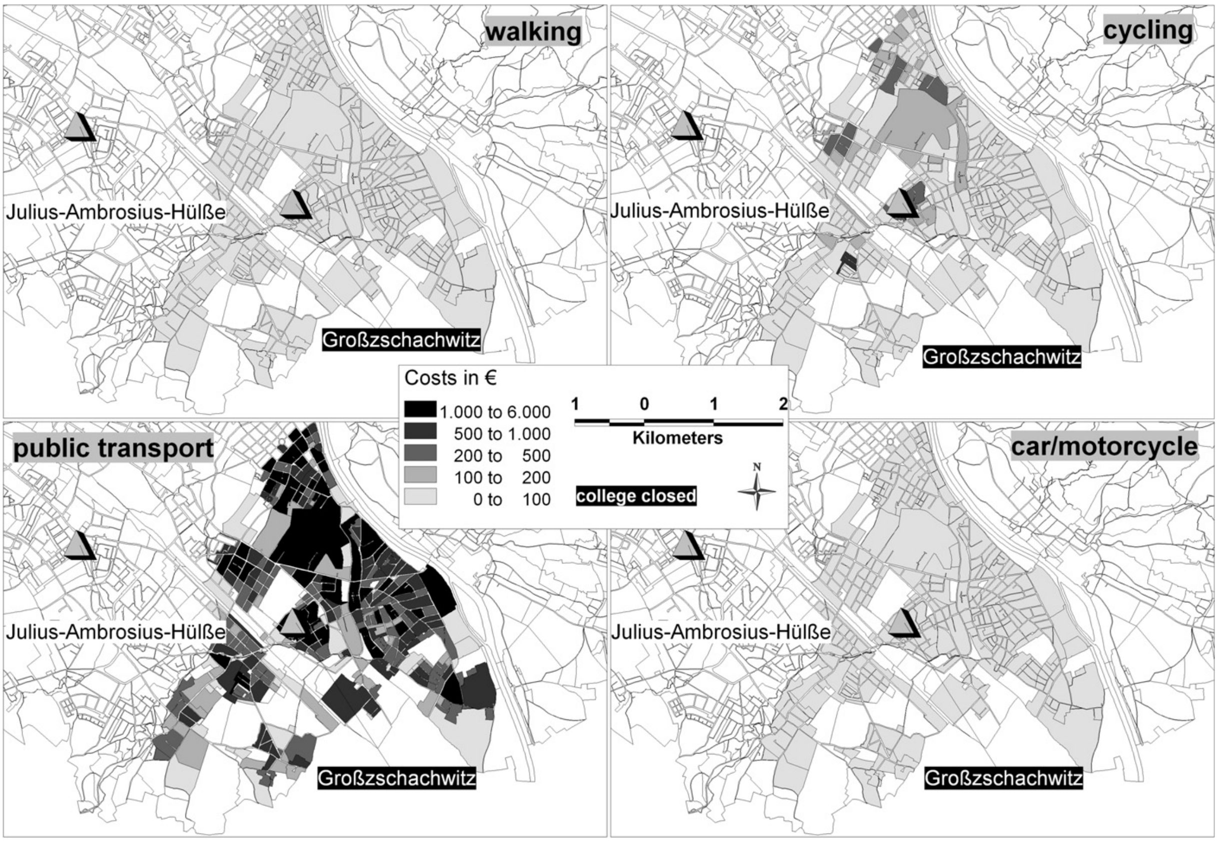


Fig. 14. Transport costs per mode in 2008.

given in Table 5. As plain costs we consider the average usage and consumption costs for cars and bicycles covering fuel usage, insurance as well as purchase and maintenance costs. The bus or tram fare reflects the costs of one trip using a standard seasonal ticket. We carefully assume that a student makes 2.2 trips to school per day.[4] There are usually 200 days of school per year in Saxony. The travel-times are derived from the distance matrix and the assumed average speeds (see Section 2). For public transport travel-times we use a travel-time matrix based on the bus and tram line network. We do not explicitly consider congestion costs because we assume these are included in the value of travel-time, pollution and noise costs. Furthermore, there arise costs due to decreased physical activity which is related to the transport modes public transport and car. An increase in the number of students commuting to school by car or bus yields increased levels of obesity, type 2 diabetes, heart disease etc. Unfortunately, we cannot obtain information about the relationship between student illness and students choosing motorized transport modes, nor do we have costs figures available based on diseases. Figs. 13 and 14 present the mode specific transport costs allocated to the location of the originator (student) for 2004 and 2008. In 2004 the walking costs are due to the value of travel-time of a lot of students walking to school with distances up to 1.5 km. Due to longer distances the number of students walking is very low in 2008 – and so are the walking costs. There is an increase in cycling costs observable, particularly within proximity of Großzschachwitz College. This is reasonable since there is a strong increase in student numbers in this area. Moreover, more students go by bike due to longer commuting distances. The increased number of students is a cogent reason for the increase in public transport costs as well. But most of all of this is because of the modal-shift due to longer commuting distances caused by the closure of Großzschachwitz College. There is a remarkable increase in students commuting by public transport, in particular within proximity to Großzschachwitz College. Because of the low level of car availability this transport mode and its costs are neglectable. In absolute numbers the transport costs rise from nearly 80,000€ in 2004 to more than 200,000€ in 2008 (increase by 150%). This increase is

[4] One trip to school in the morning and one trip back home at midday per school day. On some days there are additional trips necessary in the afternoon, for example sports.

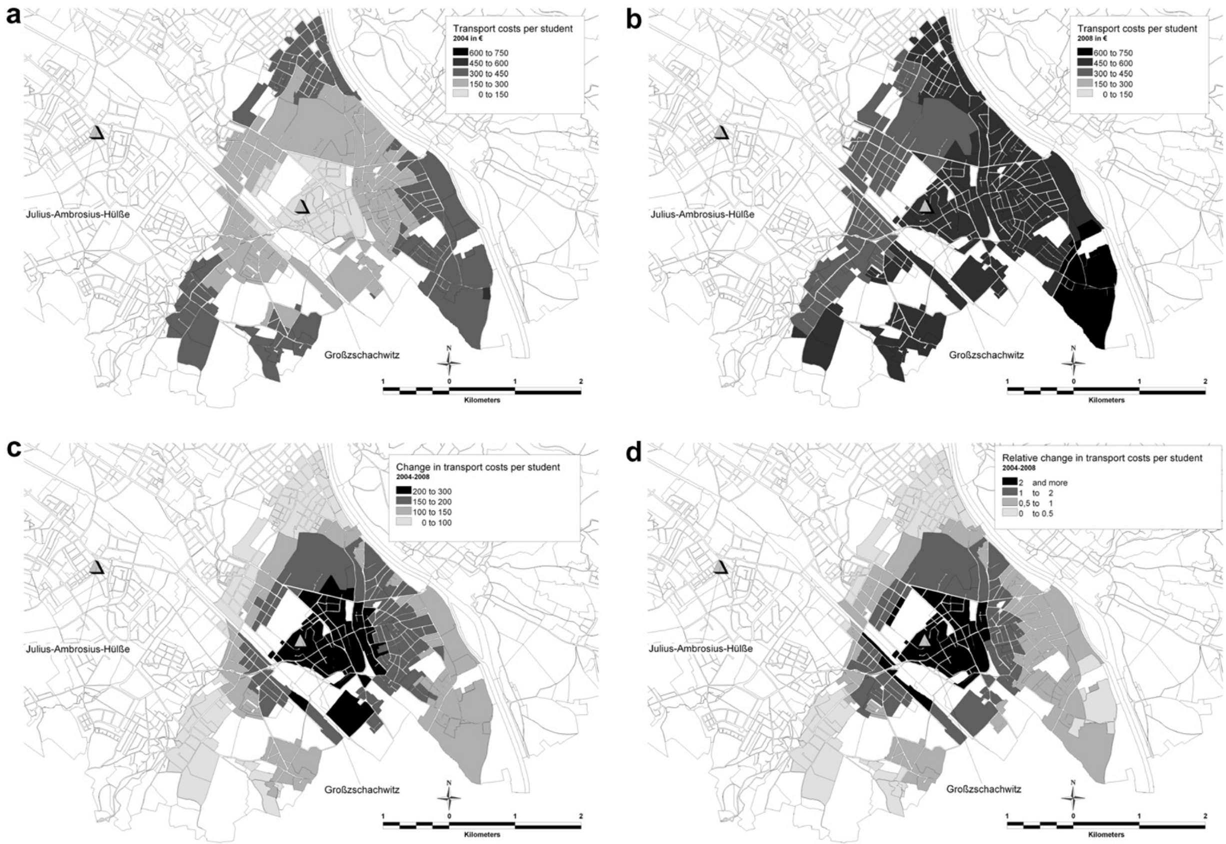


Fig. 15. Zone 1 of Großzschachwitz college: Absolute (c) and relative (d) change in transport costs from 2004 (a) to 2008 (b).

mainly due to public transport costs and focuses spatially on the proximate area (radius of 1 km) of Großzschachwitz College (see Fig. 15).

Assuming realistic location costs of a college of more than 1 million euros per year, the increase in transport costs does not justify the decision to keep an underused college running. This will probably hold true even if one considers additional external diseconomies (health, loss of community). From an economic point of view it is therefore not advisable to maintain a dense school network which is not appropriate for a smaller number of students. But if we consider other interests like ecological and social benefits, the example gives some evidence that local neighborhood schools are desirable.

6. Conclusion

We have presented a method utilizing GIS to disaggregate travel survey data. Particularly for travel-to-school analysis this could be a useful procedure to gain better and even more realistic modelling results. Mostly, students choose public transportation and thus detailed spatial information is available. In our analysis we have shown that besides the well-known factors like distance and authority responsible, the school's profile is affecting the school choice as well. The results of the multivariate analysis illustrate that weather or season, respectively, have a strong influence on transport mode choice for students' travel-to-school. Furthermore, we show that distance is the most important factor for discrimination between modes of transport linked with costs (public transport and car/motorcycle) and those with lower costs (walking and cycling). Our findings are consistent with the literature in the field. Moreover, the findings generate robust empirical evidence due to the extent of the sample.

By using an example we have made the attempt to quantify the costs of a modal-shift due to school closure. Although the increase in transport costs is remarkable this is not a substantial reason – from an economic point of view – against a school closure within an urbanized area. If we mostly consider other factors like the health of the students or ecological aspects, the costs of a modal-shift become apparent. Note, these findings are only valuable for an urban area. The closure of a school location in rural areas will have much more dramatic effects on travel-times and modal-shift as well as other socio-economic consequences.

Acknowledgements

We like to thank Dresdner Verkehrsbetriebe AG (DVB AG) for financial support of the survey. The very helpful comments of three anonymous referees are grateful acknowledged. Finally, we thank the Editor of this Journal, Prof. Richard Knowles, for copy editing the paper.

References

Axhausen, K., Haupt, T., Fell, B., Heidl, U., 2001. Searching for the rail bonus – results from a panel SP/RP study. European Journal of Transport and Infrastructure Research 1 (4), 353–369.

Bach, L., 1981. The problem of aggregation and distance for analyses of accessibility and access opportunity in location–allocation models. Environment and Planning A 13, 955–978.

Backhaus, K., Erichson, B., Plinke, W., Weiber, R., 2003. Multivariate Analysemethoden, 10th ed. Springer, Berlin.

Baum, H., Esser, K., Höhnscheid, K., 1998. Volkswirtschaftliche Kosten und Nutzen des Verkehrs. In: Forschungsarbeiten aus dem Strassen- und Verkehrswesen, No. 108.

Ben-Akiva, M., Lerman, S., 1985. Discrete Choice Analysis, Theory and Applications to Travel Demand. MIT Press, Cambridge, MA.

Bhat, C., 1997. A nested logit model with covariance heterogeneity. Transportation Research B 31, 11–21.

Bickel, P., Friedrich, R., 1995. Was kostet uns die Mobilität? Externe Kosten des Verkehrs. Heidelberg.

Black, C., Collins, A., Snell, M., 2004. Encouraging walking: the case of journey-to-school trips in compact urban areas. Urban Studies 38 (7), 1121–1141.

City Council of Dresden (=Landeshauptstadt Dresden), 2003. Dresden in Zahlen. Technical Report, Dresden.

de Boer, E., 2005. The dynamics of school location and school transportation: Illustrated with the Dutch town of Zwijndrecht. Transportation Research News (237), 11–16.

Ewing, R., Forinash, C., Schroeer, W., 2005. Neighborhood schools and sidewalk connections: what are the impacts on travel mode choice and vehicle emission?. Transportation Research News (237) 4–10.

Ewing, R., Schroeer, W., Greene, W., 2004. School location and student travel, analysis of factors affecting mode choice. Journal of the Transportation Research Board (TRB) (1895), 55–63.

Federal Environment Agency Germany, 2007. Nachhaltige Entwicklung. <http://www.umweltbundesamt.de/verkehr/verkehrstraeg/fussfahrad/texte/foerdnmiv.htm> (7 September 2007).

Federal Ministry of Transport, Building and Urban Affairs, 2002. Mobilität in Deutschland. <http://www.mid2002.de/> (15 June 2006).

FGSV (Ed.), 2002. Wirtschaftlichkeitsuntersuchungen an Straßen: Stand und Entwicklung der EWS. Forschungsgesellschaft für Strassen- und Verkehrswesen, Köln.

Fotheringham, A., Densham, P., Curtis, A., 1995. The zone definition problem in location allocation modelling. Geographical Analysis 27 (1), 60–77.

German Aerospace Center, 2007. Verkehrslage mit floating-car-daten (fcd). <http://www.taxifcd.de/Lage/HAMBURG/Barometer.html> (7 September 2007).

Gimona, A., Geddes, A., Elston, D., 2000. Localised areal disaggregation for linking agricultural census data to remotely sensed land cover data. In: Innovations in GIS, Computational Issues, vol. 7. Francis and Taylor, Philadelphia.

Goodchild, M., 1979. The aggregation problem in location–allocation. Geographical Analysis 11 (3), 240–255.

Greaves, S., 2004. GIS and the collection of travel survey data. In: Hensher, D. et al. (Eds.), Handbook of Transport Geography and Spatial Systems. Elsevier, Amsterdam, pp. 375–390.

Greene, W., 2003. Econometric Analysis, fifth ed. Pearson Education, New Jersey.

Hastings, J., Kane, T., Staiger, D., 2005. Parental preferences and school competition, evidence from a public school choice program. Technical Report, National Bureau of Economic Research.

Hoxby, C. (Ed.), 2003. The Economics of School Choice. University of Chicago Press.

Infras/IWW, 2004. External costs of transport – update study, on behalf of International Union of Railways (UIC). Infras/IWW, Zürich/Karlsruhe.

Koppelmann, F., Sethi, V., 2000. Closed-form discrete choice models. In: Hensher, D., Button, K. (Eds.), Handbook of Transport Modelling. Elsevier, Amsterdam, pp. 211–226.

Longley, P., Goodchild, M., Maguire, D., Rhind, D., 2001. Geographic Information Systems and Science. Wiley, Chichester.

Mahr-George, H., 1999. Determinanten der Schulwahl beim Übergang in die Sekundarstufe 1. In: Forschung Soziologie, vol. 28. Leske+Budrich, Opladen.

McDonald, M., 2005. Children's travel, patterns and influences. Ph.D. Thesis, University of California, Berkeley.

McFadden, D., 1973. Conditional logit analysis of quantitative choice behaviour. In: Zaremmbka, P. (Ed.), Frontier of Econometrics. Academic Press, New York.

McMillan, T., 2003. Walking and urban form: Modelling and testing parental decisions about children's travel. Ph.D. Thesis, University of California, Irvine.

Oosterhaven, J., 2005. Spatial interpolation and disaggregation of multipliers. Geographical Analysis 37 (1), 69–84.

Planco Consulting GmbH (Ed.), 1993. Gesamtwirtschaftliche Bewertung von Verkehrswegeinvestitionen: Bewertungsverfahren für den Bundesverkehrswegeplan 1992. Essen.

Schneider, T., 2004. Der Einfluss des Einkommens der Eltern auf die Schulwahl. Discussion Papers, vol. 446. DIW, Berlin.

Speiser, I., 1993. Determinanten der Schulwahl, Privatschulen – öffentliche Schulen. In: Europäische Hochschulschriften 513, Pädagogik, vol. 11. Lang, Frankfurt a.M.

Spiekermann, K., Wegener, M., 2000. Freedom from the tyranny of zones: towards new GIS-based spatial models. In: Fotheringham, A., Wegener, M. (Eds.), Spatial Models and GIS – New Potential and New Models. Taylor & Francis, London, pp. 45–62.

Van der Horst, D., 2002. The benefits of more spatial detail in regional bioenergy models and the problems of disaggregating agricultural census data. Options Méditerranéennes A 48, 131–138.

Verkehrsverbund Oberelbe, 2007. Fahrpreise auf einen Blick. <http://www.vvoonline.de/de/ticketsundnetz/fahrpreise/fahrpreiseaufeinenblick/index.aspx> (7 September 2007).

Woodside, A., Gunay, B., Seymour, J., 2002. Travel to school: a study of the effect of socioeconomic differences in secondary schools on the choice of 'travel to school' modes. In: Proceedings of the PTRC Conference. Cambridge.

geographical analysis

Geographical Analysis (2012) **44**, 65–86

Exposing Unobserved Spatial Similarity: Evidence from German School Choice Data

Sven Müller, Knut Haase, Frauke Seidel

Institute of Transport Economics, University of Hamburg, Hamburg, Germany

In a spatial context, flexible substitution patterns play an important role when modeling individual choice behavior. Issues of correlation may arise if two or more alternatives of a selected choice set share characteristics that cannot be observed by a modeler. Multivariate extreme value (MEV) models provide the possibility to relax the property of constant substitution imposed by the multinomial logit (MNL) model through its independence of irrelevant alternatives (IIA) property. Existing approaches in school network planning often do not account for substitution patterns, nor do they take free school choice into consideration. In this article, we briefly operationalize a closed-form discrete choice model (generalized nested logit [GNL] model) from utility maximization to account for spatial correlation. Moreover, we show that very simple and restrictive models are usually not adequate in a spatial choice context. In contrast, the GNL is still computationally convenient and obtains a very flexible structure of substitution patterns among choice alternatives. Roughly speaking, this flexibility is achieved by allocating alternatives that are located close to each other into nests. A given alternative may belong to several nests. Therefore, we specify a more general discrete choice model. Furthermore, the data and the model specification for the school choice problem are presented. The analysis of free school choice in the city of Dresden, Germany, confirms the influence of most of the exogenous variables reported in the literature. The estimation results generally indicate the applicability of MEV models in a spatial context and the importance of spatial correlation in school choice modeling. Therefore, we suggest the use of more flexible and complex models than standard logit models in particular.

Introduction

Space plays an important role in evaluating individual choices for several goods and services. School choice decisions especially exhibit features of choice situations that are highly influenced by spatial factors. In this article, we first give a brief introduction and a short overview of literature concerned with spatial choice modeling. Later, we turn our attention to the main aspects of the German school system and school choice modeling in particular.

Spatial Choice Modeling

A frequently used statistical model to analyze discrete choices is the multinomial logit (MNL) model. Its popularity is owed to, among other reasons, its utility-maximizing behavior and closed-form choice

Correspondence: Sven Müller—Institute of Transport Economics, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany
e-mail: sven.mueller@wiso.uni-hamburg.de

Submitted: August 17, 2009. Revised version accepted: January 20, 2011.

doi: 10.1111/j.1538-4632.2011.00836.x

probabilities. This model exhibits the property of independence of irrelevant alternatives (IIA), which (according to some researchers) is seen as a major shortcoming. This property may lead to model misspecification or false prediction of market shares. Haynes, Good, and Dignan (1988) argue that spatial choice problems especially show characteristics (e.g., random taste variation) that are difficult to handle with the MNL. Hunt, Boots, and Kanaroglou (2004) further point out that some researchers emphasize that spatial choice models have to be seen as distinct from discrete choice modeling due to incapabilities introduced by space. Meanwhile, developments in discrete choice analysis now allow existing models to account for a wide range of substitution patterns, including features of space (Bolduc, Fortin, and Fournier 1996; Train 1999; Walker and Li 2007). However, regarding the specifics of spatial choice (i.e., correlation in unobserved utility), little attention in the geographic literature is paid to the application of choice models other than the MNL. Hunt, Boots, and Kanaroglou (2004) state that discrete choice models should be increasingly applied in geographic contexts in order to evaluate their possible benefits. Attempts to account for spatial correlation involve the adjustment of the systematic component of utility or the implementation of a choice model that exhibits more flexible substitution patterns. The standard logit model enables constant substitution among alternatives. In contrast, the generalized nested logit (GNL) model allows for correlations in unobserved attributes by grouping alternatives that share unobserved (spatial) variability into common nests. Our focus here is on applying a GNL within the framework of random utility theory for school choice in the city of Dresden, Germany.

School Choice

Fluctuating student numbers over time and space force municipalities to adjust the number, the locations, and the capacities of schools. Within the framework of (long-term) school network planning, officials need to know factors that influence students to choose a certain school in order to derive expected utilization. The literature about school choice modeling usually focuses on racial mix, tuition fees, and travel-to-school distance (see section “School Choice Modeling”). Because free school choice is seldom found in many countries, most school location planning approaches do not account for spatial substitution (Müller 2008; Müller, Haase, and Kless 2009), but some references lead one to believe that spatial substitution patterns between school locations exist (Manski and Wise 1983; Borgers et al. 1999; Müller 2009).

The concept of utility entails a compensatory decision process. It presumes that students’ choices involve trade-offs among the attributes characterizing schools. For example, a student may choose a school located far away from her location if the profile offered by that school (e.g., math and languages) compensates for the increased travel distance. Based on such trade-offs, each student selects the school with the highest utility value. The focus on utility maximization in this article arises from its strong theoretical background.

The utility-maximization rule is robust; that is, it provides a good description of choice behavior even if students use different rules (Koppelman and Bhat 2006, pp. 12–13). German students are free to choose a secondary school in which to enroll. This means enrollment is not determined by location of the students, because school districts do not exist. In general, after 4 years in primary school, a student enrolls in a secondary school. Based on their academic ability, students are allowed to enroll in either *Mittelschule* or *Gymnasium*. The latter can be seen as a special type of secondary school that prepares graduates to attend the university. The degree that students have when graduating from Gymnasium, therefore, is equivalent to a high school degree in the United States. Figure 1 shows the structure of the German school system as well as the number of grades and the corresponding students’ ages.

Students rarely switch from one school to another on the same educational level (i.e., switch from one Gymnasium to another). The secondary school choice decision is strongly dependent on an educational recommendation a student receives after having finished 4 years of primary school. As a result, students showing good scholastic performance are allowed to enroll in Gymnasium, while less capable students have to attend Mittelschule. In general, enrolling in Gymnasium is prohibited if the educational recommendation is not for Gymnasium. Students rarely choose to enroll in Mittelschule when their educational recommendation qualifies them for Gymnasium.

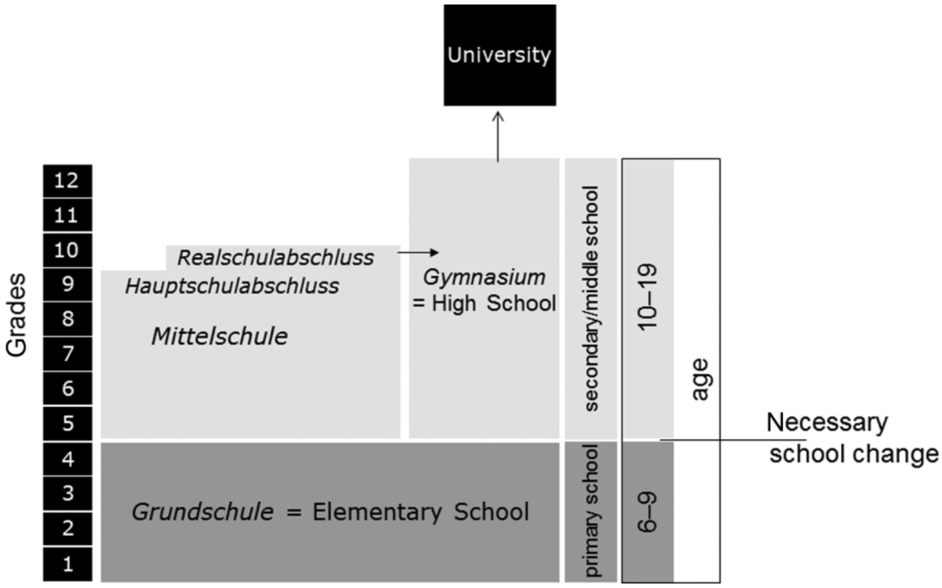


Figure 1. Main features of the educational system of Germany.

Unlike Gymnasium, Mittelschule schools generally are more homogeneous regarding their spatial distribution and offered profiles. Hence, in this article, we focus on students choosing a Gymnasium school in the city of Dresden, Germany. Gymnasium schools exhibit varying characteristics regarding the amount of education offered in subjects like sciences, languages, and music/arts. The objective here was to describe the development of a school choice model embedded in the framework of discrete choice analysis, considering spatial dependencies between the school locations under study. Although this is a specific application, the modeling framework for spatial choices presented can be easily applied to a wide range of spatial contexts, like demand modeling for recreational sites and other (non market) recreational goods and services. Valuable applications regarding the subject of spatial choice include recreational demand models (Train 1999) or housing location choice models (Guo and Bhat 2007).

Multivariate Extreme Value Models

The choice models we employ in this article are based on the assumptions of random utility theory. A decision maker n is assumed to choose from a set of available alternatives C_n alternative i such that utility $U_{ni} \geq U_{nj}\ \forall\ j \in C_n, j \neq i$. Note that $C_n \subseteq C$ with C: set of all alternatives under study. Because we do not observe all effects on utility-maximizing behavior, we decompose utility U_{ni} into a deterministic (or systematic) part V_{ni} and a stochastic part ϵ_{ni}:

$$U_{ni} = V_{ni} + \epsilon_{ni}. \tag{1}$$

Usually V_{ni} is linear in parameters:

$$V_{ni} = \sum_h \beta_{ih} x_{nih}. \tag{2}$$

The H independent variables x_{nih} describe alternative i and characteristics of decision maker n. The x_{nih} variables are weighted by coefficients β_{ih}. Because ϵ_{ni} is a random variable, we can only determine the probability that an individual n chooses i from her choice set of available alternatives C_n by

$$P_n(i \mid C_n) = \mathrm{Prob}(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj}, \quad \forall \quad j \in C_n, j \neq i) \tag{3}$$

$$= \mathrm{Prob}(\epsilon_{nj} < V_{ni} - V_{nj} + \epsilon_{ni}, \quad \forall \quad j \in C_n, j \neq i). \tag{4}$$

Now we have to make assumptions about the joint probability distributions for the random components of utility ϵ_{ni} in equation (1).

Multivariate extreme value (MEV) models constitute a large class of discrete choice models whose unifying attribute is that the stochastic part of utility ϵ_{ni} is distributed as a generalized extreme value for all alternatives (Train 2003, p. 80). Following McFadden (1978), different kinds of discrete choice models can be developed as special cases of the more general MEV model formulation. The generating function for different types of models (e.g., MNL and nested logit [NL]) is obtained by making specific assumptions about the cumulative distribution of the vector of unobserved utility $\epsilon_n = \langle \epsilon_{n1}, \ldots, \epsilon_{nJ} \rangle$ (Train 2003, pp. 83–100). Each instance of the MEV family is derived from a continuous and differentiable generating function,

$$G: \mathbb{R}_+^{C_n} \to \mathbb{R}_+, \tag{5}$$

which defines the cumulative distribution function (CDF) of the error terms and the choice model, respectively. The CDF of an MEV model takes the form

$$F_{\varepsilon_n}(\xi_1, \ldots, \xi_{C_n}) = e^{-G(e^{\xi_1, \ldots, \xi_{C_n}})}, \tag{6}$$

whereas, in order for F to be a CDF, the μ-MEV-generating function G needs to exhibit the following properties:

(1) $G(y)$ is a nonnegative function, $G(y) \geq 0 \ \ \forall y_i \in \mathbb{R}_+^{C_n}$;
(2) $G(y)$ is homogeneous of degree $\mu > 0$; that is, $G(\lambda y) = \lambda\mu G(y)$, for $\lambda > 0$;
(3) $G(y)$ asymptotically tends to infinity for each y_i tending to infinity: $\lim_{y_i \to \infty} G(y_1, \ldots, y_i, \ldots, y_{C_n}) = \infty$, for each $i = 1, \ldots, C_n$; and,
(4) the mth partial derivative of $G(y)$ with respect to m distinct y_i is nonnegative if m is odd and non positive if m is even, for any distinct indices $i_1, \ldots, i_m \in \{1, \ldots, C_n\}$.

The probability of choosing alternative i from a choice set C_n for an MEV model may be written as

$$P_n(i \mid C_n) = \frac{y_i G_i(y_1, \ldots, y_{C_n})}{\mu G(y_1, \ldots, y_{C_n})}, \tag{7}$$

where $y_i = e^{V_{ni}}$. By explicitly assuming that the generating function $G(y_1, y_2, \ldots, y_{Cn})$ takes the form

$$G(y) = \sum_{i=1}^{C_n} y_i^{\mu}, \tag{8}$$

the MNL model is derived. Substituting equation (8) into (7) yields

$$P_n(i \mid C_n) = \frac{e^{\mu V_{ni}}}{\sum_{j \in C_n} e^{\mu V_{nj}}}, \tag{9}$$

where μ is a scale parameter that is not identified and has to be set to an arbitrary value (e.g., one) for model identification purposes. Equation 9 is the logit choice probability. In the case of the MNL, the random components of utility ϵ_{ni} in equation (1) are assumed to be independently and identically distributed extreme value (iid EV), which is a special case of the assumption made for the error terms in MEV models. Note that equation (9) has a closed form and that the unknown coefficients β_{ik} of equation (2) can be provided relatively simply through maximum likelihood estimation.

Failure of IIA and Substitution Patterns

Although the MNL is applied in various situations, it has some severe shortcomings, particularly in a spatial choice context. The main issue concerning spatial choice (such as school choice) is the well-known IIA property (Luce 1959), a direct outcome of the assumption that the ϵ_{ni} are iid (Haynes, Good, and Dignan 1988). The IIA property ensures that the ratio of choice probabilities for any two alternatives is unaffected by the presence or change of any other alternative and its attributes. Therefore, a change in the probability of one alternative leads to identical changes in relative choice probabilities for all other alternatives. For example, let us assume, that a school network consists of five school locations available to a student and that the predicted choice probabilities from equation (9) equal 0.30, 0.12, 0.15, 0.18, and 0.25, respectively. Next, we assume that school location 5 is closed due to an expected overall decline in student numbers. Equation (9) predicts choice probabilities equal to 0.40, 0.16, 0.20, and 0.24 for the remaining four locations. The choice probability for every remaining alternative increases by one-third (i.e., a 33.33% relative change to choice probabilities). This rigid substitution pattern ignores the fact that some schools may be better substitutes for the closed site (e.g., because of spatial proximity to that school). Although whether IIA holds for given data is an empirical question and a matter of the specification of V_{ni} set, many geographers suggest that IIA is unlikely to hold in spatial choice applications. For example, Haynes and Fotheringham (1990) note that size, aggregation, dimensionality, spatial continuity, and variation and location characteristics of spatial choice data are likely to produce substitution patterns that violate IIA. In its strict form, IIA applies only to an individual student n and not to all students as a population. As Ben-Akiva and Lerman (1985, pp. 109–11) state, IIA often is misinterpreted as implying that the ratio of the shares of the population choosing any two alternatives is unaffected by the utilities of other alternatives (schools).

Many attempts in the past tried to overcome IIA weaknesses and to account for a richer pattern of substitution than that offered by the MNL (Hunt, Boots, and Kanaroglou [2004] for a more detailed overview). Unfortunately, most of these attempts were based on the logit model specified by McFadden (1975) (Timmermans and van der Waerden 1992). In general, the models used have not been consistent with random utility theory (Koppelman and Sethi 2000). As Hunt, Boots, and Kanaroglou (2004) point out, developments in discrete choice modeling are considerable, and today various models exist that are able to cope with spatial complexity. These models are classified as *closed-form* models, such as the MNL, and *open-form* models, such as the multinomial probit (MNP). The advantage of the closed-form models is their computational tractability, whereas the advantage of the open-form models is their flexibility. In the remainder of this article, we consider a closed-form model with a maximum of flexibility for considering (spatial) substitution patterns of choice alternatives (i.e., schools).

Two Closed-Form Discrete Choice Models with Flexible Substitution Patterns

The NL is a model that accounts for a wide range of substitution patterns that arise when alternatives share unobserved attributes. Its implementation is appropriate when alternatives faced by a decision maker can be grouped into subsets, or *nests*, in such a way that IIA holds between alternatives within each nest but not across nests. Due to the nesting of alternatives, the NL overcomes the proportional substitution across alternatives imposed by the MNL through IIA. Following Train (2003 p. 84), the NL is a more general formulation of the MNL that allows for correlation in unobserved utility. The generating function to derive the NL from equation (7) is

$$G(y) = \sum_{k=1}^{K} \left(\sum_{i=1}^{C_n} y_i^{\mu_k} \right)^{\frac{\mu}{\mu_k}}, \tag{10}$$

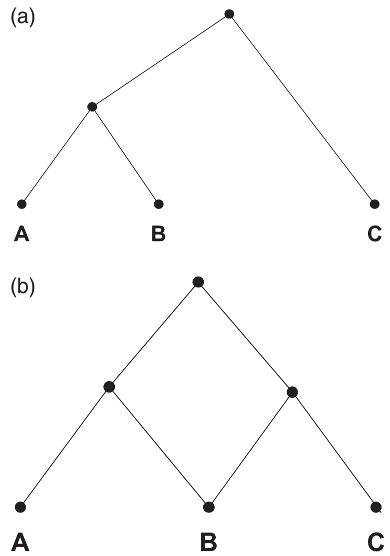


Figure 2. (a) Nesting structure of the nested logit. (b) Nesting structure of the generalized nested logit.

where K depicts the number of existing nests B_k. A separate scale parameter μ_k exists for each nest, so that only the ratios μ/μ_k are identified. Thus, a normalization of the scale parameter is required for model identification purposes.

Normalizing $\mu = 1$ is good practice, which is referred to as normalization from the top, although other normalization for the NL can be considered as well (Bierlaire 2006). Furthermore, $\mu/\mu_k = \sqrt{1-\rho_{ij}}$, where ρ_{ij} denotes the correlation coefficient $corr(U_i, U_j)$. This is the correlation of the total utilities for any pair of alternatives in C_n that share the same nest (Ben-Akiva and Lerman 1985; Heiss 2002, p. 289). The scale parameter generally serves as an indicator for the independence among alternatives within a nest. Thus, a higher μ_k translates into a higher correlation between alternatives in that particular nest. Substituting equation (10) into (7) yields the following NL choice probability that individual n chooses alternative i:

$$P_n(i/C_n) = \frac{e^{\mu_k V_{ni}}}{\sum_{j \in C_{nk}} e^{\mu_k V_{nj}}} \frac{\left(\sum_{\in C_{nk}} e^{\mu_k V_{nj}}\right)^{\mu/\mu_k}}{\sum_{m=1}^{K}\left(\sum_{j \in C_{nm}} e^{\mu_m V_{nj}}\right)^{\mu/\mu_m}}, \tag{11}$$

where $C_{nk} = B_k \cap C_n$, and k denotes the nest that contains alternative i. Figure 2a shows the nesting structure for an NL model with three alternatives, A, B, and C, available; that is, $C_n = \{A, B, C\}$. Alternatives that have similar unobserved attributes (here, alternatives A and B) are assigned to one nest.

For the NL model, every alternative belongs to only one nest. This aspect imposes an important restriction on the model insofar as this assumption might be inappropriate in some situations. Assume, for example, that alternative B shares some unobserved attributes not only with alternative A but also with alternative C. Such a nesting structure is presented in Fig. 2b and belongs to the GNL[1] model.

The proposed analytical formulation is derived from the MEV model in equation (7). An alternative may be a member of more than one nest to varying degrees. An allocation parameter α_{ik} reflects the extent to which alternative i is a member of nest k. The parameter α_{ik} is nonnegative, and $\sum_k \alpha_{ik} = 1 \quad \forall i$ for identification purposes. Further, α_{ik} may be interpreted as the portion of alternative i that is allocated to each nest k. If $\alpha_{ik} = 0$, alternative i does not belong to nest k, and if $\alpha_{ik} = 1$, the alternative belongs to nest k only. Values of α_{ik} between zero and one indicate a membership of an alternative i to multiple nests. A larger value of α_{ik} means that alternative i shares a larger amount of common unobserved attributes with alternatives in nest k than with alternatives in other nests. The generating function to derive the choice probability for the GNL is

$$G(y) = \sum_{k=1}^{K} \left(\sum_{i=1}^{C_n} \alpha_{ik}^{\frac{\mu_k}{\mu}} y_i^{\mu_k} \right)^{\frac{\mu}{\mu_k}}. \tag{12}$$

Substituting equation (12) into (7) yields the probability function of the GNL

$$P_n(i \mid C_n) = \sum_{k=1}^{K} \frac{\left(\sum_{j \in C_{nk}} \alpha_{jk}^{\mu_k/\mu} e^{\mu_k V_{nj}} \right)^{\mu/\mu_k}}{\sum_{m=1}^{K} \left(\sum_{j \in C_{nm}} \alpha_{jm}^{\mu_m/\mu} e^{\mu_m V_{nj}} \right)^{\mu/\mu_m}} \frac{\alpha_{ik}^{\mu_k/\mu} e^{\mu_k V_{ni}}}{\sum_{j \in C_{nk}} \alpha_{jk}^{\mu_k/\mu} e^{\mu_k V_{nj}}}. \tag{13}$$

Due to the nest structure and the flexible allocation of alternatives to nests, the GNL does not exhibit the IIA of the MNL. Nevertheless, this advantage comes at the expense of an a priori assumption about the underlying correlation structure. If each alternative enters only one nest, with $\alpha_{ik} = 1 \ \forall i \in B_k$ and zero otherwise, the model becomes the NL of equation (11). If, in addition, $\mu_k = 1 \ \forall k$, the model becomes the MNL as in (9) (Train 2003, p. 95).

School Choice Modeling

The next section summarizes studies concerned with the modeling of school choice decisions and their influencing factors. In the past, several types of choice models have been employed. All studies identify distance to school as an important factor in individuals' choice decisions. Based on the findings in the literature and some data-related issues, we specifiy a spatial choice model for school choice in section "Data-related issues and model specification."

Literature Review

Manski and Wise (1983) initiated the growing now body of literature about the choice of educational facilities such as schools and universities. Borgers et al. (1999) employ an MNL based on stated choice data to identify the choice between Protestant, Catholic, and public schools in the Netherlands. They find evidence that school type (e.g., Montessori), religious affiliation, school size, and the distance between a student's location and a school are the most important decision-making factors. Moreover, they include substitution and availability effects to account for (spatial) competition between schools. Lankford, Lee, and Wyckoff (1995) model the choice across public, religious, and independent schools. Their MNP analysis reveals that school choice is affected by the racial composition of public schools, the crime rate, and the religious orientation of a school, as well as by the socioeconomic characteristics of a household, particularly the location of a household in a central city. Lankford and Wyckoff (2006) use a sequentially estimated NL to identify the effect of school choice on the racial segregation of students. They find that the racial composition of a school and the distance between a student's home and school influence school choice. They also find similarities between Catholic and private schools in unobserved factors. The mixed MNL (MMNL) model of Hastings, Kane, and Staiger (2006) furnishes evidence that distance traveled to school is the most important factor influencing the school decision. The combination of schools' mean test scores, household incomes, and parents' academic abilities results in a negative correlation between distance to school and mean test score. For German schools, Schneider (2004) shows that besides distance to school, household income has a strong influence on school choice. Finally, Jepsen and Montgomery (2009) use an NL to show that distance is the most important factor in deciding whether to enroll at a community college and about which school to choose. This finding is uncovered after controlling for tuition fees, school size, and socioeconomic variables.

This short review shows that distance seems to be by far the most important factor in the school choice process, indicating the possibility of spatial substitution between proximate schools. In our analysis, we apply the GNL model, which in contrast to MNP and MMNL is computationally easy to handle in identifying such substitution patterns. We control for most of the variables used in the studies mentioned here.[2]

Data-Related Issues and Model Specification

We aim to model the school choice of students living in the city of Dresden, Germany, and we use the survey by Müller, Tscharaktschiew, and Haase (2008). This study has been designed to model the travel-to-school mode choice. The data were collected at the schools under study, representing the endogenous variable in our study. The sample was stratified to l subsets of students with $l = 1, \ldots, L$: $L \leq C$ contains all individuals who have chosen one particular alternative. Hence, this sample is choice based, which leads to problems in estimating GNL with standard maximum likelihood methods. Fortunately, we acquired data on the actual market share of each school. Therefore, we are able to employ the weighted exogenous sampling maximum likelihood (WESML) estimator[3]

$$\max_{\theta} \sum_{l=1}^{L} \sum_{n=1}^{N_l} \sum_{i \in C_n} y_{ni} \left(\frac{W_l}{H_l} \right) \ln\left[P_n(i \mid X_{ni}, \theta) \right], \tag{14}$$

where N_l denotes the set of students having chosen school l, y_{ni} is the choice by student n concerning school i (i.e., equals one, if student n chooses school i, zero otherwise), W_l denotes the known actual market shares, and H_l represent sample market shares, X_{ni} is the vector of exogenous variables, and θ is the vector of unknown coefficients β_{ih} in equation (2). The fraction W_l/H_l is reported in the last column of Table 1. As stated by Ben-Akiva and Lerman (1985, pp. 238–9), this estimator yields a consistent estimate for θ. However, the WESML estimates are not necessarily asymptotically efficient.

From the survey sample, we select all students enrolled at Gymnasium ($N = 5{,}215$). Information about a chosen school, address (Fig. 3), and sex (about 44% of all students are male) is directly available from the survey. Table 1 reports the average travel distance to each school (based on street network) and its corresponding standard deviation. Moreover, from local authority statistics, we add the average income of the city district where a student is located. Average income is intended to account for differences among students' neighborhoods (Cullen, Jacob, and Levitt 2005).

Although we know that the median would be more appropriate, median income data are not available. Explanations in the remainder of this section are based on the following assumptions:

(1) The likelihood of attending a private school is generally higher for a student originating from a wealthier city district than for a student living in a poorer district; and
(2) As average income of a city district increases, the affinity of inhabitants toward education tends to increase.

City districts having a low average income are assumed to exhibit a large number of blue-collar workers, directly translating into a poorer social standing for the respective districts (Neu 2007). We also assume that the majority of households in a wealthier city district can afford tuition. Consequently, children of these households are more likely to enroll at private schools. If a student lives in a wealthy district, the student either stems from a wealthy household that enables him or her to enroll at private school or, if not, at least some in his or her peer group belong to a wealthy household. Hence, peer group pressure may influence the school choice decision of students from less wealthy families living in a well-off district. Finally, we have some attributes of the schools themselves, mainly profiles and size (Table 1 and Fig. 3). In the school choice context, one can imagine that one school is more similar to a second one than to other schools due to the same profile offered, authority, and spatial proximity.

While some of these similarities could be incorporated in V_{ni}, spatial similarity is particularly difficult to operationalize in V_{ni}. As Hunt, Boots, and Kanaroglou (2004) point out, spatial effects may be accounted

Table 1 School Attributes, Availability, Frequencies, and Weights

Short name	Profile a	b	c	d	e	f	g	Size*	Distance to CBD (*km*)	Travel distance (*m*) Mean	SD	Avail. 2004	Number of observations	WESML weight
STBE						1		3	0.8322	5,508.94	2,659.39	5,215	400	0.7806
MC	1							4	0.3966	5,560.22	2,641.01	5,215	199	1.5868
ANNE	1							3	0.8555	5,997.95	2,876.50	1,200		
BB				1				4	1.8995	5,455.92	2,811.18	5,215	566	0.5837
DKS					1			3	1.9790	6,140.89	2,564.16	5,215	120	1.6644
WALD							1	2	2.7259	6,518.75	2,541.77	5,215	56	2.3189
RORO			1					3	1.4277	5,837.08	2,576.96	5,215	373	0.6265
PEST	1							3	3.5188	7,429.34	2,776.76	5,215		
KLOT				1				3	7.5818	10,383.76	3,634.24	5,215	340	0.8869
DKS2					1			3	7.8815	9,176.73	3,645.66	5,215	32	2.8000
MAN		1						3	4.1813	6,124.14	3,333.74	5,215	261	0.9376
EVKZ						1		3	3.8280	5,904.65	3,268.67	5,215	364	0.8762
JOHA					1			4	2.4011	5,427.87	2,919.61	2,088	218	0.4689
HE				1				5	3.6305	5,811.50	3,343.25	5,215		
GZW	1							4	8.1933	8,874.82	4,597.43	3,538	168	1.0972
JAH				1				4	5.2381	6,732.89	3,945.26	5,215	619	0.5248
WUST				1				3	6.8566	7,451.84	4,222.19	2,055		
FL				1				3	1.9912	6,146.49	3,200.91	3,524	169	1.0697
VITZ					1			4	2.7463	6,023.21	3,276.62	5,215	266	0.9912
PLAU				1				5	3.7615	7,085.79	3,659.79	5,215	439	0.9171
COTT					1			5	4.2186	7,767.03	3,726.21	5,215	625	0.6322
JAS	1							5	4.2871	7,986.27	3,931.19	2,055		

*Measured in number of classes per grade.

a, Math profile; b, Math core; c, Languages core; d, Math and languages; e, Math and music/arts; f, Math and languages and music/arts (private school); g, no profile (private school); CBD, central business district; WESML, weighted exogenous sampling maximum likelihood; SD, standard deviation.

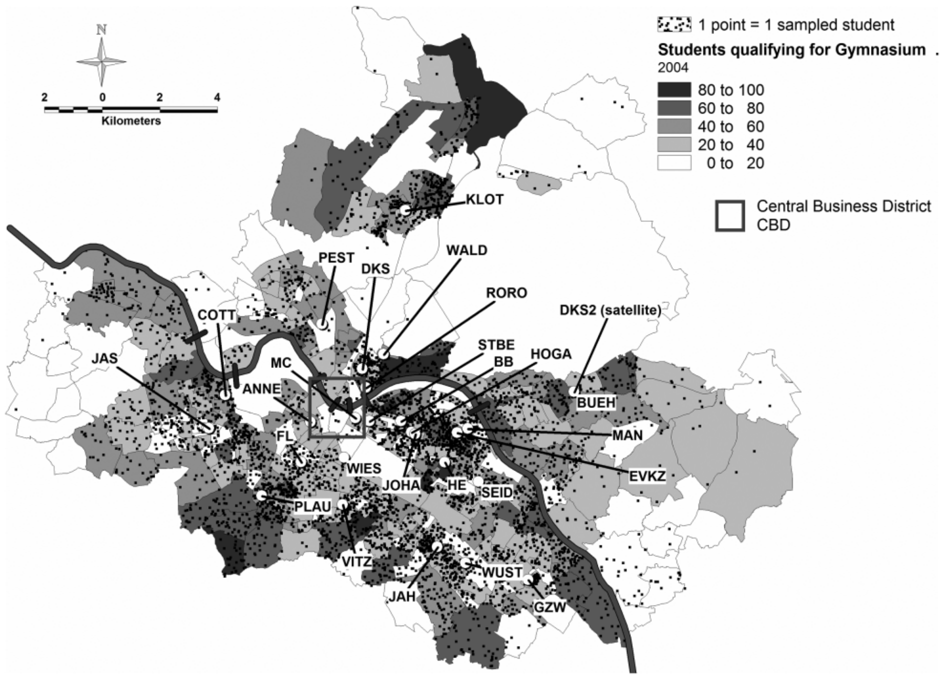


Figure 3. Locations of Gymnasien sampled students, and student numbers in Dresden, 2004.

for by adjusting the systematic utility of an alternative but at the expense of possibly affecting the behavioral underpinnings of the choice model. To operationalize spatial similarity in the deterministic component of utility, one would have to define explanatory variables that describe every kind of relation and spatial dependency that might exist between any pair of school locations. Such a model specification suffers from a remarkable increase in degrees of freedom (= number of coefficients to be estimated) and thus the tractability of the model. Furthermore, the corresponding model might not be consistent with the utility-maximizing theory of MEV models. In our model, spatial dependencies are at least partially incorporated in the stochastic part of utility, which causes correlation of certain alternatives and hence leads to the implementation of the models described in section "Two closed-form discrete choice models with flexible substitution patterns." Spatial correlation could arise because of various reasons: for example, if two or more schools

- are located along the route of parents taking their children to school during their commute to work;
- are located near a transit stop served by many transit lines:
- use the same (sports) facilities; or
- are located in the same neighborhood or district.[4]

To account for spatial substitution, we group nearby schools into nests (spatial nests). These nests are imposed in order to capture spatial similarity or correlation only. Additional purposes are conceivable, but this would lead to even more complicated (multilevel) nesting structures (Daly and Bierlaire 2006; Müller 2008). For the basic specification, spatially proximate schools have been pooled into one of six nests, depending on the distance between pairs of school locations (Hartigan and Wong 1979). Within the process of specifying the model structure, we found that $K = 6$ results in a reasonable grouping of schools. A large number of nests would probably yield restrictive substitution patterns (i.e., only two or three schools are assumed to share unobserved common attributes). This, in turn, may result in a remarkable number of insignificant nest parameters. A large number of feasible nesting structures can be found, and other values for *K* are possible. However, to have more or fewer nests results in fewer or more alternatives per nest, which complicates the finding of similarities between alternatives (a trade-off exists between number of nests and schools per nest). In our study, the ratio nest/alternatives is 6/26 = 0.23, which is close to the nesting ratio of 3/15 = 0.20 employed by Berkovec and Rust (1985), for example, who analyze the car choice of households. Similar ratios for nesting structures can be found in Gelhausen (2006) and Bhat (1998), with respective values of 6/21 = 0.28 and 3/15 = 0.20.

Finally, the decision about the overall nesting structure is subject to the discretion of a modeler. A researcher can impose an a priori structure. To determine the initial nesting structure shown in Fig. 4a, we employ R 2.10.0 and the *k-Means* function of the package *stats* version 2.10.1.[5] An alternative approach is to search all possible nesting structures that might result in a large number of distinct structures for even moderate choice sets (Hensher and Button 2000, p. 216). For the estimation of the model coefficients and parameters, we use the public domain software package Biogeme 1.8 (Bierlaire 2003, 2008).

Results

Throughout the model-building process, we found several more or less equivalent specifications (for both V_{ni} and the nesting structure). Table 2 summarizes the estimation results for the MNL, NL, and GNL models. Maranzo and Papola (2008) show that for a given feasible substitution pattern, an infinite number of associated GNL specifications may exist. The nesting structures for the NL and GNL model specifications are as follows. For the NL model,

- Nest 1: ANNE, FL, VITZ, MC
- Nest 2: MAN, EVKZ, HE
- Nest 3: STBE, BB, JOHA
- Nest 4: PLAU, COTT, JAS
- Nest 5: DKS, DKS2, WALD, RORO, PEST, KLOT
- Nest 6: GZW, JAH, WUST,

and for the GNL model,

- Nest 1: ANNE, FL, VITZ, MC, **PLAU**
- Nest 2: MAN, EVKZ, **HE**
- Nest 3: **STBE**, BB, JOHA, **HE**
- Nest 4: **PLAU**, COTT, JAS
- Nest 5: DKS, DKS2, WALD, RORO, PEST, KLOT, **STBE**
- Nest 6: GZW, JAH, WUST.

As can be seen from these nesting structures, the GNL model allows three alternatives to enter two nests (the corresponding short names are in bold). For the NL structure, we found the following relationships: the variable distance between a student's home and the school location is considered to be semialternative-specific instead of generic.[6] Hence, three different coefficients have been estimated for this variable: one for magnet schools ($\beta_{1.2} = -0.477$), which offer a unique profile, one for private schools ($\beta_{1.3} = -0.454$), and one for all others ($\beta_{1.1} = -0.573$). The corresponding coefficients (1.1–1.3) are significant, as measured by the value of the asymptotic *t*-test. This result indicates that for a two-tailed test, the respective coefficients differ from zero at the frequently used significance level of 0.05 (Ben-Akiva and Lerman 1985, pp. 161–2). Because we expected distance to be nonlinear, we tested selected modifications of the distance variable, such as the log of distance, a power series, and piecewise linearization. However, the linear specification presented in Table 2 yields the best model fit.

These results show that students enrolled in private or magnet schools are less sensitive to distance than others. This finding implies that a trade-off exists between distance traveled to school and the degree of

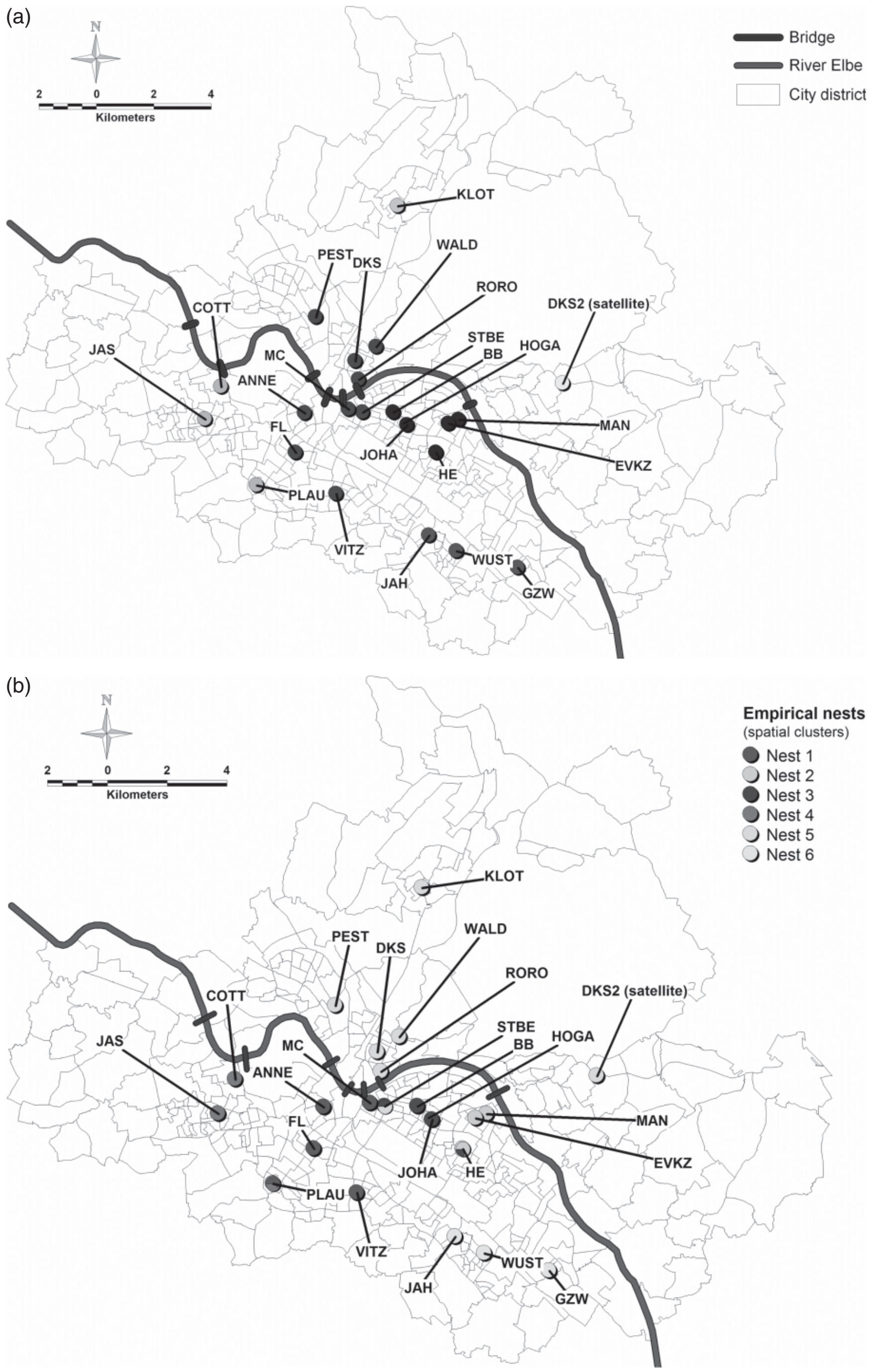


Figure 4. (a) Spatial nests of schools predetermined by cluster analysis. Schools that are allocated to the same nest show the same shading. (b) Empirically determined nests (by the generalized nested logit model). HE, PLAU, and STBE are proportionately allocated to different nests.

Table 2 Estimates for MNL, NL, and GNL Models

Variable number	Description	MNL Coeff. estimate	MNL t-stat	NL Coeff. estimate	NL t-stat	GNL Coeff. estimate	GNL t-stat
Distance variable							
1.1	Distance student-school in km (other schools)	–0.660	–45.44	–0.573	–48.55	–0.571	–48.89
1.2	stance student-school in km (magnet schools)	–0.544	–28.79	–0.477	–38.33	–0.488	–40.05
1.3	Distance student-school in km (private schools)	–0.465	–23.42	–0.454	–38.65	–0.470	–40.99
Other spatial variables							
2	Distance to school ≤1 km*	0.452	7.77	0.165	3.61	0.0703	1.93
3	Distance school-CBD in km	–0.142	–8.42	–0.0322	–2.71	–0.093	–7.28
4	School and student on same side of Elbe*	0.337	4.95	0.554	10.67	0.399	8.54
5	Site location of school's main campus*	–0.789	–2.97	–1.06	–8.54	–0.964	–8.34
6	School size (no. classes per grade)	0.230	10.23	0.273	13.77	0.234	13.80
7	Private school: no profile*	–2.14	–4.63	–1.22	–6.41	–1.03	–7.06
8	Private school: profile math & languages & music/arts*	–1.81	–4.45	–0.720	–4.17	–0.555	–4.50
9	Private school × avg. neighborhood income in thousand euro	0.951	4.20	0.821	9.75	0.708	11.57
10	Profile math & languages*	0.111	1.58	0.0819	1.59	0.192	3.75
11	Profile math & music/arts*	0.359	5.03	0.386	7.34	0.466	9.54
12	Profile languages core*	0.560	4.52	0.867	9.54	0.775	8.87
13	Profile languages core × gender male*	–0.983	–7.14	–0.839	–7.03	–0.799	–6.87
14	Profile math core*	0.221	–1.57	0.867	11.72	0.822	11.98
15	Profile math × gender male*	0.566	3.95	0.112	3.15	0.087	3.23
Nest parameters‡							
$\mu C1$	ANNE, FL, VITZ, MC (PLAU)			1.00	§	1.04	19.82†
$\mu C2$	MAN, EVKZ, HE			4.73	10.00†	6.29	11.71†
$\mu C3$	STBE, BB, JOHA (HE)			1.12	19.58†	3.82	7.37†
$\mu C4$	PLAU, COTT, JAS			1.76	17.95†	4.32	7.54†
$\mu C5$	DKS, DKS2, WALD, RORO, PEST, KLOT (STBE)			1.29	26.05†	1.29	23.55†
$\mu C6$	GZW, JAH, WUST			1.18	13.71†	1.16	12.79†

Table 2 *Continued*

Variable number	Description	MNL		NL		GNL	
		Coeff. estimate	*t*-stat	Coeff. estimate	*t*-stat	Coeff. estimate	*t*-stat
α's							
	αC1_PLAU					0.599	10.60
	αC4_PLAU					0.401	7.10
	αC3_STBE					0.410	7.10
	αC5_STBE					0.590	10.23
	αC2_HE					0.710	30.15
	αC3_HE					0.290	12.33
	Number of observations	5,215		5,215		5,215	
	$\mathcal{L}(0)$	−15,233.325		−15,233.325		−15,233.325	
	$\mathcal{L}(\hat{\beta})$	−9,883.091		−9,672.115		−9,572.170	
	$\bar{\rho}^2$	0.350		0.364		0.370	

*Dummy variables.

†t-test for nest parameters against 1.

‡Alternatives in brackets are considered in GNL only.

§Nest 1 is fixed due to modeling reasons.

MNL, multinomial logit; NL, nested logit; GNL, generalized nested logit.

specialization.[7] The models with a generic distance variable and the semialternative-specific distance variable have been tested with a log-likelihood ratio test. The semialternative-specific distance variable outperforms the generic one at a 0.05 significance level. Due to space restrictions, we report the first specification only. Moreover, we consider a dummy variable that indicates whether a school is situated less than 1 km from a student's home. The positive sign of the corresponding coefficient indicates that schools within walking distance are favored. Variable 3 denotes the distance between a school and the central business district (CBD), which is a measure of location of that school. The negative sign of the corresponding coefficient indicates that schools located near the city center are more attractive than others. This measure is a proxy for the accessibility of schools. In Dresden, the transit system has a more or less radial network, and hence, schools near the central node are more accessible than others.[8] In addition, parents who bring their children to school on their work commute are more likely to go to the CBD or at least pass the CBD. The trade-off between the distance school-CBD and the distance traveled to school supports this interpretation:

$$\frac{\partial U_{ni}/\partial \text{ distance school} - \text{CBD}}{\partial U_{ni}/\partial \text{ distance student} - \text{school}} = \frac{\hat{\beta}_5}{\hat{\beta}_1}, \qquad (15)$$

which is 0.056 for the NL and 0.163 for the GNL model specification. In the GNL model, each 1-km increase in distance between a school and the CBD tends to be compensated for by a decrease of 0.163 km in the distance traveled by a student to school without affecting the utility of the student. For constant utility, more peripherally located schools have smaller catchment areas than central ones. According to equation (15), 1 km traveled to a central school has higher utility compared with 1 km traveled to a school located in the outskirts. Variables 4 and 5 have not been considered in the first place in the estimation. An outlier analysis points to observations with poorly predicted choice probabilities. First, there was an overprediction for DKS2, which is a site location of DKS's main campus. Without variable 7, the predicted market share for DKS2 was far too high. The negative coefficient of this variable corrects for this misprediction. Furthermore, for some observations, the predicted choice probabilities for nonchosen schools are remarkably high, an issue especially related to schools located on the opposite side of the river Elbe, based on a student's place of residence. Introducing variable 4, corrects the predicted choice probabilities of the nonchosen schools.

The remaining school attributes suggest a number of implications. Larger schools tend to be more attractive than smaller ones (variable 6) because large schools are less likely to deny enrollment based on capacity constraints. The dummy variables 7 and 8 indicate that private schools are less preferred than public schools, because most private schools often are associated with school fees, and some of them additionally have religious affiliation restrictions. For private schools that offer a wide range of profiles, the disutility is less remarkable (variable 8). Nevertheless, the higher the average neighborhood income, the more attractive private schools become (variable 9). The most requested profile is the math core (variable 14), particularly by male students (variables 14 and 15). This preference is followed by the languages core profile if students are female (variable 12). Less attractive are standard profiles (variables 10 and 11) and the languages core profile if students are male (variables 12 and 13).[9]

As expected, the nest parameters μ_{C2} to μ_{C6} are significantly different from one, indicating similarities between nearby schools and the failure of IIA in the standard logit case. Significant nest parameters with values consistent with utility maximization (i.e., if $0 < 1/\mu_k < 1$) are a sufficient indicator that IIA does not hold for an MNL in the school choice context (Train 2003, p. 54).The nesting coefficients presented in Table 2 for both the NL and the GNL models indicate that IIA holds for alternatives that are in the same nest but not for those across nests (Train 2003, p. 82). Nest 1 was fixed for the estimation of the NL model. Although alternative feasible nesting structures other than the one presented for the NL model in Table 2 could have been found, this option was abandoned here in favor of comparability between the NL and GNL models. Both models (NL and GNL) are normalized from the top (i.e., $\mu = 1$). The presumed spatial substitution pattern (Fig. 4a) is empirically confirmed to some extent (Fig. 4b). The strongest spatial

Table 3 Cross-Elasticities with Respect to the *Distance to CBD*-Variable for Nest 3 (GNL) and Two Alternatives from Other Nests*

	HE	BB	STBE	JOHA	ANNE	COTT
HE	–4.738	2.284	1.437	2.691	0.045	0.0133
BB	0.105	–6.285	2.499	6.664	0.045	0.0133
STBE	0.024	0.899	–0.266	1.059	0.045	0.0133
JOHA	0.105	5.657	2.499	–5.217	0.045	0.0133
ANNE	0.0135	0.292	0.810	0.344	–2.780	0.0133
COTT	0.0135	0.292	0.810	0.344	0.045	–17.9581

*School names printed in bold indicate the nest membership.
CBD, central business district; GNL, generalized nested logit.

Table 4 Cross-Elasticities with Respect to the *Distance to CBD*-Variable for Nest 3 (NL) and Three Alternatives from Other Nests*

	HE	BB	STBE	JOHA	ANNE	COTT
HE	–11.667	0.175	0.298	0.179	0.073	0.024
BB	0.066	–3.230	0.443	0.277	0.073	0.024
STBE	0.066	0.271	–2.358	0.277	0.073	0.024
JOHA	0.066	0.271	0.443	–3.206	0.073	0.024
ANNE	0.066	0.175	0.298	0.179	–2.649	0.024
COTT	0.066	0.175	0.298	0.179	0.073	–6.955

*School names printed in bold indicate the nest membership.
CBD, central business district; NL, nested logit.

similarity exists between schools in nest C2 (μ_{C2}). However, particularly near the CBD, the substitution patterns and thus correlation between schools are somewhat more complicated than had been assumed. Instead of having two large nests, as presumed, we empirically find six smaller nests. Moreover, all schools north of the river Elbe are allocated in one nest. Although we account for the separation effect of the river with variable 6, schools north of the river share some unobserved spatial factors. This indicates that spatial similarity exhibits substitution effects that are difficult to account for in the deterministic part of utility (V_{ni}). Comparing the results between the NL and the GNL models, the coefficients of the deterministic utility functions are mostly similar under different error structures. This outcome signifies the reliability of the model specification. However, concerning the small difference in coefficients between the NL and the GNL models, most of the GNL model coefficients are smaller in magnitude than the NL model coefficients.

The relationship among pairs of alternatives in the NL and the GNL models can be examined further by comparing the respective cross-elasticities, or the proportional change in the choice probability of an alternative with respect to a proportional change in an explanatory variable of another alternative (Koppelman and Bhat 2006, p. 50). The elasticity increases between pairs of alternatives as the corresponding value of $1/\mu_k$ decreases from one. The magnitude of this effect is further related to the choice probability of the respective nest and the conditional probability of the alternatives in that nest. This effect can also be seen in Tables 3 and 4, which include the cross-elasticities for alternatives in nest 3 and alternatives outside the nest associated with a change in the distance to the CBD variable.

The elasticity measure for distance to CBD, for example, can be used to evaluate a relocation. A change in the distance to CBD for a given school occurs if that school is relocated for a certain period of time due to extensive renovation of the original school building. In the GNL model, a 1% change in the respective attribute of alternative BB, for instance, causes a 2.5% change in the choice probability of STBE, which is in the same nest (Table 3). For an alternative outside the nest, like ANNE, the respective change in choice probability is only 0.045%, which is disproportionately small as these alternatives do not share a common nest. Thus, alternatives that share a common nest are much better substitutes for each other than alternatives from different nests. The given values of the elasticities quantify this distinction concerning substitutability. Furthermore, in case of the GNL model, the fraction of each alternative included in one or more common nests determines the implied correlation and substitution between alternatives (Hensher and Button 2000, p. 218). For our analysis, we chose schools allocated to more than one nest that are located nearby schools of a different nest. Hence, STBE, HE, and PLAU are assigned to a second nest, as displayed in Fig. 4b. Several GNL model specifications lead to the feasible, reasonable, and easy-to-interpret model presented in Table 2. STBE, located south of the river, exhibits spatial similarity with schools north of the river. Allocation parameters α_{C3_STBE} and α_{C5_STBE} indicate that STBE is allocated to nest C5 by nearly 60% and to nest C3 by 40%. Thus, STBE shares stronger common unobservable attributes with schools of nest C5 than with schools of nest C3. This is a reasonable finding that may be explained by the bridges across the river Elbe surrounding the area around STBE. We further derive correlation matrices from the NL and the GNL models, which are displayed in Tables A1 and A2, respectively, in the Appendix. The eighth row of Table A2 (STBE) documents the advantage of the GNL model. Because STBE is allocated to nests C3 and C5, STBE is correlated with many more schools than indicated by a simple NL model. Besides this flexibility in substitution patterns, the GNL model yields a higher log-likelihood ($\mathcal{L}\left[\hat{\beta}\right]$ in Table 2). We can reject the null hypothesis that the NL and the GNL models are equivalent at the 0.05 level of significance using a nonnested hypothesis test (Ben-Akiva and Lerman 1985, p. 171ff.).

Conclusion

Due to the possibility of free school choice and fluctuating student numbers, schools in Germany face increasing competition, which can be seen particularly in the expanding number of profiles (e.g., math, languages, sciences, arts) or extracurricular activities offered by schools to stimulate enrollment and, thus, avoid school closings. To analyze the effects of changes in the school network, mid- and long-term forecasts of demographic trends as well as of students' decisions about school choice needs to be taken into account to derive possible future scenarios. Therefore, we feel that choice models reproducing the decision-making processes of individuals that are as realistic as possible (i.e., choice models accounting for spatial substitution) are a valuable instrument in school planning and school assignment. Until now, literature about school network planning seems to have ignored spatial substitution between competing school locations. Moreover, school choice literature with a focus on spatial substitution is scarce. The model presented here explicitly accounts for spatial substitution. Fortunately, the model still takes a computationally convenient closed-form. We can verify most of the findings in the literature concerning the variables that enter the systematic part of utility, like school size and travel-to-school distance. Moreover, we find new evidence about spatial effects. First, we see that the catchment area of a school (based on constant utility values) decreases in relation to increased distance from the CBD. Second, our analysis shows that a significant and remarkable correlation exists between schools within proximity to one another. Furthermore, correlation patterns are allowed to vary due to a flexible allocation of schools to nests. From a methodological perspective, more sophisticated approaches are worth using (i.e., discrete choice models based on utility-maximizing behavior) in order to attain more insights into the spatial patterns of locational choice. Through relaxation of the distinct membership of a school to one nest, we incorporate spatially overlapping substitution effects. This analysis strongly suggests that spatial substitution should be focused on more when designing a school network. Accordingly, empirically determined substitution between locations should be accounted for in location–allocation problems and urban models in general.

Acknowledgements

We thank four anonymous referees for their comments on an earlier version of this article. We are also very grateful to the editor for his editing of this article. Of course, we retain responsibility for all remaining errors, omissions, and opinions.

Notes

1 This model is similar to the cross-nested logit (CNL) model by Vovsha 1997.
2 We do not, however, incorporate the variables of mean test score and religious affiliation in our study due to lack of information. Race does not play an important role within the school choice process in most eastern German cities (except Berlin) because the percentage of students of color is very low (≤10%).
3 If actual market shares are not known, one can use (under certain conditions) the weighted conditional maximum likelihood (WCML) estimator by Bierlaire, Bolduc, and McFadden (2008).
4 If two schools are located in the same neighborhood, as perceived by a student (Guo and Bhat 2007), we expect that they are more correlated to each other than to other schools.
5 **R** is a programming language and software environment for statistical computing and graphics.
6 A full alternative specific specification yields $I - 1$ coefficients.
7 We consider private and magnet schools as specialized schools.
8 Nearly 60% of students enrolled at Gymnasium schools in Dresden choose public transport for their commute to school (Müller, Tscharaktschiew, and Haase 2008).
9 Profile math is the reference category.

Appendix

Table A1 Correlation Matrix of the NL Model

	MC	ANNE	FL	VITZ	MAN	EVKZ	HE	STBE	BB	JOHA	PLAU	COTT	JAS	DKS	DKS2	WALD	RORO	PEST	KLOT	GZW	JAH	WUST
MC	1																					
ANNE		1																				
FL			1																			
VITZ				1																		
MAN					1	0.96	0.96															
EVKZ					0.96	1	0.96															
HE					0.96	0.96	1															
STBE								1	0.20	0.20												
BB								0.20	1	0.20												
JOHA								0.20	0.20	1												
PLAU											1	0.68	0.68									
COTT											0.68	1	0.68									
JAS											0.68	0.68	1									
DKS														1	0.40	0.40	0.40	0.40	0.40			
DKS2														0.40	1	0.40	0.40	0.40	0.40			
WALD														0.40	0.40	1	0.40	0.40	0.40			
RORO														0.40	0.40	0.40	1	0.40	0.40			
PEST														0.40	0.40	0.40	0.40	1	0.40			
KLOT														0.40	0.40	0.40	0.40	0.40	1			
GZW																				1	0.28	0.28
JAH																				0.28	1	0.28
WUST																				0.28	0.28	1

Table A2 Correlation Matrix of the GNL Model

	MC	ANNE	FL	VITZ	MAN	EVKZ	HE	STBE	BB	JOHA	PLAU	COTT	JAS	DKS	DKS2	WALD	RORO	PEST	KLOT	GZW	JAH	WUST
MC	1	0.754	0.754	0.754							0.578											
ANNE	0.754	1	0.754	0.754							0.578											
FL	0.754	0.754	1	0.754																		
VITZ	0.754	0.754	0.754	1																		
MAN					1	0.8912	0.794															
EVKZ					0.8912	1	0.794															
HE					0.794	0.794	1	0.2875	0.426	0.426												
STBE							0.2875	1	0.5356	0.5356				0.2997	0.2997	0.2997	0.2997	0.2997	0.2997			
BB							0.426	0.5356	1	0.923												
JOHA							0.426	0.5356	0.923	1												
PLAU	0.578	0.578	0.578	0.578							1	0.5349	0.5349									
COTT											0.5349	1	0.9283									
JAS											0.5349	0.9282	1									
DKS								0.2997						1	0.3991	0.3991	0.3991	0.3991	0.3991			
DKS2								0.2997						0.3991	1	0.3991	0.3991	0.3991	0.3991			
WALD								0.2997						0.3991	0.3991	1	0.3991	0.3991	0.3991			
RORO								0.2997						0.3991	0.3991	0.3991	1	0.3991	0.3991			
PEST								0.2997						0.3991	0.3991	0.3991	0.3991	1	0.3991			
KLOT								0.2997						0.3991	0.3991	0.3991	0.3991	0.3991	1			
GZW																				1	0.2568	0.2568
JAH																				0.2568	1	0.2568
WUST																				0.2568	0.2568	1

References

Ben-Akiva, M., and S. Lerman. (1985). *Discrete Choice Analysis, Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.
Berkovec, J., and J. Rust. (1985). "A Nested Logit Model of Automobile Holdings for One Vehicle Households." *Transportation Research Part B: Methodological* 19(4), 275–85.
Bhat, C. (1998). "Analysis of Travel Mode and Departure Time Choice for Urban Shopping Trips." *Transportation Research Part B: Methodological* 32(6), 361–71.
Bierlaire, M. (2003). "BIOGEME: A Free Package for the Estimation of Discrete Choice Models." Proceedings of the Swiss Transport Research Conference. Ascona, Switzerland.
Bierlaire, M. (2006). "A Theoretical Analysis of the Cross-Nested Logit Model." *Annals of Operations Research* 144(1), 287–300.
Bierlaire, M. (2008). Estimation of Discrete Choice Models with BIOGEME 1.8, biogeme.epfl.ch.
Bierlaire, M., D. Bolduc, and D. McFadden. (2008). "The Estimation of Generalized Extreme Value Models from Choice-Based Samples." *Transportation Research B: Methodological* 42(4), 381–94.
Bolduc, D., B. Fortin, and M. Fournier. (1996). "The Effect of Incentive Policies on the Practice Location of Doctors: A Multinomial Probit Analysis." *Journal of Labor Economics* 14(4), 703–32.
Borgers, A., H. Oppewal, M. Ponjé, and H. Timmermans. (1999). "Assessing the Impact of School Marketing: Conjoint Choice Experiments Incorporating Availablility and Substitution Effects." *Environment and Planning A* 31, 1949–64.
Cullen, J., B. Jacob, and S. Levitt. (2005). "The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools." *Journal of Public Economics* 89(5–6), 729–60.
Daly, A., and M. Bierlaire. (2006). "A General and Operational Representation of Generalised Extreme Value Models." *Transportation Research Part B: Methodological* 40(4), 285–305.
Gelhausen, M. (2006). "Flughafen- und Zugangsverkehrsmittelwahl in Deutschland—Ein verallgemeinerter Nested Logit-Ansatz." Proceedings of the Deutscher Luft-und Raumfahrtkongress. Braunschweig.
Guo, J. Y., and C. R. Bhat. (2007). "Operationalizing the Concept of Neighborhood: Application to Residential Location Choice Analysis." *Journal of Transport Geography* 15, 31–45.
Hartigan, J., and M. Wong. (1979). "Algorithm AS 136: A K-Means Clustering Algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28(1), 100–8.
Hastings, J., T. Kane, and D. Staiger. (2006). "Parental Preferences and School Competition, Evidence from a Public School Choice Program." National Bureau of Economic Research Working Paper No. 11805.
Haynes, K., and A. Fotheringham. (1990). "The Impact of Space on the Application of Discrete Choice Models." *The Review of Regional Studies* 20, 39–49.
Haynes, K., D. Good, and T. Dignan. (1988). "Discrete Spatial Choice and the Axiom of Independence from Irrelevant Alternatives." *Socio-Economic Planning Science* 22(6), 241–51.
Heiss, F. (2002). "Structural Choice Analysis with Nested Logit Models." *The Stata Journal* 2(3), 227–52.
Hensher, D., and K. E. Button. (2000). *Handbook of Transport Modelling*. Oxford, U.K: Pergamon.
Hunt, L., B. Boots, and P. Kanaroglou. (2004). "Spatial Choice Modelling: New Opportunities to Incorporate Space into Substitution Patterns." *Progress in Human Geography* 28, 746–66.
Jepsen, C., and M. Montgomery. (2009). "Miles to Go before I Learn: The Effect of Travel Distance on the Mature Person's Choice of Community College." *Journal of Urban Economics* 65, 64–73.
Koppelman, F., and C. Bhat. (2006). A Self-Instructing Course in Mode Choice Modelling: Multinomial and Nested Logit Models. Tech. rep. U.S. Department of Transportation Federal Transit Administration.
Koppelman, F. S., and V. Sethi. (2000). "Closed-Form Discrete-Choice Models." In *Elsevier's Handbooks in Transport 1*, 211–27, edited by D. A. Hensher and K. J. Button. Vol. Handbook of Transport Modelling. Amsterdam, Netherlands: Pergamon, Chap. 13.
Lankford, H., E. Lee, and J. Wyckoff. (1995). "An Analysis of Elementary and Secondary School Choice." *Journal of Urban Economics* 38, 236–51.
Lankford, H., and J. Wyckoff. (2006). "The Effect of School Choice and Residential Location on the Racial Segregation of Students." In *Improving School Accountability—Check-Ups or Choice, Advances in Applied Microeconomics, Volume* 14. 185–239, edited by T. Gronberg and D. Jansen. Albany, NY: Elsevier.
Luce, R. (1959). *Individual Choice Behaviour: A Theoretical Analysis*. New York: Wiley.
Manski, C., and D. Wise. (1983). *College Choice in America*. Cambridge, MA: Harvard University Press.
Maranzo, V., and A. Papola. (2008). "On the Covariance Structure of the Cross-Nested Logit Model." *Transportation Research B: Methodological* 42(2), 83–98.
McFadden, D. (1975). On Independence Structure and Simultaneity in Transportation Demand Analysis. Working Paper 7511. Berkeley: Department of Economics, University of California.
McFadden, D. (1978). "Modeling the Choice of Residential Location." *Spatial Interaction Theory and Planning Models* 25(477), 75–96.

Müller, S. (2008). *Dynamic School Network Planning in Urban Areas*, 5. Stadt- und Raumplanung Stadt- und Raumplanung. Berlin: LIT Verlag.

Müller, S. (2009). "A Spatial Choice Model Based on Random Utility." *Journal of the Institute of Transport and Economics* 11(2), 2–25.

Müller, S., K. Haase, and S. Kless. (2009). "A Multi-Period School Location Planning Approach with Free School Choice." *Environment and Planning A* 41(12), 2929–45.

Müller, S., S. Tscharaktschiew, and K. Haase. (2008). "Travel-to-School Mode Choice Modelling and Patterns of School Choice in Urban Areas." *Journal of Transport Geography* 16(5), 342–57.

Neu, M. (2007). "Sozialraumstrukturen im Wandel: Eine Längsschnittanalyse des Essener Stadtgebietes 1970-1987-2006." Diskussionspapiere aus der Fakultät für Sozialwissenschaft, Ruhr-Universität Bochum 2007-1.

Schneider, T. (2004). Der Einuss des Einkommens der Eltern auf die Schulwahl. Vol. 446. Discussion Papers. Berlin: DIW.

Timmermans, A., and P. van der Waerden. (1992). "Mother Logit Analysis of Substitution Effects in Consumer Shopping Destination Choice." *Journal of Business Research* 24(2), 177–89.

Train, K. (1999). "Mixed Logit Models for Recreation Demand." In *Valuing Recreation and the Environment*, 121–40, edited by J. Herriges and C. Kling. Northhampton, MA: Edward Elgar.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge, MA: Cambridge University Press.

Vovsha, P. (1997). "Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel Metropolitan Area." *Transportation Research Record: Journal of the Transportation Research Board* 1607, 6–15.

Walker, J., and J. Li. (2007). "Latent Lifestyle Preferences and Household Location Decisions." *Journal of Geographical Systems* 9(1), 77–101.

Economics of Innovation and New Technology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gein20

The adoption of photovoltaic systems in Wiesbaden, Germany

Sven Müller [a] & Johannes Rode [b]
[a] Institute for Transport Economics, University of Hamburg , Von-Melle-Park 5, 20146 , Hamburg , Germany
[b] Graduate School for Urban Studies , Technische Universität Darmstadt , Bleichstrasse 2, 64283 , Darmstadt , Germany
Published online: 09 Jun 2013.

Economics of Innovation and New Technology, 2013
Vol. 22, No. 5, 519–535, http://dx.doi.org/10.1080/10438599.2013.804333

The adoption of photovoltaic systems in Wiesbaden, Germany

Sven Müller[a] and Johannes Rode[b]*

[a]*Institute for Transport Economics, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany;* [b]*Graduate School for Urban Studies, Technische Universität Darmstadt, Bleichstrasse 2, 64283 Darmstadt, Germany*

(*Received 4 September 2012; final version received 3 May 2013*)

The purpose of this study is to investigate factors that determine the adoption of photovoltaic (PV) systems. Our case study of the city of Wiesbaden, Germany, is based on a geocoded data set of the grid-connected PV systems set up through 2009. We aim to determine whether the decision to install can be explained by peer effects measured by preexisting installations in the vicinity, i.e. the installed base which is determined for each decision-maker individually. We employ a binary panel logit model and control for spatial variations in buying power and population density. Our analysis reveals a significantly positive influence of previously installed systems located nearby on the decision to install a PV system.

Keywords: photovoltaics; solar; diffusion; adoption; imitation; discrete choice; installed base; binary panel logit

JEL Classification: O33; C35; Q55; R10

1. Introduction

Photovoltaics (PV), i.e. solar cell systems to produce electric power, may be a sustainable alternative to finite and fossil fuels causing climate change. Therefore, factors influencing the adoption decision of this technology are interesting for political decision-makers, managers and economists. The purpose of this study is to investigate some of these factors.

In general, the decision to adopt a technology might be hampered by uncertainty regarding, e.g. the technology's reliability and its net present value. The uncertainty might be reduced if there is an information spillover from peers to the decision-maker. To put it in the words of Rogers' (1983) pioneering work on innovation diffusion: as PV systems are highly *observable*, previous installations nearby may decrease the uncertainty of the utility of an installation by illustrating that PV are *compatible* with prevailing norms, *incomplex* to adopt and indeed work effectively, the latter also referred to as *trialability*.

Bollinger and Gillingham (2012) study the diffusion of PV in California. They analyze the aggregate level and find evidence for peer effects, thus 'at the average number of owner-occupied homes in a zip code, an additional installation increases the probability of an adoption in the zip code by 0.78 percentage points' (900). In Bollinger and Gillingham's study, peers are defined as other decision-makers, e.g. in the same zip code area. However,

*Corresponding author. Email: rode@stadtforschung.tu-darmstadt.de

aggregating individual adoption decisions to larger spatial units will bias the 'peer effect'. In contrast, analyses on the individual level help to identify the influence of peers on the decision to adopt (Valente 1996). Therefore, we analyze peer effects in PV adoption on the individual decision level.

Based on this, we might assume that the choice of a decision-maker to install PV is influenced by peers, and the propensity of being a peer is higher for proximate locations than for locations far away. Now – in line with Tobler's (1970) First Law of Geography – our hypothesis concerning peer effects is: the more proximate PV systems exist in a preceding period, the higher the probability that a decision-maker will choose to install PV in the current period. The number of proximate PV systems (i.e. the number of peers who already use the new technology, that is, the number of adopters from earlier periods) is usually referred to as the installed base (see Farrell and Saloner [1986], Majumdar [1996] or Koski [1999], for example). In contrast to Bollinger and Gillingham (2012), we determine the installed base for each potential adopter of PV individually and use a distance-based measure.

When studying the adoption of innovations on the individual level, economic scholars often employ the theory of social learning. This theory emphasizes differences between individuals and assumes that utility is maximized by those adopting (Young 2009). Some simulation studies on social learning account for a spatial dimension (Ellison and Fudenberg 1993) or learning from neighbors (Bala and Goyal 1998), also revealing that space matters. Delre, Jager, and Janssen (2007) show that considering decisions to adopt a new product or technology on the individual level yields promising results regarding the speed of adoption in a (social) network. Managers or political decision-makers might use these results in order to justify 'installation seeds' to foster the adoption of new technologies (in certain regions).

One possibility to study adoption decisions on the individual level are threshold models of innovation diffusion (Bonus 1973; Kemp 1998). This approach assumes that when a stimulus variate exceeds a critical threshold the innovation is adopted. It would be reasonable to assume that a potential user will purchase the innovation when its net present value is positive. Unfortunately, we lack information on the net present value of PV for our period of analysis. Furthermore, since only less than 1% of the potential users adopt the technology during our period of study, we may not predefine under which circumstances the innovation is adopted but may model the adoption decision itself.

Rogers (1983) describes several stages of the innovation-decision process. In his terminology the decision stage is the point at which – after information on the innovation has been acquired – the actual decision on whether the innovation is adopted or rejected is made. Since the decision to adopt a new technology (in a certain period of time) is binary (Karshenas and Stoneman 1992), we consider a discrete (binary) choice panel model in our analysis. Discrete choice analysis is the standard approach to analyzing individual discrete decision-making (McFadden 2001). Particularly, Geroski (2000) highlights that when focusing on differences in adopter characteristics a probit model, which is one specific discrete choice model, is appropriate. Discrete choice models have been widely applied to analyze decision-making in energy markets (see, for example, Bernard, Bolduc, and Bélanger 1996; Nesbakken 2001; Wilson and Dowlatabadi 2007). We estimate a binary panel logit model (BPL) for different periods of PV adoption.[1]

The temporal dimension of our data also allows us to test whether peer effects vary over time. Certainly, for very early adopters a peer effect can not be observed as no or too few preexisting users exist. However, according to Rogers (1983, 166–167) early adopters 'have more exposure to interpersonal channels', i.e. we hypothesize that the influence of peer effects may increase (in the very early phases of the diffusion path) over time and may decrease later on.[2] Furthermore, early adopters have 'higher socioeconomic status than late[r]' adopters. Accordingly, we hypothesize that measures indicating high income and low population density – i.e. a high share of single- and double-family homes – are associated with early adopters.

We apply our model to data from the city of Wiesbaden, Germany. Wiesbaden is selected as a case study since: (i) the yearly average of global radiation that every square meter in Wiesbaden receives is close to the German average (DWD 2010); (ii) roughly 70% of the German population lives in urban regions; (iii) in 2010, Wiesbaden had roughly 275,000 inhabitants, which means that it was one of the 25 largest German cities (BBSR 2010), i.e. Wiesbaden is an urban area which is neither exceptionally large nor small; (iv) similar to many German cities, Wiesbaden is situated next to a river, the River Rhine; and (v) comprehensive data is available for Wiesbaden.

Our geocoded data set covers the 324 grid-connected PV systems which were installed in Wiesbaden through the end of the year 2009. Thirteen of these were installed before the year 2000, when the incentive to install was low. Since the year 2000 – when the Renewable Energy Sources Act (Erneuerbare-Energien-Gesetz) put in place a strong legal incentive system – Germany has been attractive to PV investors. The legal incentive system for PV is based on a feed-in tariff. The feed-in tariff forces the electricity grid operators to accept that electricity produced by renewable energy is fed into the grid and guarantees a fixed remuneration for each kilowatt-hour produced by PV for 20 years (Altrock, Oschmann, and Theobald 2008).[3] In consequence, the installation of PV systems increased strongly: in Wiesbaden, 311 PV systems were installed between 2000 and 2009.

The remainder of the paper is as follows: we give a brief introduction to discrete choice analysis for binary panel data and random utility theory in Section 2. A description of the data and the model specification can be found in Section 3. Section 4 comprises the results and Section 5 summarizes the paper and gives an outlook on further research.

2. Discrete choice analysis for binary panel data

Consider a choice maker n (an individual or a household) who chooses in a certain period t one and only one alternative $i \in C_{nt}$. C_{nt} is the choice set, i.e. all alternatives n faces in t. Since we consider only two alternatives, i.e. 'to install' (PV) and 'to not install' (no PV) the choice problem is binary. We assume n perceives utility U_{nti} from choosing i in t. The choice maker n chooses to install a PV system in period t only if

$$U_{nt,\mathrm{PV}} > U_{nt,\mathrm{no\,PV}}. \tag{1}$$

For our behavioral model (1) we assume utility maximizing behavior. In discrete choice analysis the latent construct utility is decomposed into a deterministic (or systematic) part V_{nti} and a stochastic part ε_{nti}:

$$U_{nti} = V_{nti} + \varepsilon_{nti}. \tag{2}$$

Usually V_{nti} is linear in parameters:

$$V_{nti} = \sum_{h \in H} \beta_{ith} x_{ntih}. \tag{3}$$

The H independent variables x_{ntih} describe alternative i and characteristics of choice maker n in period t. Of course, lagged variables can also be used. The exogenous variables x_{ntih} are weighted by coefficients β_{ith}. Obviously, utility of Equation (2) is random and hence only probability statements on our behavioral model of Equation (1) can be made:

$$\begin{aligned} P_{nt,\mathrm{PV}} &= \mathrm{Prob}(U_{nt,\mathrm{PV}} > U_{nt,\mathrm{no\,PV}}) \\ &= \mathrm{Prob}(\varepsilon_{nt,\mathrm{PV}} < V_{nt,\mathrm{PV}} - V_{nt,\mathrm{no\,PV}} + \varepsilon_{nt,\mathrm{PV}}). \end{aligned} \tag{4}$$

Equation (4) denotes the choice probability of choice maker n to choose to install in period t. In order to operationalize the choice model of Equation (4), we have to make assumptions about the random components of utility ε_{nti} in Equation (2). If we assume ε_{nti} are independent, identically extreme-value distributed, then we obtain from Equation (4) by means of certain algebraic transformations

$$P_{nt,\mathrm{PV}} = \frac{e^{\mu V_{nt,\mathrm{PV}}}}{e^{\mu V_{nt,\mathrm{PV}}} + e^{\mu V_{nt,\mathrm{no\,PV}}}}, \tag{5}$$

where μ is a scale parameter >0 that is not identified and has to be set to an arbitrary value (e.g. 1) for model identification purposes. Equation (5) is the binary panel logit model (BPL). Note that Equation (5) has a closed form (in contrast to a probit model), and that the unknown coefficients β_{ih} of Equation (3) can be provided through maximum likelihood estimation:

$$\max_{\beta\in\mathbb{R}}^{H} \sum_{n=1}^{N}\sum_{t=1}^{T}(o_{nt,\mathrm{PV}} \ln P_{nt,\mathrm{PV}} + (1 - o_{nt,\mathrm{PV}}) \ln(1 - P_{nt,\mathrm{PV}})). \tag{6}$$

$o_{nt,\mathrm{PV}}$ equals 1 if we observe that choice maker n has installed a PV system in period t (0 otherwise). Therefore, $o_{nt,\mathrm{PV}}$ is the dependent variable and hence we are not able to measure U_{nti} directly. Furthermore, we are only able to identify utility differences due to Equation (1). As stated before, utility U_{nti} is a latent variable: the observable choices are manifestations of the underlying utilities described by exogenous variables. The modeling framework is given in Figure 1.

Discrete choice models like the BPL of Equation (5) are the workhorse in analyzing individual choice behavior (McFadden 1986, 2001). However, the BPL might exhibit a severe shortcoming depending on the specific choice situation and data to be analyzed: the assumption of independence of the error terms of Equation (2).

In our case the independence of the error terms over alternatives seems to be uncritical because the choice situation is binary and the two alternatives are antipodal. Although the assumption of independence (over periods and choice makers) might be violated in our

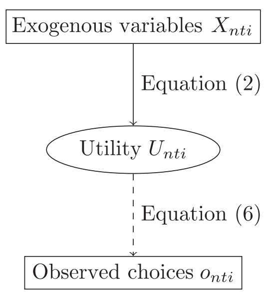


Figure 1. Modeling framework: rectangles represent observed data and the ellipse denotes a latent variable. The solid line represents a structural equation while the dashed line stands for a measurement equation.

case, it is well known from empirical studies that the inferences based on the estimates of (binary) logit models are fairly robust (Hensher and Greene 2003).

3. Building the model

In this section we discuss the operationalization of the installed base in order to account for peer effects on the decision to adopt PV systems. We also present details of the data used for model estimation and we specify the utility functions of the BPL.

3.1. *Data*

This study builds on a unique data set including the location and date of installation of the PV systems set up in Wiesbaden, Germany, through 2009.[4] Wiesbaden covers an area of somewhat more than 200 sqkm and is located in the middle of Germany, next to the river Rhine.

Since 80% of the PV systems in Germany are installed on roofs (BMU 2011), we consider buildings as the predominantly potential locations for PV systems.[5] Therefore, we are only interested in PV systems on buildings. Further we assume that each building is owned by someone. Eventually, this (artificial) person – the owner or the owner group – makes the decision to install in a certain period or not. However, due to data-related issues, we do not observe the choice directly. We only observe whether there is a PV system at a given location (building) in a given period or not. Of course, whether the building is owned by a private household or a house cooperation makes a difference. Unfortunately, we cannot account for that. To the best of our knowledge no other study on PV adoption, particularly those using aggregate approaches, accounts for this difference, e.g. by considering land owner properties (see Bollinger and Gillingham 2012; Rode and Weber 2012).

Our geocoded data set covers the 324 grid-connected PV systems, which were installed in Wiesbaden through the end of the year 2009. Totally, 297 PV systems are geocoded with an address level accuracy, 10 with street-level accuracy and 17 with post code level accuracy. Figure 2 illustrates the spatial dimension of the data.

We assume that each building in Wiesbaden can be equipped with one PV system. Wiesbaden's (2009) land surveying office provided spatial data for the 42,352 main buildings. Data clearance yields 41,666 observations left for analysis. As shown in Figure 3, not all PV systems fall inside a building-polygon. Therefore, 149 out of 324 PV systems are allocated to their most proximate building.[6] This procedure results in 324 PV systems allocated to 286 buildings, i.e. in 25 cases two or more PV systems are allocated to one building. This is due to inaccurate geocoding at the street or post code level or the fact that several PV systems are installed on one building. In order to follow our assumption of one PV system per building, 38 systems are reallocated to the nearest free building. As a consequence, our data set comprises 324 PV systems allocated to 324 buildings.

3.2. *Utility functions and operationalization*

The choice problem under consideration is for each n to choose to install PV in a given period or to not install (see Section 2). In order to test our hypothesis of peer effects as outlined in Section 1, we have to consider the influence of the choice of $m \in N$ on the choice of $n \in N$ in period t, that is, the dependencies between choice makers n and m.

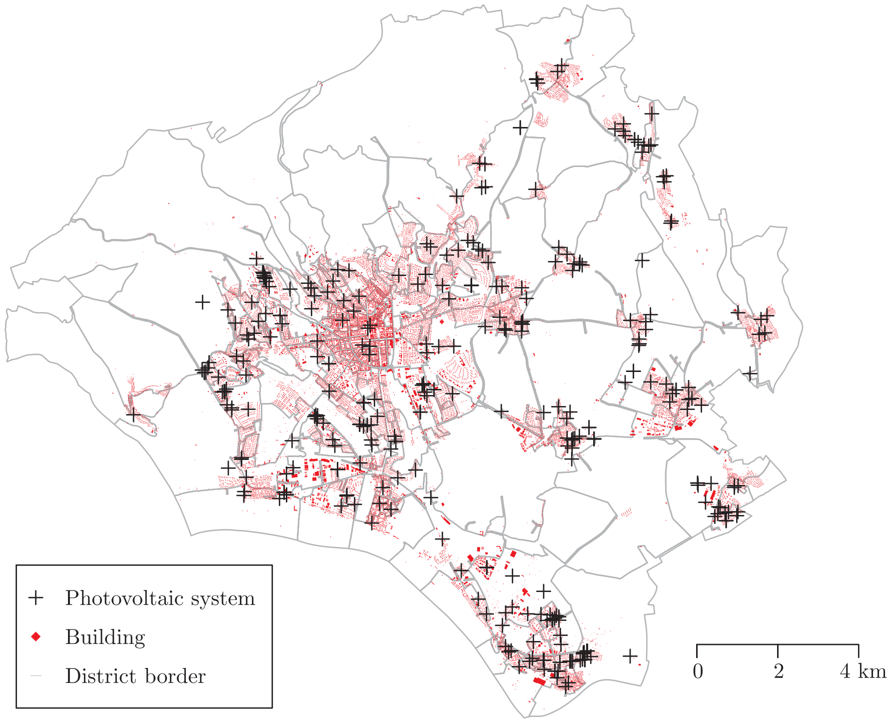


Figure 2. PV systems, buildings and borders of statistical districts in Wiesbaden.
Source: Figure created using R (2013).

We define this peer effect on decision-maker n in period t as

$$\mathrm{IBASE}_{nt} = \sum_{\substack{m \in N \\ m \neq n | \\ d_{nm} \leq D}} o_{m,t-1,\mathrm{PV}} f(d_{nm}), \tag{7}$$

the installed base (Bollinger and Gillingham 2012). With $d_{nm} > 0$ as the Euclidean distance in meters between the location of n and the location of m. D is a cut-off parameter to be set by the analyst. We may assume that there is no remarkable influence of PV installations farther away from location n than D. For simplification, we choose $f(d_{nm}) = 1/d_{nm}$, and $D = 1000$ m. Further, we assume that a decision-maker is most likely influenced by users of PV systems that installed in the previous period: $t - 1$. Since the incentive system for PV was changed during our period of study, only these peers may pass over reliable information regarding, e.g. the reliability and the net present value of PV at present. Obviously, Equation (7) enables us to test our hypothesis that preexisting PV systems stimulate further installations nearby.

We consider periods defined according to shifts in the annex of PV shown in Figure 4 caused by changes in the legal incentive system. The feed-in tariff, based on the Renewable Energy Sources Act, was put in place in 2000 and made installing PV systems interesting to decision-makers. We define the years before 2000 as our first period. During this period, 13 PV systems were installed in Wiesbaden. Due to the 100,000 roofs program – an additional incentives program which offered subsidized interest rates (e.g. see Jacobsson and Lauber 2006) – the annex of PV installations peaked in 2000 and 2001. These two years form our second period, in which 60 systems were installed. In the following years, when

Figure 3. Exemplary detailed view of PV systems, buildings and statistical districts.
Source: Figure created using R (2013).

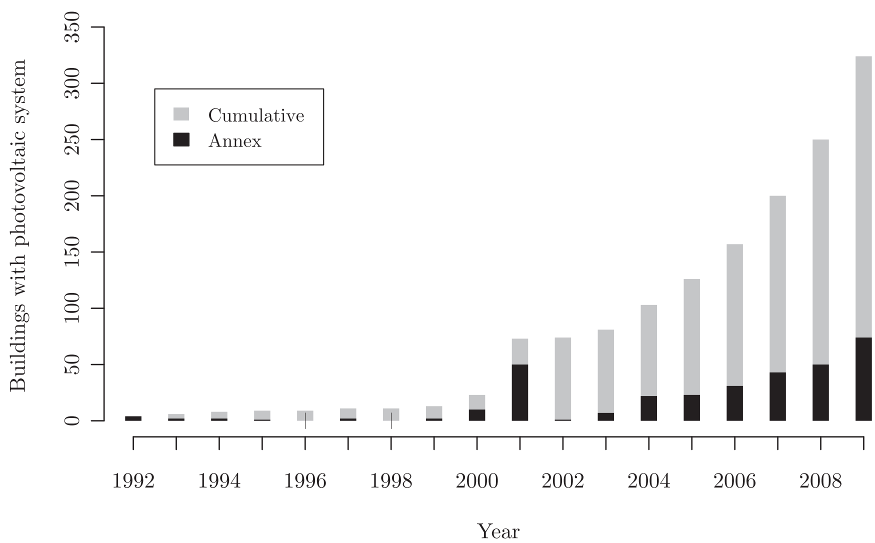


Figure 4. Analyzed cumulative and new PV systems in Wiesbaden per year.
Source: Figure created using R (2013).

Table 1. Frequency of choice alternatives.

Period t	Years	Alternative	Frequency
1	Before 2000	0: No PV system installed	41,653
		1: PV system installed	13
2	2000–2001	0: No PV system installed	41,606
		1: PV system installed	60
3	2002–2005	0: No PV system installed	41,613
		1: PV system installed	53
4	2005–2009	0: No PV system installed	41,468
		1: PV system installed	198

the 100,000 roofs program expired and only the feed-in tariff persisted, the yearly annex of PV installations increased (without large peaks), starting, however, from a smaller yearly level than during the peak in 2000 and 2001. For this reason we set our third period of study to cover 2002 through 2005, when 53 systems were installed. Finally, we define our last period of study to include all installations between 2006 and 2009 since we observe a strong but comparatively balanced annex then: 198 new installations. Table 1 shows the corresponding frequencies.

Since the feed-in tariff for electricity from PV was adapted yearly between 2000 and 2009 and the costs of installing PV decreased during this period, the profitability of PV systems changed over the course of time.[7] Therefore, we allow for different gains in utility from different installation periods by considering temporal fixed effects for each period of analyses.

Of course, the utility a decision-maker gains by installing in a certain period or not may also be influenced by factors other than peer effects (IBASE_{nt}) and temporal fixed effects covering shifts in the profitability of PV systems. Unfortunately, we do not have information about the characteristics of the decision-maker itself but we have information on characteristics of the decision-makers' location. We assume that it is very likely that the characteristics of both are comparable and therefore use locational characteristics as control variables. We obtain data from 177 statistical districts in Wiesbaden in the year 2009 (Infas 2009). The average area of a statistical district is about 1.13 sqkm. We consider the following variables:

- BUYPOW_n is the buying power index of the statistical district according to decision-maker n's location. An index value of 0.1 corresponds to the median buying power of German households in 2009. Since PV installations are expensive, one might assume that wealthier decision-makers are more likely to install. Furthermore, according to Rogers (1983) early adopters may have a higher socio-economic status than followers, which leads us to the hypothesis that measures indicating high income are associated with early adopters.
- POPDEN_n denotes the population density times 10 of the statistical district where n is located in 2009. A low measure of POPDEN_n may refer to places with a high share of single- and double-family homes. For decision-makers located at these places the decision to install a PV system may be easier as fewer parties have to agree upon the installation on a certain building. As a consequence, we expect a negative impact of POPDEN_n on the propensity to install PV. In line with Rogers (1983), early adopters may have a higher socio-economic status than later adopters. Therefore,

Table 2. Descriptive statistics.

Variable name	Min.	Mean	Max.	Expected sign	Expected change over time
$IBASE_{n2}$	0	0.0013	0.2025	+	+
$IBASE_{n3}$	0	0.0055	0.2324	+	−
$IBASE_{n4}$	0	0.0045	0.2353	+	+/−
$BUYPOW_n$	0.0936	0.1238	0.1602	+	−
$POPDEN_n$	0.0003	0.0485	0.3204	−	+

we test whether early adopters are located in districts with a low population density, referring to locations with a high share of single-and double-family homes.

The descriptive statistics for the variables, the expected sign of the corresponding coefficients and their expected change through time are shown in Table 2. In order to cover time-invariant spatial effects, we also use spatial fixed effects according to the land use classification of CORINE Land Cover (CLC 2009).[8]

According to Section 2, we normalize the deterministic part of the utility from an observation n to choose alternative $i = \text{no PV}$ in period t as

$$V_{nt,\text{no PV}} = 0, \tag{8}$$

while the deterministic part of the utility from an observation n to choose alternative $i = \text{PV}$ in period t is measured by

$$\begin{aligned} V_{nt,\text{PV}} = \beta_0 &+ \beta_1 \text{IBASE}_{n2} + \beta_2 \text{IBASE}_{n3} + \beta_3 \text{IBASE}_{n4} + \beta_4 \text{BUYPOW}_n \text{PER}_2 \\ &+ \beta_5 \text{BUYPOW}_n \text{PER}_3 + \beta_6 \text{BUYPOW}_n \text{PER}_4 + \beta_7 \text{POPDEN}_n \text{PER}_2 \\ &+ \beta_8 \text{POPDEN}_n \text{PER}_3 + \beta_9 \text{POPDEN}_n \text{PER}_4 + \Sigma_{h=8}^{H} \beta_h W_{nth}, \end{aligned} \tag{9}$$

with β_h to be estimated by maximum likelihood of Equation (6) and IBASE_{nt} given as Equation (7). PER_t is a dummy variable that equals 1 for period t and 0 otherwise. Spatial and period fixed effects are denoted by dummy variables W_{nth}.[9]

4. Results and discussion

Table 3 displays the coefficient estimates of the BPL of Equation (5) using the utility specification of Equation (9) in Section 3.2. Specification M2, M4 and M6 include spatial fixed effects. In general they do not largely differ from M1, M3 and M5. However, likelihood ratio tests confirm that M2, M4 and M6 are significantly superior models.

First of all, the installed base (IBASE_{nt}) has a significantly ($p < .01$) positive influence on the decision to install PV in all the specifications. That is, the more proximate PV systems in the preceding period, the higher the propensity of a decision-maker to obtain a PV system in the current period.[10] Since every building is owned by someone, and these persons might be influenced by the decisions of their peers, we find imitation of spatially close precursors is indeed an explaining factor in PV adoption; i.e. our results confirm a localized peer effect in the adoption of PV. The utility to install a PV system in the current period increases by 15.8 units (see M2) per installed PV system in a previous period relative to the distance to the previous installations. Imagine a given owner of a building and the situation that in a previous period only one PV system has been installed on a building located 100 m away. Then, according to Equation (7) the increase in utility is $15.8/100 = 0.158$.

Table 3. Coefficient estimates of utility functions.

Number	Model Variable Description	M1 $\hat{\beta}$ (t-Stat)	M2 $\hat{\beta}$ (t-Stat)	M3 $\hat{\beta}$ (t-Stat)	M4 $\hat{\beta}$ (t-Stat)	M5 $\hat{\beta}$ (t-Stat)	M6 $\hat{\beta}$ (t-Stat)
0	Intercept for PV	−3.54 (−6.52)		−3.53 (−6.49)		−2.90 (−4.32)	
1,2,3	IBASE_{nt} $\beta_1 = \beta_2 = \beta_3$	16.0 (5.32)	15.8 (5.26)				
1	IBASE_{n2}			19.8 (3.82)	20.9 (4.39)	18.3 (3.50)	19.3 (3.93)
2	IBASE_{n3}			16.7 (2.75)	17.3 (3.28)	15.3 (2.35)	16.0 (2.67)
3	IBASE_{n4}			15.1 (4.05)	14.1 (3.58)	16.6 (4.55)	15.9 (4.30)
4,5,6	BUYPOW_n $\beta_4 = \beta_5 = \beta_6$	−12.8 (−2.98)	−23.7 (−11.26)	−12.8 (−2.99)	−23.6 (−11.24)		
4	$\text{BUYPOW}_n\text{PER}_2$					5.05 (0.52)	6.92 (0.69)
5	$\text{BUYPOW}_n\text{PER}_3$					−17.7 (−1.75)	−16.3 (−1.61)
6	$\text{BUYPOW}_n\text{PER}_4$					−16.5 (−3.10)	−27.0 (−10.27)
7,8,9	POPDEN_n $\beta_7 = \beta_8 = \beta_9$	−7.71 (−5.30)	−6.12 (−4.25)	−7.71 (−5.31)	−6.12 (−4.26)		
7	$\text{POPDEN}_n\text{PER}_2$					0.183 (0.08)	2.46 (1.04)
8	$\text{POPDEN}_n\text{PER}_3$					−2.53 (−1.24)	−0.259 (−0.12)
9	$\text{POPDEN}_n\text{PER}_4$					−13.8 (−5.77)	−11.8 (−5.34)
Period dummies		Yes	Yes	Yes	Yes	Yes	Yes
CLC dummies		No	Yes	No	Yes	No	Yes
Final log-likelihood $\mathcal{L}$		−2093.3	−2068.3	−2093.3	−2068.1	−2082.4	−2054.0
$-2[\mathcal{L}(M1) - \mathcal{L}(Ml)], l = 2, \dots, 6$			50.0	0.0	50.4	21.8	78.6
$-2[\mathcal{L}(M2) - \mathcal{L}(Ml)], l = 3, \dots, 6$				−50.0	0.4	−28.2	28.6
$-2[\mathcal{L}(M3) - \mathcal{L}(Ml)], l = 4, 5, 6$					50.4	21.8	78.6
$-2[\mathcal{L}(M4) - \mathcal{L}(Ml)], l = 5, 6$						−28.6	28.2
$-2[\mathcal{L}(M5) - \mathcal{L}(M6)]$							56.8

Notes: We employ Biogeme 2.2 for estimation. See Bierlaire (2003). N = 124,998.

When comparing M2 and M4 one observes that allowing IBASE_{nt}'s influence to vary over time only slightly changes the estimates' magnitudes. We hypothesized that the importance of peer effects may at first increase and then decrease over time (in the early phase of the diffusion path). The estimates for M4 show that β_1 is larger than β_2 which in turn is larger than β_3. However, the evidence is weak since a likelihood ratio test does not confirm M4's superiority over M2. Furthermore, simply comparing the magnitude of the estimated coefficients yields limited information. Instead, we study the general impact of the installed base. We consider the aggregate direct elasticities for each period. Generally, the direct elasticity measures the percentage change in the choice probability of a given alternative, with respect to a percentage change in a variable of that alternative.[11] In our case, the direct elasticity helps illustrate the change in choice probabilities of the whole population due to a percentage change of the installed base. Table 4 shows the respective results.

Obviously, the choice probabilities are inelastic but positive to the peer effect (IBASE_{nt}). This result confirms the effect of installation seeds in fostering PV adoption. For our second

Table 4. Aggregate direct elasticities.

Variable	$\mathcal{E}$
IBASE_{n2}	0.04
IBASE_{n3}	0.12
IBASE_{n4}	0.09
$\text{BUYPOW}_n\text{PER}_2$	0.86
$\text{BUYPOW}_n\text{PER}_3$	−1.97
$\text{BUYPOW}_n\text{PER}_4$	−3.23
$\text{POPDEN}_n\text{PER}_2$	0.12
$\text{POPDEN}_n\text{PER}_3$	−0.01
$\text{POPDEN}_n\text{PER}_4$	−0.31

period, the elasticity to the peer effect is smallest, for period 3 it shows the highest value and in period 4 it takes a value in the middle, i.e. its change over time is at first increasing and then decreasing as expected.

Since we find an inelastic elasticity over all periods it may be advisable to facilitate the peer effect. Cardwell (2012) documents marketing strategies fostering PV adoption: having in mind the relevance of peer effects, the PV industry encourages enthusiastic customers to inform their neighbors about the benefits of PV (during gatherings similar to Tupperware parties). Being informed about the reliability of PV and the low complexity of installing by an acquaintance may be more convincing than the arguments of a salesperson. In general, it may be rewarding to create incentives for PV users to share their experience with the technology. Besides industry, such rewards for recommending PV may also be implemented by the state. It is conceivable that such rewards could – as installation seeds – be used to steer the location of PV systems: e.g. a grid-oriented location plan may help to overcome bottlenecks in grid capacity, as observed in Germany.

In contrast to our expectation, M2 shows that the impact of the buying power (BUYPOW_n) on the utility is significantly negative. The utility to install PV decreases for a decision-maker with an average buying power (i.e., $\text{BUYPOW}_n = 0.1$) by -2.37 units. However, allowing BUYPOW_n's estimate to vary over time (see M6) reveals that its impact is only significantly negative in period 4. This finding is in line with our hypothesis and indicates that early adopters have a higher socio-economic status than followers.

The impact of the population density (POPDEN_n) is negative as expected (see M2), indicating that decision-makers located in less densely populated areas are more likely to install a PV system in a certain period. This finding is reasonable as the propensity is high that decision-makers located in areas with low population density own a house. In these areas we expect a high share of single- and double-family homes. Again, comparing M6 and M2 illustrates that POPDEN_n's impact on the utility is insignificant in periods 2 and 3 and only significantly negative in period 4: i.e. we observe a decreasing trend in the impact of POPDEN_n on utility. This result implies that early decisions to adopt are less influenced by population density than later decisions. We might conclude that it is more likely for early installations to be found in areas with higher population density than for later installations. However, this finding is not in line with our hypothesis that a low population density characterizes early adopters.

We consider M6 to be the least worse model because it provides the best model fit (i.e. the highest final log-likelihood) and the specification (mainly) corresponds to our expectations (see Section 3.2 and below). Comparing M2 and M6 confirms that the coefficient estimates are robust for IBASE and that the actual impact of BUYPOW_n and POPDEN_n is only revealed if allowed to be time-variant.

Table 4 also reveals that the elasticity to BUYPOW_n and POPDEN_n varies over time: decision-makers located in low-income districts are more likely to be later adopters. Possibly, later adopters are more interested in the earnings they may gain from PV than early adopters: i.e. early adopters might be driven by a more intrinsic motivation than later adopters. The same coherence can be observed for population density: later adopters are more likely to be located in districts outside the core city region. We conclude that installation seeds may be more effective in fostering PV adoption for later adopters (in the early phases of the diffusion path) in less densely populated areas with a high income level.

The robustness of our findings is tested by employing 25 urban district dummies instead of CLC dummies. The urban district dummies also cover time-invariant spatial effects but according to a different spatial structure. See Table 5, M7 for details. As in M6 the coefficients for IBASE_{nt} are positive for all three periods of analysis. However, the estimate for period 4 is insignificant. This result corresponds with our previous finding that the importance of peer effects initially increases and subsequently decreases (revealed by the aggregate direct elasticities for M7 which are for clarity reasons not shown). The estimates of BUYPOW_n are positive in M7 but only for period 2 significant at the 5% level. In M6 the estimated coefficient for BUYPOW_n is significantly negative in period 4. We are therefore unsure about the influence of BUYPOW_n on the decision to adopt. Note that the influence of buying power on the decision to adopt may be covered by the district dummies and the estimated coefficient may be biased in M7. Similar to M6, the estimated coefficient for POPDEN_n is only significant (and negative) for period 4. Thus, our previous result is confirmed: in less densely populated areas within Wiesbaden – where a high share of single- and double-family homes may be situated – the propensity to adopt PV is comparatively high for later adopters.

We also confirm our results for private installations only, since the decision process of firms may be different from that of private homeowners. The results are shown in Table 5, M8. We assume that all PV systems under a nominal power of 10 kilowatt-peak are private installations. Under this assumption 240 installations are left: 13 of these were installed through 1999 (period 1), 58 in 2000 or 2001 (period 2), 44 between 2002 and 2005 (period 3) and 125 between 2006 and 2009 (period 4). In general, our findings are verified. The estimation results are close to those shown in M6 of Table 3 but the coefficient for IBASE_{nt} in period 4 is insignificant. This result is in line with our finding that the importance of peer effects increases initially and subsequently decreases (in the early phase of technology diffusion).

We tried other spatial measures than the simple non-linear specification: $f(d_{nm}) = 1/d_{nm}$. When employing $1/\log(d_{nm})$, or $(d_{nm})^{-2}$, we obtain inferior log-likelihood values. To maintain clarity these results are not shown.

Finally, we test the robustness of our findings by splitting up the data set into cross sections. This procedure confirms our findings. Again, for reasons of clarity we do not show the estimates.

For verification purposes, we compare our results and findings with selected studies that employ different methods. We are able to confirm the findings of Bollinger and Gillingham (2012), who analyze the adoption of PV in California, concerning the negative impact of population density on the propensity to install a PV system. Further, their study reveals that the median income has a negative impact on adoption. Here, we find a comparable effect, but only for later adopters. Most importantly, Bollinger and Gillingham (2012) find a positive effect of proximate PV systems previously installed (the installed base), indicating a significant peer effect. They consider a constant effect of all previous installations within a certain area (street-level and zip code), yielding a decrease in the effect toward the larger

Table 5. Coefficient estimates of utility functions.

Number	Model Variable Description	M7 $\hat{\beta}$ (t-stat)	M8 $\hat{\beta}$ (t-stat)
1	$IBASE_{n2}$	17.2 (3.19)	16.4 (2.64)
2	$IBASE_{n3}$	16.1 (2.25)	18.0 (2.39)
3	$IBASE_{n4}$	5.94 (0.77)	−2.36 (−0.24)
4	$BUYPOW_nPER_2$	26.2 (2.37)	6.30 (0.63)
5	$BUYPOW_nPER_3$	1.59 (0.14)	−18.1 (−1.97)
6	$BUYPOW_nPER_4$	3.04 (0.40)	−32.4 (−16.96)
7	$POPDEN_nPER_2$	2.38 (0.81)	1.74 (0.69)
8	$POPDEN_nPER_3$	−0.523 (−0.19)	0.474 (0.23)
9	$POPDEN_nPER_4$	−12.9 (−4.61)	−10.3 (−4.68)
Period dummies		Yes	Yes
District dummies		Yes	No
CLC dummies		No	Yes
Final log-likelihood $\mathcal{L}$		−2054.4	−1623.0
$-2[\mathcal{L}(M6) - \mathcal{L}(Ml)], l = 7, 8$		−0.8	862.0

Notes: We employ Biogeme 2.2 for estimation. See Bierlaire (2003).

area (zip code). In contrast, we consider a continuous measure that leads to a more precise proxy of peer effects and we witness the same overall effect: a decrease of the peer effect with distance.

Rode and Weber (2012) employ an epidemic diffusion model to analyze the diffusion of PV in Germany between 1992 and 2009. They add a spatial component to the epidemic model, control for changes in the incentive system and also find an attenuating peer effect with distance. The peer effect can only be significantly identified up to a range of 1.2 km, which is in line with our cut-off distance of 1 km. However, Rode and Weber (2012) find a positive influence of population density, which may be caused by the less detailed population data they employ (i.e. they use more aggregated data).

Welsch and Kühling's (2009) analysis of solar thermal equipment – a closely related technology to PV – also supports the importance of peer effects and therefore our findings. Their study builds on individual level data from a survey in Hanover, Germany, and reveals that the behavior of reference groups is of major importance for the adoption decision.

5. Summary and conclusion

We set out to study the adoption decision to install a PV system in Wiesbaden. In contrast to previous studies on PV adoption, we analyze individual decision level data and determine a measure of peer effects for each decision-maker individually. In consequence, our results should be less influenced by a measurement bias due to data aggregation. Based on the year and location of the PV systems set up in Wiesbaden through 2009 and locational data of buildings (on which we assume all PV systems to be installed), we employ a binary panel choice model for our study. In line with Bollinger and Gillingham (2012) and Rode and Weber (2012), our analysis reveals that the propensity to install PV increases with the number of previously installed systems in spatial proximity. Our proxy for the installed base is built on a continuous non-linear measure of distance in order to account for peer effects. The impact of each previously installed PV system on the propensity of a new installation decreases continuously with the distance between them. Although the impact is significant, it is inelastic. We find a decreasing elasticity in the propensity of a new installation on the expected income level for later adopters. For more recent periods the adoption is more elastic than for earlier periods. We control our results for spatial and temporal fixed effects.

From the viewpoint of a political decision-maker our results reveal that it may be more efficient to influence the adoption of PV systems by installation seeds in less densely populated areas with high income levels at early (but not the first) periods of the diffusion phase. Installation seeds may also be used to steer PV to certain regions.

Although our case study provides valuable insights into factors influencing the decision to install PV, a larger unit of study than the city of Wiesbaden (with a higher number than 324 PV systems installed within almost 20 years) may be worthwhile: besides more detailed results, e.g. allowing for yearly changes in some estimates, it may be interesting to account for further controls. In addition, it may be rewarding to conduct similar studies for other German cities and to test if city specific factors influence the decision to adopt PV (see Löw 2012). Certainly, investigating if our results hold true in other institutional settings, i.e. other countries (e.g. see Lüthi and Wüstenhagen 2012), would also be a natural step. Further, more detailed data on the peer effect would allow us to focus on the network structure or the type of network intervention (e.g. see Valente 2012). Studying the influence of marketing efforts as, e.g. suggested by Kalish (1985) or Delre, Jager, and Janssen (2007) may also be of interest.

Acknowledgements

Johannes Rode gratefully acknowledges funding provided by the Graduate School for Urban Studies (URBANgrad) at Technische Universität Darmstadt (kindly supported by LOEWE-Schwerpunkt “Eigenlogik der Städte”). We greatly appreciate that the base map of Wiesbaden was provided by Wiesbaden’s land surveying office. Our sincere gratitude goes to the developers of Biogeme, PostGIS, PostgreSQL, and R. Thanks a lot for making great software available for free. Further, we are very grateful for the helpful comments of two anonymous referees and the editorial board. Finally, we highly appreciate proofreading by Philip Savage.

Notes

1. Using a probit approach instead would yield high-dimensional integrals for the corresponding choice probabilities. Because our sample comprises more than 40,000 observations, a probit model is not advisable.
2. At the end of our period of study, still less than 1% of the potential users had adopted PV. Thus, when referring to *early adopters*, we mean those who adopt in the very first periods of diffusion. In contrast, we define *later adopters* as those who adopt in early but not in the very first periods of diffusion.
3. The remuneration is financed through an apportionment by all consumers of electricity; thus the costs are born by all consumers.
4. The data on PV systems in Wiesbaden is extracted from Rode and Weber’s (2012) data set. Primary data comes from the German transmission system operators, which are by law obliged to publish address data and the date of grid connection for all the PV systems being financed through the feed-in tariff since the amendment of the Renewable Energy Sources Act on October 25, 2008.
5. Dewald and Truffer (2011) confirm the dominance of roof-top installations in Germany.

6. Since disparities between the spatial data used for the geocoding of the PV systems and the spatial data of the buildings exist, it is not surprising that some PV systems do not fall inside a building-polygon. Still, 100 out of the 149 PV systems are allocated to a building within less than 10 m distance, illustrating the high quality of the data set. The distance calculations are based on a spherical model of the earth. As all the geographical calculations, they are conducted using PostGIS 1.5.
7. See Altrock, Oschmann, and Theobald (2008) for more details on the level of the feed-in tariff per year, capacity and type of installation.
8. Our reference category consists of road and rail networks and associated land (CLC code: 122), airports (124), mineral extraction sites (131), land principally occupied by agriculture, with areas of natural vegetation (243), coniferous forest (312), mixed forest (313), natural grasslands (312), and stream courses (511). We set spatial dummies for continuous urban fabric (111), discontinuous urban fabric (112), industrial or commercial units (121), dump (132), green urban areas (141), sport and leisure facilities (142), non-irrigated arable land (211), vineyards (221), fruit trees and berry plantations (222), pastures (231), complex cultivation (242), and broad-leaved forest (311). The minimum mapping unit for the polygons of the CORINE Land Cover data set is 25 m.
9. Note that potential coefficients for period 1 have to be fixed to zero for identification purposes.
10. According to our data set, less than 1% of potential adopters install PV. We therefore argue that the diffusion of PV in Wiesbaden is still to be found in the first half of the diffusion curve, which implies that saturation has not yet been reached. In this early phase of diffusion, an increasing likelihood of adoption with increasing numbers of pre-installed systems is also expected by diffusion theory (Griliches 1957; Mansfield 1961; Bass 1969, 1980; Stoneman, forthcoming). Having a look at Figure 4 (while neglecting the outlier in 2001) underpins this view.
11. In other words, the direct elasticity measures the responsiveness of the choice probability of an alternative to a change in an attribute (variable) of the same alternative. The aggregate direct elasticity is simply a weighted average of the individual level elasticities using the choice probabilities as weights over all decision-makers (i.e. $n \in N$). We refer to Ben-Akiva and Lerman (1985) for details on elasticities derived from discrete choice models.

References

Altrock, M., V. Oschmann, and C. Theobald. 2008. *EEG: Erneuerbare-Energien-Gesetz; Kommentar*. 2nd ed. Munich: Beck.

Bala, Venkatesh, and Sanjeev Goyal. 1998. “Learning from Neighbours.” *The Review of Economic Studies* 65 (3): 595–621. doi:10.1111/1467-937X.00059.

Bass, Frank M. 1969. “A New Product Growth Model for Consumer Durables.” *Management Science* 15 (5): 215–227. http://www.jstor.org/stable/2628128

Bass, Frank M. 1980. “The Relationship Between Diffusion Rates, Experience Curves, and Demand Elasticities for Consumer Durable Technological Innovations.” *The Journal of Business* 53 (3): S51–S67. http://www.jstor.org/stable/2352211

BBSR (Bundesinstitut für Bau-, Stadt- und Raumforschung). 2010. “Stadt-Land-Regionen: Großstadtregionen 2010.” Bonn. Database, accessed November 14, 2012. http://www.bbsr.bund.de/cln_032/nn_1051708/BBSR/DE/Raumbeobachtung/Downloads/downloads__node.html

Ben-Akiva, M., and S. R. Lerman. 1985. *Discrete Choice Analysis, Theory and Applications to Travel Demand*. Cambridge, MA: MIT Press.

Bernard, Jean-Thomas, Denis Bolduc, and Donald Bélanger. 1996. “Quebec Residential Electricity Demand: A Microeconometric Approach.” *The Canadian Journal of Economics* 29 (1): 92–113. http://www.jstor.org/stable/136153

Bierlaire, Michel. 2003. “BIOGEME: A Free Package for the Estimation of Discrete Choice Models.” Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland. Accessed May 2, 2013. http://www.strc.ch/conferences/2003/bierlaire.pdf

BMU (Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit). 2011. *Vorbereitung und Begleitung der Erstellung des Erfahrungsberichtes 2011 gemäß § 65 EEG – Vorhaben II c – Solare Strahlungsenergie – Endbericht. Im Auftrag des BMU, Projektleitung: Matthias Reichmuth*. Berlin: Leipziger Institut für Energie GmbH. Accessed April 10, 2013. http://www.erneuerbare-energien.de/fileadmin/ee-import/files/pdfs/allgemein/application/pdf/eeg_eb_2011_solare_strahlung_bf.pdf

Bollinger, Bryan, and Kenneth Gillingham. 2012. "Environmental Preferences and Peer Effects in the Diffusion of Solar Photovoltaic Panels." *Marketing Science* 31 (6): 900–912. doi:10.1287/mksc.1120.0727.

Bonus, Holger. 1973. "Quasi-Engel Curves, Diffusion, and the Ownership of Major Consumer Durables." *Journal of Political Economy* 81 (3): 655–677. http://www.jstor.org/stable/1831030

Cardwell, Diane. 2012. "Solar Industry Borrows a Page, and a Party, From Tupperware." *New York Times*, December 1, B1. http://www.nytimes.com/2012/12/01/business/energy-environment/solar-industry-borrows-a-page-and-a-party-from-tupperware.html

CLC (Coordinated Information on the European Environment (CORINE) Land Cover). 2009. "CORINE Land Cover from the year 2006, Umweltbundesamt, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Deutsches Fernerkundungsdatenzentrum (DFD)." Database, accessed June 13, 2012, Weßling. http://www.corine.dfd.dlr.de/datadescription_2006_de.html

Delre, Sebastiano A., Wander Jager, and Marco A. Janssen. 2007. "Diffusion Dynamics in Small-World Networks with Heterogeneous Consumers." *Computational and Mathematical Organization Theory* 13 (2): 185–202. doi:10.1007/s10588-006-9007-2.

Dewald, Ulrich, and Bernhard Truffer. 2011. "Market Formation in Technological Innovation Systems – Diffusion of Photovoltaic Applications in Germany." *Industry & Innovation* 18 (3): 285–300. doi:10.1080/13662716.2011.561028.

DWD (Deutscher Wetterdienst). 2010. "1-km-Rasterdaten der mittleren jährlichen Globalstrahlung (kWh/qm, Zeitraum 1981–2000) der gesamten Bundesrepublik Deutschland." Data provided after personal request, July 26, 2010, Offenbach. http://www.dwd.de/

Ellison, Glenn, and Drew Fudenberg. 1993. "Rules of Thumb for Social Learning." *Journal of Political Economy* 101 (4): 612–643. http://www.jstor.org/stable/2138741

Farrell, Joseph, and Garth Saloner. 1986. "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation." *American Economic Review* 76 (5): 940–955. http://www.jstor.org/stable/1816461

Geroski, P. A. 2000. "Models of Technology Diffusion." *Research Policy* 29 (4–5): 603–625. doi:10.1016/S0048-7333(99)00092-X.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25 (4): 501–522. http://www.jstor.org/stable/1905380

Hensher, D. A., and W. H. Greene. 2003. "The Mixed Logit model: The State of Practice." *Transportation* 30 (2): 133–176. http://www.ingentaconnect.com/content/klu/port/2003/00000030/00000002/05106267

Infas (Infas Geodaten GmbH). 2009. "Wohnquartiere, Kreis-Gemeinde-Schlüssel (KGS) 22, 2009." Purchased data, provided after personal request, December, 2010, Bonn. http://www.infas-geodaten.de/

Jacobsson, S., and V. Lauber. 2006. "The Politics and Policy of Energy System Transformation – Explaining the Diffusion of Renewable Energy Technology." *Energy Policy* 34 (3): 256–276. doi:10.1016/j.enpol.2004.08.029.

Kalish, Shlomo. 1985. "A New Product Adoption Model with Price, Advertising, and Uncertainty." *Management Science* 31 (12): 1569–1585. doi:10.1287/mnsc.31.12.1569.

Karshenas, Massou D., and Paul Stoneman. 1992. "A Flexible Model of Technological Diffusion Incorporating Economic Factors with an Application to the Spread of Colour Television Ownership in the UK." *Journal of Forecasting* 11 (7): 577–601. doi:10.1002/for.3980110702.

Kemp, René. 1998. "The Diffusion of Biological Waste-Water Treatment Plants in the Dutch Food and Beverage Industry." *Environmental and Resource Economics* 12 (1): 113–136. doi:10.1023/A:1016078930151.

Koski, Heli A. 1999. "The Installed Base Effect: Some Empirical Evidence from the Microcomputer Market." *Economics of Innovation and New Technology* 8 (4): 273–310. doi:10.1080/10438599900000012.

Löw, Martina. 2012. "The Intrinsic Logic of Cities: Towards a New Theory on Urbanism." *Urban Research & Practice* 5 (3): 303–315. doi:10.1080/17535069.2012.727545.

Lüthi, Sonja, and Rolf Wüstenhagen. 2012. "The Price of Policy Risk – Empirical Insights from Choice Experiments with European Photovoltaic Project Developers." *Energy Economics* 34 (4): 1001–1011. doi:10.1016/j.eneco.2011.08.007.

Majumdar, Sumit K. 1996. "Bandwagon Influences and Installed-Base Conversion in U.S. Telecommunications." *Economics of Innovation and New Technology* 4 (2): 113–122. doi:10.1080/10438599600000003.

Mansfield, Edwin. 1961. “Technical Change and the Rate of Imitation.” *Econometrica* 29 (4): 741–766. http://www.jstor.org/stable/1911817

McFadden, D. 1986. “The Choice Theory Approach to Market Research.” *Marketing Science* 5 (4): 275–297. doi:10.1287/mksc.5.4.275.

McFadden, D. 2001. “Economic Choices.” *American Economic Review* 91 (3): 351–378. doi:10.1257/aer.91.3.351.

Nesbakken, Runa. 2001. “Energy Consumption for Space Heating: A Discrete-Continuous Approach.” *Scandinavian Journal of Economics* 103 (1): 165–184. doi:10.1111/1467-9442.00236.

R (R Development Core Team). 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Accessed May 2, 2013. http://www.R-project.org/

Rode, Johannes, and Alexander Weber. 2012. “Does Localized Imitation Drive Technology Diffusion? A Case Study on Solar Cells in Germany.” Working paper. Accessed September 3, 2012. http://www.stadtforschung.tu-darmstadt.de/media/loewe_eigenlogik_der_staedte/dokumente_download/artikel/Rode_Weber_2012.pdf

Rogers, Everett M. 1983. *Diffusion of Innovations*. 3rd ed. New York: Free Press.

Stoneman, Paul. Forthcoming. “The Impact of Prior Use on the Further Diffusion of New Process Technology.” *Economics of Innovation and New Technology*. doi:10.1080/10438599.2012.708133.

Tobler, W. 1970. “A Computer Movie Simulating Urban Growth in the Detroit Region.” *Economic Geography* 46 (June): 234–240. http://www.jstor.org/stable/143141

Valente, Thomas W. 1996. “Social Network Thresholds in the Diffusion of Innovations.” *Social Networks* 18 (1): 69–89. doi:10.1016/0378-8733(95)00256-1.

Valente, Thomas W. 2012. “Network Interventions.” *Science* 337 (6090): 49–53. doi:10.1126/science.1217330.

Welsch, Heinz, and Jan Kühling. 2009. “Determinants of Pro-Environmental Consumption: The Role of Reference Groups and Routine Behavior.” *Ecological Economics* 69 (1): 166–176. doi:10.1016/j.ecolecon.2009.08.009.

Wiesbaden (Landeshauptstadt Wiesbaden, Vermessungsamt). 2009. “Stadtgrundkarte Wiesbaden, Hauptgebäude und Nebengebäude (Stand: März 2009).” Data was provided for the research unit LOEWE ‘Eigenlogik der Städte’. Accessed December 14, 2011. http://www.wiesbaden.de/vv/produkte/66/141010100000008503.php

Wilson, Charlie, and Hadi Dowlatabadi. 2007. “Models of Decision Making and Residential Energy Use.” *The Annual Review of Environment and Resources* 32 (November): 169–203. doi:10.1146/annurev.energy.32.053006.141137.

Young, H. Peyton. 2009. “Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning.” *American Economic Review* 99 (5): 1899–1924. doi:10.1257/aer.99.5.1899.

2.2 Spatial Management and Planning

2.2.1 Location Planning

1. Haase, K., Müller, S. (2014): **A comparison of linear reformulations for multinomial logit choice probabilities in facility location models**. *European Journal of Operational Research* **232**(3), 689-691.

2. Haase, K., Müller, S. (2014): **Insights in clients' choice in preventive health care facility location planning**. *OR Spectrum*, forthcoming.

3. Müller, S., Haase, K. (2014): **Customer segmentation in retail facility location planning**. *Business Research*, in print.

4. Haase, K., Müller, S. (2013): **Management of school locations allowing for free school choice**. *Omega - The International Journal of Management Science* **41**(5), 847-855.

European Journal of Operational Research 232 (2014) 689–691

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Short Communication

A comparison of linear reformulations for multinomial logit choice probabilities in facility location models

Knut Haase [1], Sven Müller *

Institute for Transport Economics, University of Hamburg, D-20146 Hamburg, Germany

ARTICLE INFO

Article history:
Received 8 April 2013
Accepted 7 August 2013
Available online 15 August 2013

Keywords:
Multinomial logit model
Facility location
Linear reformulations

ABSTRACT

In the last decade several papers appeared on facility location problems that incorporate customer demand by the multinomial logit model. Three linear reformulations of the original non-linear model have been proposed so far. In this paper, we discuss these models in terms of solvability. We present empirical findings based on synthetic data.

1. Introduction

When customer choice behavior is considered in facility location planning we generally assume utility maximizing behavior. Probabilistic demand based on utility maximization implies that only the probability of a customer choosing a given facility is known. Traditionally, in facility location literature, gravity-type demand models are used to consider probabilistic demand (Serra, Eiselt, Laporte, & ReVelle, 1999). More recently, there is a growing body of literature that considers the multinomial logit model (MNL) in facility location models (see for example Aros-Vera, Marianov, & Mitchell, 2013; Benati & Hansen, 2002; Haase, 2009; Haase & Müller, 2013; Müller, Haase, & Kless, 2009; Zhang, Berman, & Verter, 2012). The MNL is a well-known discrete choice model (random utility model) to describe (spatial) customer choice behavior (Ben-Akiva et al., 2002; McFadden, 2001; Müller, Haase, & Seidel, 2012). Using MNL within a mathematical program for locational decision making probably yields a non-linear model formulation (see Benati, 1999; Marianov, Rfos, & Icaza, 2008, for example). If we consider only locational decisions, then there exist three different linear reformulations of the MNL yielding mixed-integer linear programs. Benati and Hansen (2002) where the first who proposed a linear reformulation based on variable substitution. Haase (2009) has proposed to employ the constant substitution pattern of the MNL in order to enable a linear integer formulation (see also Aros-Vera et al., 2013). Finally, Zhang et al. (2012) presented an alternative approach based on variable substitution.

As the three model formulations are discussed independently so far, we compare them in this contribution. Therefore, we first give a brief discussion of the MNL and its incorporation into the maximum capture problem (Section 2). Based on this, we discuss the three linear reformulations using a unified notation in order to make them more comparable (Sections 2.1, 2.2, 2.3). In Section 3 we compare the models in a numerical study using synthetic data. Our conclusions can be found in Section 4.

2. Mathematical formulations

Consider a firm that wishes to enter a market where customers are located in zones denoted by nodes I. Potential facilities of the firm and facilities of competitors are located in nodes M. Now, the problem of the firm is to select r facilities from all potential facilities J with $J \subset M$ such that the patronage of the facilities of the firm is maximized. In order to determine the patronage of a located facility we assume the customers located in $i \in I$ to be homogeneous in their observable characteristics like income and so on. We further assume that the customers located in $i \in I$ maximize their utility

$$u_{ij} = v_{ij} + \epsilon_{ij}, \qquad (1)$$

when choosing a facility located in $j \in M$.[2] v_{ij} is the deterministic part of utility containing measures of distance, cost, and other attributes. ϵ_{ij} is assumed to be a random term that is independent iden-

* Corresponding author. Tel.: +49 40 42838 9021; fax: +49 40 42838 2924.
E-mail addresses: knut.haase@wiso.uni-hamburg.de (K. Haase), sven.mueller@wiso.uni-hamburg.de (S. Müller).
[1] Tel.: +49 40 42838 9021; fax: +49 40 42838 2924.

[2] M might contain a dummy facility denoting a so-called "no-choice" alternative indicating that customers might patronize no facility at all.

0377-2217/$ - see front matter
http://dx.doi.org/10.1016/j.ejor.2013.08.009

tically extreme value distributed. Then, according to McFadden (1973), the probability of customers located in $i \in I$ choosing a facility located in $j \in M$ is given by the MNL

$$p_{ij} = \frac{e^{v_{ij}}}{\sum_{k \in M} e^{v_{ik}}}. \tag{2}$$

Let us further denote the locational decision variable y_j, taking value 1 if a new facility is located in j and 0 otherwise. For simplicity reasons we assume the demand (number of customers, for example) in node i to be 1 for all nodes $i \in I$. Then, the problem of the firm can be formalized as a mixed-integer non-linear program

$$\max F = \sum_{i \in I} \frac{\sum_{j \in J} e^{v_{ij}} y_j}{\sum_{j \in M \setminus J} e^{v_{ij}} + \sum_{j \in J} e^{v_{ij}} y_j}, \tag{3}$$

$$\text{s.t.} \sum_{j \in J} y_j = r, \tag{4}$$

$$y_j \in \{0, 1\} \quad \forall j \in J. \tag{5}$$

It has been discussed by Benati (1999) that (3)–(5) is NP-hard. In the following we present three different, exact linear reformulations to this problem known from literature (Aros-Vera et al., 2013; Benati & Hansen, 2002; Haase, 2009; Zhang et al., 2012). For comparison reasons we use a unified notation. Thus, the three models appear here in a different fashion compared to the original contributions. We provide the corresponding algebraic transformations in the Appendix. For a detailed discussion of the model formulations we refer to the corresponding references.

2.1. Linear reformulation by Benati and Hansen (2002)

Let us consider the parameter

$$\varphi_{ij} = \frac{e^{v_{ij}}}{\sum_{k \in M \setminus J} e^{v_{ik}}}. \tag{6}$$

Of course, this parameter is only defined if $\{M \setminus J\} \neq \emptyset$. This assumption is quite reasonable because customers usually face at least the choice alternative of "non-consumption" (i.e., the no-choice alternative). In all four references $\{M \setminus J\} \neq \emptyset$. Further, let x_{ij} be a non-negative variable. Applying variable substitution, Benati and Hansen (2002) show that the problem (3)–(5) can be reformulated as

$$\max F = \sum_{i \in I} \sum_{j \in J} \varphi_{ij} (y_j - x_{ij}). \tag{7}$$

s.t. (4), (5), and

$$y_j - x_{ij} + \sum_{k \in J} \varphi_{ik} (y_k - x_{ik}) \leqslant 1 \quad \forall i \in I, \ j \in J \tag{8}$$

$$x_{ij} \geqslant 0 \quad \forall i \in I, \ j \in J. \tag{9}$$

2.2. Linear reformulation by Haase (2009)

We further introduce the non-negative variable $\tilde{x}_i$. Then, (3)–(5) can be reformulated as

$$\max F = \sum_{i \in I} \sum_{j \in J} x_{ij}. \tag{10}$$

s.t. (4), (5), (9), and

$$\tilde{x}_i + \sum_{j \in J} x_{ij} \leqslant 1 \quad \forall i \in I, \tag{11}$$

$$x_{ij} - \frac{\varphi_{ij}}{1 + \varphi_{ij}} y_j \leqslant 0 \quad \forall i \in I, \ j \in J, \tag{12}$$

$$x_{ij} - \varphi_{ij} \tilde{x}_i \leqslant 0 \quad \forall i \in I, \ j \in J, \tag{13}$$

$$\tilde{x}_i \geqslant 0 \quad \forall i \in I. \tag{14}$$

The model of Haase (2009) has been analogously presented by Aros-Vera et al. (2013).

2.3. Linear reformulation by Zhang et al. (2012)

Finally, we consider the non-negative variable z_{ijk} and reformulate (3)–(5) as (10) subject to (4), (5), (9), and

$$x_{ij} - \varphi_{ij} y_j + \sum_{k \in J} \varphi_{ik} z_{ijk} = 0 \quad \forall i \in I, \ j \in J, \tag{15}$$

$$z_{ijk} - x_{ij} \leqslant 0 \quad \forall i \in I, \ j, k \in J, \tag{16}$$

$$z_{ijk} - y_k \leqslant 0 \quad \forall i \in I, \ j, k \in J, \tag{17}$$

$$x_{ij} + y_k - z_{ijk} \leqslant 1 \quad \forall i \in I, \ j, k \in J, \tag{18}$$

$$z_{ijk} \geqslant 0 \quad \forall i \in I, \ j, k \in J. \tag{19}$$

3. Numerical investigations

We compare the models using artificially generated data. Therefore, we have implemented the three models in GAMS (23.7) and use CPLEX 12 on a 64-bit Windows Server 2008 with 1 Intel Xeon 2.4 gigahertz processor and 24 gigabyte RAM to solve the problems. The Cartesian coordinates of the nodes I and M are randomly generated by a uniform distribution in the interval $[0, 30]$. We consider the rectangular distances between $i \in I$ and $j \in M$ denoted by d_{ij}. The deterministic part of utility of (1) is defined as $v_{ij} = -0.2 \cdot d_{ij}$. Further, we consider $|\{M \setminus J\}| = 10$ and $r = \lceil \alpha |J| \rceil$ with $0 < \alpha < 1$. Table 1 in the appendix displays the computational results for different problem sizes ($|I| \times |J|$) and α. For each problem set we have computed ten random instances. The maximal computational time is set to one hour. The table reports the

Table 1
Numerical study. The values for F^*, CPU, GAP and DEV are average values over ten randomly generated instances. CPU denotes the time in seconds used by CPLEX to solve the respective problem. We set the maximum computational time to 3600 seconds. GAP denotes the gap reported by CPLEX in %. DEV denotes the deviation of the objective function value from the optimal value in %.

$\|I\|$	$\|J\|$	α	Benati and Hansen (2002) (4), (5), (7)–(9)				Haase (2009) (4), (5), (9), (10)–(14)				Zhang et al. (2012) (4), (5), (9), (10), (15)–(19)			
			F^*	CPU	GAP	DEV	F^*	CPU	GAP	DEV	F^*	CPU	GAP	DEV
200	25	0.2	44.55	10.52	0	0	44.55	8.04	0	0	44.55	1348.36	0	0
		0.4	69.66	267.27	0	0	69.66	64.22	0	0	65.89	3600.00	28.91	5.44
	50	0.2	43.13	153.40	0	0	43.13	67.10	0	0	41.26	3600.00	19.01	4.39
		0.4	68.85	3591.20	11.99	0.05	68.88	443.97	0	0	58.14	3600.00	–	15.72
400	25	0.2	87.12	35.41	0	0	87.12	38.10	0	0	85.71	3323.68	9.79	1.65
		0.4	136.46	976.47	0	0	136.46	213.06	0	0	116.37	3600.00	–	14.62
	50	0.2	84.82	745.41	0	0	84.82	241.86	0	0	67.97	3600.00	–	19.16
		0.4	134.93	3600.00	29.12	0.81	136.01	1548.51	0	0	110.46	3600.00	–	18.48

average values of the objective function value F^*, the computational effort in seconds CPU and the gap reported by CPLEX (denoted by GAP) as well as the deviation from the optimal solution (denoted by DEV) over ten instances for each model. The objective function value of the model proposed by Zhang et al. (2012) is 20% below the optimal value for some instances. In contrast, using the model of Haase (2009) enables CPLEX to find the optimal solution for all instances within 30 min. For the model by Benati and Hansen (2002) CPLEX usually finds the optimal solution within one hour. However, the computational effort is remarkably higher compared to the model of Haase (2009).

4. Conclusion

In the last decade three linear reformulations for MNL-based demand in facility location models have been proposed. All references appear to be independent from each other. In order to compare the model formulations in terms of solvability we first present the models using a unified notation. This is followed by a computational study using artificial data. We find that the approach proposed by Haase (2009) seems to be promising for solving large problems using GAMS/CPLEX. This finding has practical meaning, because real world problems can be solved without using tailored methods. Practitioners are enabled to use state-of-the-art commercial software to solve their problems. This finding has implications for researchers, too. One might be more interested to find tight bounds for the model of Haase (2009) to increase solvability rather than to simply rely on heuristics. Moreover, the reformulations are valid for gravity-type demand models as well.

However, one major issue remains unsolved so far: The MNL (and hence the models outlined in this paper) exhibits constant substitution patterns, i.e., each facility is an equal substitute to every other facility. It is very likely in applications that substitution is not constant. So far, only Haase and Müller (2013) and Müller et al. (2009) have proposed approximate approaches to deal with flexible substitution patterns.

Acknowledgments

The authors gratefully thank an anonymous referee for helpful comments on an earlier version of the paper.

Appendix A

A.1. Algebraic transformations according to the model of Benati and Hansen (2002)

Following the definition of a_{ij} and b_i by Benati and Hansen (2002), $\varphi_{ij} = a_{ij}/b_i$. If we rearrange (7) and (8) we get by substitution the original formulation.

A.2. Algebraic transformations according to the model of Haase (2009)

Obviously, the variable $\tilde{x}_i$ is defined as the cumulative choice probabilities of the competitors (see Eq. (11)), i.e.,

$$\tilde{x}_i = \sum_{m \in M \setminus J} x_{im} \quad \forall i \in I. \tag{20}$$

Therefore, (13) can be written as

$$x_{ij} \leqslant \varphi_{ij} \sum_{m \in M \setminus J} x_{im} \quad \forall i \in I,\ j \in J. \tag{21}$$

Substituting φ_{ij} and rearranging yields

$$x_{ij} \leqslant \sum_{m \in M \setminus J} x_{im} \frac{e^{v_{ij}}}{\sum_{k \in M \setminus J} e^{v_{ik}}} \quad \forall i \in I,\ j \in J. \tag{22}$$

Without loss of generality, we further assume $|\{M \setminus J\}| = 1$. Then, (22) becomes

$$x_{ij} \leqslant \frac{e^{v_{ij}}}{e^{v_{im}}} x_{im} \quad \forall i \in I,\ j \in J, \text{ and } m \text{ as the competitor}. \tag{23}$$

This is equivalent to

$$p_{im} x_{ij} \leqslant p_{ij} x_{im} \quad \forall i \in I,\ j \in J, \text{ and } m \text{ as the competitor}, \tag{24}$$

with p_{ij} given as (2). Obviously, (24) is valid for all $m, j \in M$. Hence,

$$p_{im} x_{ij} \leqslant p_{ij} x_{im} \quad \forall i \in I,\ j, m \in M, \tag{25}$$

given $y_m = 1$. If y_m has to be determined (25) becomes

$$p_{im} x_{ij} \leqslant p_{ij} x_{im} + 1 - y_m \quad \forall i \in I,\ j, m \in M, \tag{26}$$

Since (12) is just a tighter bound on x_{ij} than simply using $x_{ij} \leqslant y_j$, (10)–(14) can be easily transferred to the original notation using (26) instead of (13). Note, (26) is valid, even if $\{M \setminus J\} = \emptyset$.

A.3. Algebraic transformations according to the model of Zhang et al. (2012)

If we multiply (15) by $\sum_{m \in M \setminus J} e^{v_{im}}$ and rearrange we get

$$x_{ij} \sum_{m \in M \setminus J} e^{v_{im}} + \sum_{k \in J} e^{v_{ik}} z_{ijk} = e^{v_{ij}} y_j \quad \forall i \in I,\ j \in J. \tag{27}$$

Since in Zhang et al. (2012) $|\{M \setminus J\}| = 1$ and $v_{im} = 0$ for $m \in M \setminus J$, (27) reduces to

$$x_{ij} + \sum_{k \in J} e^{v_{ik}} z_{ijk} = e^{v_{ij}} y_j \quad \forall i \in I,\ j \in J. \tag{28}$$

Now, (28), (16), (17), and (18) can be easily transferred to the original notation.

References

Aros-Vera, F., Marianov, V., & Mitchell, J. E. (2013). p-Hub approach for the optimal park-and-ride facility location problem. *European Journal of Operational Research, 226*(2), 277–285.
Ben-Akiva, M., Mc Fadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., et al. (2002). Hybrid choice models: Progress and challenges. *Marketing Letters, 13*(3), 163–175.
Benati, S. (1999). The maximum capture problem with heterogeneous customers. *Computers & Operations Research, 26*(14), 1351–1367.
Benati, S., & Hansen, P. (2002). The maximum capture problem with random utilities: Problem formulation and algorithms. *European Journal of Operational Research, 143*, 518–530.
Haase, K. (2009). *Discrete location planning. Technical Report WP-09-07*. Institute for Transport and Logistics Studies, University of Sydney.
Haase, K., & Müller, S. (2013). Management of school locations allowing for free school choice. *Omega, 41*(5), 847–855.
Marianov, V., Rfos, M., & Icaza, M. J. (2008). Facility location for market capture when users rank facilities by shorter travel and waiting times. *European Journal of Operational Research, 191*(1), 30–42.
McFadden, D. (1973). Conditional logit analysis of qualtitative choice behaviour. In P. Zarembka (Ed.), *Frontiers of Econometrics* (pp. 105–142). New York: Academic Press.
McFadden, D. (2001). Economic choices. *The American Economic Review, 91*(3), 351–378.
Müller, S., Haase, K., & Kless, S. (2009). A multi-period school location planning approach with free school choice. *Environment and Planning A, 41*(12), 2929–2945.
Müller, S., Haase, K., & Seidel, F. (2012). Exposing unobserved spatial similarity: Evidence from german school choice data. *Geographical Analysis., 44*, 65–86.
Serra, D., Eiselt, H., Laporte, G., & ReVelle, C. (1999). Market capture models under various customer-choice rules. *Environment and Planning B, 26*, 741–750.
Zhang, Y., Berman, O., & Verter, V. (2012). The impact of client choice on preventive healthcare facility network design. *OR Spectrum, 34*, 349–370.

OR Spectrum
DOI 10.1007/s00291-014-0367-6

REGULAR ARTICLE

Insights into clients' choice in preventive health care facility location planning

Knut Haase · Sven Müller

Abstract In this contribution we build on the approach proposed by Zhang et al. (OR Spectrum 34:349–370, 2012) to consider clients' choice in preventive health care facility location planning. The objective is to maximize the participation in a preventive health care program for early detection of breast cancer in women. In order to account for clients' choice behavior the multinomial logit model is employed. In this paper, we show that instances up to 20 potential locations and 400 demand points can be easily solved (to optimality or at least close to optimality) by a commercial solver in reasonable time if the problem is modeled by an alternative formulation. We present an intelligible approach to derive a lower bound to the problem. Our paper provides interesting insights into the trade-off between minimum workload requirement (to ensure quality of care) and participation (and thus early diagnosis of disease). We present a general definition of clients' utility (which allows for clients' characteristics, for example) and discuss some fundamental issues (and pitfalls) concerning the specification of utility functions.

Keywords Facility location · Multinomial logit model · Random utility · Preventive health care · Congestion · Capacity constraints

1 Introduction

In Zhang et al. (2012) a probabilistic-choice model and a so-called optimal-choice-model for locating preventive health care facilities are proposed. The objective of both

Accepted for Publication in OR Spectrum.

K. Haase · S. Müller ( )
Fakultät für Betriebswirtschaft, Institut für Verkehrswirtschaft, Lehrstuhl für Betriebswirtschaftslehre, insbesondere Verkehr, Universität Hamburg,
Von-Melle-Park 5, 20146 Hamburg, Germany
e-mail: sven.mueller@wiso.uni-hamburg.de
URL: http://www.wiso.uni-hamburg.de/verkehr

Published online: 25 May 2014

models is the maximization of the participation in a preventive health care program for early detection of breast cancer in women, i.e. the yearly expected number of women who access a medical checkup at a mammography center. A mammography machine is generally called a server. A mammography center (facility) can operate more than one server. A minimum workload requirement must be considered at each mammography facility to ensure that doctors are sufficiently experienced. At the same time one has to account for a maximum number of clients per unit of time to ensure a given service level (i.e., maximum waiting time). Zhang et al. (2012) model each facility as an $M/M/d$ queuing system, with d as the number of servers at the facility. For a more detailed discussion about preventive health care facility location planning we refer to Zhang et al. (2012). In contrast with sick people who need urgent medical attention, the clients of preventive health care choose whether to take part in a preventive health care program. That is, clients choose to patronize a certain facility location or not to take part in the program. For the so-called "probabilistic-choice model" Zhang et al. (2012) employ a specific discrete choice model, namely the multinomial logit model (MNL) to model the clients' choice. Probabilistic models other than MNL might be used as well (see Achabal et al. 1982; Drezner 1994; and Berman and Krass 2002, for example). Using MNL within a mathematical program possibly yields a non-linear formulation as outlined by Benati and Hansen (2002), Marianov et al. (2008), and Haase and Müller (2013). Since Zhang et al. (2012) assume that waiting time and the quality of care do not influence the choice behavior of the clients, the decision is only about the locations of the facilities and the number of servers per facility. For this problem Zhang et al. (2012) present a set of linear constraints which are capable to reproduce the non-linear choice probabilities described by the MNL. The authors report difficulties to solve problem sets with more than 10 potential locations and 100 demand points to optimality using CPLEX (IBM ILOG 2009). Therefore, heuristic approaches are used to solve the problem efficiently.

In this contribution we describe a set of tight linear constraints in order to incorporate the MNL in a linear mixed-integer program. These constraints are based on the approach of Haase (2009). We present an intelligible approach to derive a lower bound for the health care facility location problem. We show by a numerical investigation that using our approach enables to solve so-called mid-sized instances with 20 potential facilities and 400 demand points to optimality (or at least close to optimality) within one hour using GAMS/CPLEX (McCarl et al. 2008). Additionally, we discuss an interesting trade-off between the minimum required workload of a server and the participation in the preventive health care program. Although this trade-off has important impacts on policy implementations it has not been discussed in the presence of probabilistic choice behavior of clients. Finally, we show that purely generic specifications of utility are quite restrictive. In this contribution we present a more general specification of the utility function that is applicable to a wide range of case studies. In the next section we discuss clients' choice behavior and a mathematical formulation of the preventive health care facility location planning problem. This is followed by an approach to derive a lower bound to the problem. The section ends with a discussion of the negative impact of the minimum workload requirement on the participation rate. In Sect. 3 we present the results of our computational studies. In addition, the impact of the minimum workload requirement on participation is examined. We compare the performance of our approach to the model proposed by Zhang et al. (2012). A conclusion is given in Sect. 4.

2 Mathematical program

Let J be the set of potential locations for the facilities. We assume that a client will always choose to patronize the facility j that maximizes its utility. Since a client might choose not to patronize any facility, the general choice set M is {"no choice"} $\cup J$. If we now further assume that all clients located in a given demand node (i.e., zone) $i \in I$ exhibit the same observable characteristics, then the deterministic part of utility of clients located in $i \in I$ choosing $j \in M$ may be given by

$$v_{ij} = \sum_{l=1}^{L} \beta_{jl} c_{ijl}, \tag{1}$$

with c_{ijl} as the value of attribute l (distance or parking space, for example) and β_{jl} as the corresponding weight. The deterministic part of utility of the "no choice" alternative is denoted by v_{i0}. From a modelers perspective the total utility of the clients is not exactly known, i.e. only the deterministic part is known. Therefore, we can only determine a probability that a client located in i patronizes a given facility location j. According to certain assumptions, the choice probability of clients located in $i \in I$ choosing $j \in M$ is given by the MNL:

$$p_{ij} = \frac{e^{v_{ij}}}{\sum_{m \in M} e^{v_{im}}}. \tag{2}$$

It is important to note that the MNL of (2) immediately implies that clients choose the alternative that maximizes utility. That is, clients choose the most attractive facility location. For a detailed reasoning and further explanations concerning utility maximization, random utility models, and MNL we refer to the Appendix. Note that $\sum_{j \in J} p_{ij} < 1$ due to the definition of J.

2.1 Model formulation

Let us consider the parameters

- g_i demand in node $i \in I$ given in number of clients who require service for the given time period,
- $\overline{p}_{ij}$ maximum choice probability of clients located in $i \in I$ patronizing a facility located in $j \in J$. That is, j is the only established facility location, i.e. $\overline{p}_{ij} = e^{v_{ij}} / (e^{v_{ij}} + e^{v_{i0}})$,
- $\widetilde{p}_i$ choice probability of clients located in $i \in I$ not demanding health care services (i.e., choosing the "no choice" alternative) given that all facilities are located. That is, at each location $j \in J$ exists a facility. For a given i: $\widetilde{p}_i = e^{v_{i0}} / \left(e^{v_{i0}} + \sum_{j \in J} e^{v_{ij}}\right)$ and hence $\widetilde{p}_i + \sum_{j \in J} p_{ij} = 1$,

q_k maximum number of clients to be processed by $k = 1, \ldots, K$ servers per time period such that a certain service level (expected waiting time for an appointment, for example) is not exceeded. At most K servers can be allocated to each facility. In Canada, for example, the waiting time for medical imaging diagnostics (that is, what mammography is about) was between 30 and 160 days in 2011 (Canadian Institute for Health Information 2011). In contrast, the nationwide benchmark was 4 weeks in 2012 (Wait Time Alliance 2012). For further information we refer to Health Council of Canada (2007). Most recently, Tejada et al. (2014) quantified the negative impact of the degree to which patients get an appointment on the probability of non adherence. Finally, consider the parameters,

$R_{\min}$ minimum workload requirement in clients per time period. This ensures that examiners are experienced and thereby a certain level of service (a minimum true-positive-rate, for example) is guaranteed, and

$Q_{\max}$ maximum number of established servers.

Then, we define the binary decision variables

$y_j = 1$, if location $j \in J$ provides a health care facility (0, otherwise), and
$w_{jk} = 1$, if location $j \in J$ has k or more servers (0, otherwise).

Further, we denote the non-negative variable

x_{ij} as the choice probability of clients at node $i \in I$ access health care service at location $j \in J$. We might define x_{ij} in a non-linear way as

$$x_{ij} = \frac{e^{v_{ij}} y_j}{e^{v_{i0}} + \sum_{m \in J} e^{v_{im}} y_m} \quad \forall\, i \in I, j \in J. \tag{3}$$

So x_{ij} denotes the choice probabilities in the solution for clients located in $i \in I$ choosing $j \in J \,|y_j = 1$. Finally, we consider the variables

$\widetilde{x}_i$ choice probability of clients located in $i \in I$ not demanding health care services (i.e., choosing the "no choice" alternative), and

F^A objective function value: expected participation in preventive health care.

In order to avoid the non-linearity in (3) Zhang et al. (2012) propose the model formulation as given in the Appendix (Model Z in Appendix A). In this paper, we present an alternative formulation of the problem as Model A:

$$\max F^A = \sum_{i \in I} g_i \sum_{j \in J} x_{ij} \tag{4}$$

subject to

$$\widetilde{x}_i + \sum_{j \in J} x_{ij} \leq 1 \qquad \forall\, i \in I \tag{5}$$

$$x_{ij} \leq \overline{p}_{ij} y_j \qquad \forall\, i \in I, j \in J \tag{6}$$

$$\widetilde{p}_i x_{ij} \leq p_{ij} \widetilde{x}_i \qquad \forall\, i \in I, j \in J \tag{7}$$

$$\widetilde{p}_i x_{ij} \geq p_{ij} \widetilde{x}_i + y_j - 1 \qquad \forall\, i \in I, j \in J \tag{8}$$

$$\sum_{i \in I} g_i x_{ij} \geq R_{\min} y_j \qquad \forall\, j \in J \tag{9}$$

$$\sum_{i \in I} g_i x_{ij} \leq \sum_{k=1}^{K} q_k w_{jk} \qquad \forall\, j \in J \tag{10}$$

$$\sum_{k=1}^{K} w_{jk} = y_j \qquad \forall\, j \in J \tag{11}$$

$$\sum_{j \in J} \sum_{k=1}^{K} k \cdot w_{jk} \leq Q_{\max} \tag{12}$$

$$x_{ij} \geq 0 \qquad \forall\, i \in I, j \in J \tag{13}$$

$$\widetilde{x}_i \geq 0 \qquad \forall\, i \in I \tag{14}$$

$$y_j \in \{0, 1\} \qquad \forall\, j \in J \tag{15}$$

$$w_{jk} \in \{0, 1\} \qquad \forall\, j \in J, k = 1, \ldots, K \tag{16}$$

Constraints (5)–(8) in conjunction with (4) are an exact linear reformulation of (3). (6) is tighter than simply using $x_{ij} \leq y_j \; \forall i, j$ yielding a smaller upper bound by the corresponding LP-relaxation. Note, (7) and (8) avoid redundant constraints used by Aros-Vera et al. (2013): (7) and (8) yield $2 \cdot |I| \cdot |J|$ rows instead of $|I| \cdot |J|^2$ rows. (9) and (10) are capacity constraints to guarantee service levels (i.e. waiting time and quality of physical examination). (11) stipulates that servers are only available at established facilities. The total number of servers is limited by (12).

Although constraints (29)–(32) of the model MODEL Z (in the Appendix) are equivalent to (5)–(8) to consider (3) in a linear way, they are less tight and thus might hamper the solvability of the problem. We get back to this issue in the computational studies in Sect. 3.

2.2 Modeling framework to derive a lower bound

Although we expect better performance of MODEL A (4)–(16) compared to MODEL Z (28)–(39) due to the tight formulation of constraints, we may encounter problems concerning solvability. Therefore, we discuss an intelligible approach to determine a lower bound for the preventive health care facility location problem (MODEL A) in this section.

Consider the variables

Ω_{ij} = 1, if node $i \in I$ is assigned to facility location $j \in J$ (0, otherwise),
Ψ_j = 1, if location $j \in J$ provides a health care facility (0, otherwise),
F^B objective function value indicating the cumulative choice probabilities,
F^C objective function value indicating the attractiveness of the located facilities, and
LB objective function value; lower bound to MODEL A.

Then we define MODEL B as

$$\max F^B = \sum_{i \in I} \sum_{j \in J} p_{ij} \Omega_{ij} \tag{17}$$

subject to

$$\sum_{j \in J} \Omega_{ij} = Q_{\max} \quad \forall\, i \in I \tag{18}$$

$$\Omega_{ij} \in \{0, 1\} \quad \forall\, i \in I, j \in J \tag{19}$$

MODEL B determines for each demand node i the $Q_{\max}$ most attractive facility locations. Now let us denote

$$b_j = \sum_{i \in I} g_i \Omega^*_{ij} \quad \forall\, j \in J \tag{20}$$

as a quantity with Ω^*_{ij} as the optimal values according to the solution of MODEL B. b_j might be interpreted as an attractiveness value for each facility location. Facility locations that are attractive to many clients obtain high values of b_j. Then MODEL C, written as

$$\max F^C = \sum_{j \in J} b_j \Psi_j \tag{21}$$

subject to

$$\sum_{j \in J} \Psi_j = Q_{\max} \tag{22}$$

$$\Psi_j \in \{0, 1\} \quad \forall\, j \in J \tag{23}$$

determines the $Q_{\max}$ most attractive facility locations. Now consider MODEL D

$$\max \text{LB} = \sum_{i \in I} g_i \sum_{j \in J} x_{ij} \tag{24}$$

subject to (5)–(16) and

$$y_j \leq \Psi^*_j \quad \forall\, j \in J \tag{25}$$

with Ψ_j^* according to the solution of MODEL C. MODEL D determines the lower bound to MODEL A. In order to employ a lower bound for MODEL A consider the following procedure:

Step 1: Solve MODEL B: (17)–(19)
Step 2: Solve MODEL C: (21)–(23)
Step 3: Solve MODEL D: (24), (5)–(16), and (25)
Step 4: Set $\underline{F}^A = LB^*$ and solve MODEL A.

2.3 Minimum workload requirement and participation

In this section we discuss an interesting trade-off between the minimum workload requirement $R_{\min}$ in (9) [and (35)] and the maximized participation F^A in (4) [and (28)]. Obviously, the true objective is to minimize the (female) breast cancer mortality. Of course, many aspects have to be considered. One aspect is to detect cancer at the most possible early stage to increase prognosis. Many different factors might influence the "detection rate". As the detection rate we consider the share of clients with a true-positive diagnosis with respect to all clients with a carcinoma. One factor is participation (F^A), i.e. how many women access preventive health care facilities, while a second factor is the quality of the preventive health care facility itself. Only experienced physicians are expected to guarantee a certain level of service, i.e. correctly identified carcinomas (this corresponds to a high true-positive-rate). It is generally agreed upon that there is a positive relationship between the number of examinations and the level of experience of a physician (Perry et al. 2006). Therefore, the second factor is given by minimum workload $R_{\min}$. However, a high service quality, i.e. a high $R_{\min}$, may lead to a suboptimal solution concerning the detection rate or rather the breast cancer mortality. Since we consider a probabilistic choice behavior of clients, the consideration of (high values of) $R_{\min}$ may yield solutions that exclude locations with very high choice probabilities, because these very attractive locations are dominating other locations in terms of choice probabilities. As a consequence, the dominated locations do not meet the necessary choice probabilities to satisfy $R_{\min}$. This yields a lower participation rate F^A compared to any solution with a lower $R_{\min}$. Roughly speaking, a high $R_{\min}$ might profoundly reduce participation. At the same time the level of service corresponding to $R_{\min}$ might not outweigh this reduction.

To view the underlying coherences in a proper perspective we consider the following very simple example. Let us consider node $i = 1$, three potential facilities $j \in \{A, B, C\}$ and $Q_{\max} = 2$, as well as the deterministic utility values $v_{1A} = -1.5$, $v_{1B} = -1$, $v_{1C} = -1.5$, and $v_{1,0} = 0$ (no-choice alternative, i.e. to not participate). We assume that only one server can be installed at a facility. The corresponding choice probabilities and participation rates are given in Table 1. Let $R_{\min}$ be given in such a way that a choice probability of 0.15 is required. Then the solution with the location set $\{A, C\}$ would be optimal. But the infeasible solution with the locations B and C would provide a higher participation rate. Considering solution $\{B, C\}$ the question is whether the lower service quality of the facility located at C outweighs

Table 1 Example: participation and detection rates. α_j is the true-positive-rate

	v_{1j}	$\{B, C\}$		$\{A, C\}$	
		x_{1j}	α_j	x_{1j}	α_j
Facility $j = A$	-1.5	0.000		0.154	0.9
Facility $j = B$	-1.0	0.231	0.9	0.000	
Facility $j = C$	-1.5	0.140	0.8	0.154	0.9
No-choice $\widetilde{x}_1$, $j = 0$	0	0.629		0.691	
Participation rate	–	0.371		0.309	
Detection rate	–	0.320		0.276	

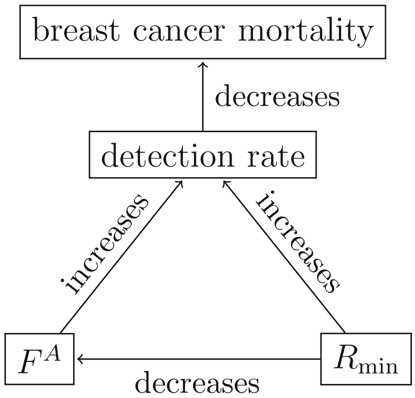


Fig. 1 Theoretical dependencies between F^A, $R_{\min}$, and detection rate. As $R_{\min}$ increases the number of located facilities decreases. In general, this yields a reduced participation F^A. However, high values of $R_{\min}$ are associated with high true-positive-rates and, therefore, it is assumed to yield high rates of detection of carcinomas. Whether high values of $R_{\min}$ increase or indirectly decrease the detection rate depends on the application and data (see Sect. 3)

the difference in participation in terms of detected carcinomas? To answer this question we have to assume a true-positive-rate α_j for each facility as stated in Table 1. The true-positive-rate may be a function of utilization (see Sect. 3). Obviously, we would prefer the solution without constraints (9), i.e. without the minimum workload requirement $R_{\min}$. Even if we assume a quite poor $\alpha_C = 0.5$, solution $\{B, C\}$ yields a better detection rate than solution $\{A, C\}$. A generalization of these dependencies is displayed in Fig. 1.

We conclude that from a public health perspective constraints (9) have to be evaluated very carefully in terms of participation (F^A) and different levels of minimum workload ($R_{\min}$). The minimum workload requirement may favor solutions with facility locations that exhibit choice probabilities which indeed satisfy the minimum workload. However, the average choice probabilities over the established facility locations (market shares) may be smaller than the average choice probabilities without (or with a lower) minimum workload requirement.

Constraints (9) seem to be questionable from an economical point of view as well: assuming identical cost per server, solution $\{B, C\}$ outperforms $\{A, C\}$ in terms of cost per client. Of course, the more clients take part in the preventive health care program the lower is the ratio of the initial cost of a server and the number of clients. The same is true for the ratio of the expected number of detected carcinomas and the initial cost of a server. These issues (costs, trade-off) have to be considered in applications and political decision making. We suggest to determine the "political price" of constraints (9) by a comparison of the objective of the solution with respect to (9) and the objective of the solution with (at least partly) relaxed (9). So the difference is the political price for a given level of $R_{\min}$.

3 Computational investigation

In this section we set out to compare MODEL A and MODEL Z in terms of computational performance. In a second experiment we analyze the impact of the minimum workload requirement on participation and detection of cancer. We have implemented all models in GAMS 23.7 (McCarl et al. 2008) and we use CPLEX 12.2 CPLEX (IBM ILOG 2009) on a 64-bit Windows Server 2008 with 4 Intel Xeon 2.4 GHz processors and 24 GB RAM for solving the problems. We only use artificial data in our experiments.

A crucial input to the model is the deterministic part of utility v_{ij} of (1). In applications, the weights (coefficients) β_{jl} should be estimated using empirical individual-level choice data. Street and Burgess (2007) demonstrate how the data could be easily obtained (particularly for health care applications). The estimation of β_{jl} and the specification of v_{ij} are illustrated for locational choice applications by Anderson et al. (1992), for example. Now, Zhang et al. (2012) define the deterministic utility component of (1) as

$$\begin{aligned} v_{i0} &= 0 \qquad \text{for dummy facility } j = 0 \text{ ("no-choice")} \\ v_{ij} &= -\beta t_{ij} \quad \text{for all non-dummy facilities } j \neq 0 \text{ with } \beta > 0, \end{aligned}$$

where attribute t_{ij} is the travel time of an individual (client) from zone i to facility $j > 0$. Hence, $p_{i0} \geq p_{ij}$ for all $j > 0$ which significantly restricts the number of potential applications. Obviously, the maximum value of p_{ij} for $j \neq 0$ is 0.5. Now, assume that 95 % of the individuals in zone i would use the facility j if $t_{ij} = 0$. The specification of the utility function in this example cannot be adapted to this assumption as $0.5 < 0.95$. Although a generic specification of v_{ij} in general and for t_{ij} in particular might be reasonable, an alternative specific specification, i.e. $v_{ij} = \text{ASC}_j - \beta_j t_{ij}$ may be reasonable as well. The alternative-specific constants ASC_j guarantee that the location-specific market shares are met (Bierlaire et al. 1997 and Train (2003), [p. 66]). In applications, it might happen that one does not obtain estimates of the alternative specific constants for all (potential) locations. Thus, it is not always possible to consider a full set of alternative specific constants. In such a case, we refer to Müller et al. (2012) and Haase and Müller (2013) for an appropriate specification.

The alternative-specific specification of travel-time $\beta_j t_{ij}$ might be useful as well. For example, consider that clients located in i have to choose between two facility locations, i.e. A and B. Assume $t_{i\text{A}} = t_{i\text{B}}$. Certainly, there are other (medical) facilities proximate to facility location A (the facilities might even share the same location). This might not be the case for facility B. Now, clients are able to "bundle purchases" (i.e. appointments with other doctors, like dentists etc.) when choosing facility location A (Carling and Haansson 2013). In an alternative specific specification of utility this yields $\beta_A < \beta_B$ if one does not include ASC_j or explicitly account for bundled purchases in the utility function. Using a generic specification, the choice probabilities may be biased. As one can easily verify, the strict generic specification of utility is rather restrictive and it does not allow for socio-economic characteristics of the clients (age and income, for example). The advantage of choice model specifications including socio-economic characteristics of the clients is shown by Müller and Haase (2014). However, to make our study comparable to the study of Zhang et al. (2012) we consider a strict generic specification of utility.

We employ the following procedure to generate artificial data for the computational studies: let $\delta_1 \in \mathbb{Z}$ denote the number of demand nodes and $\delta_2 \in \mathbb{Z}$ denote the number of nodes of potential facility locations; then consider

Step 1: nodes:
- set $| I |= \delta_1$ and $| J |= \delta_2$

Step 2: population given $i \in I$
- (a) select randomly $\phi_i \in [0, 2.4]$
- (b) $h_i := \phi_i \cdot |J| / |I|$
- (c) $g_i := h_i$

Step 3: travel times, utility, choice probabilities given $i \in I$ and $j \in J$
- (a) generate randomly Cartesian coordinates of nodes in the interval [0, 100] (uniform distribution)
- (b) t_{ij} := rectangle distance from node $i \in I$ to facility $j \in J$ divided by 60
- (c) $v_{ij} := -\beta \cdot t_{ij}$ with $\beta = 2$
- (d) $v_{i0} := 0$
- (e) $p_{ij} := \frac{e^{v_{ij}}}{\sum_m e^{v_{im}}}$
- (f) $\widetilde{p}_i := \frac{1}{1+\sum_j e^{v_{ij}}}$
- (g) $\bar{p}_{ij} := \min\{0.5, e^{v_{ij}}/(1 + e^{v_{ij}})\}$

Step 4: other parameters according to Zhang et al. (2012)
- $\lambda := 1$
- $K := 4$
- $\nabla\lambda_k := 1.05^k$
- $R_{\min} := 1.2$
- $q_k := B \cdot \nabla\lambda_k$ with $B = 1.5 \cdot R_{\min}$
- $Q_{\max} := \lceil(|J| /2)\rceil$
- $M_1 = M_2 := 1$

In order to compare performances of the models we employ the data generating procedure with $\delta_1 = \{100, 200, 400\}$ and $\delta_2 = \{10, 20, 40\}$ yielding 9 problem sets with 10 randomly generated instances each. We set the maximum CPU usage to 1 hour for each instance, model, and problem set. Capacities might not be exhausted due to parameter settings. The results can be found in Table 2. Given a problem set, CPLEX provides for each instance a gap. This gap denotes the relative deviation between the best integer solution found and the theoretical maximum (a bound for the optimal integer solution). The average of this gap over all ten instances per problem

Table 2 Computational results

$\|J\|$	$\|I\|$	Model Z "GAP"	CPU	Model A "GAP"	CPU	Model A with $\underline{F}^A = LB^*$ "GAP"	CPU (LB^*)	$100 \frac{F^{A*} - LB^*}{F^{A*}}$
10	100	0	35	0	4	0	9 (2)	1.55
	200	0	196	0	15	0	23 (2)	4.18
	400	0	831	0	60	0	51 (7)	2.97
20	100	406.44[b]	3,600	0	619	0	137 (1)	2.48
	200	–	3,600	1.66	2,858	0	604 (2)	1.66
	400	185.34[d]	3,600	7.35	3,600	0.22	1,993 (2)	0.93[a]
40	100	89.99[c]	3,600	8.58	3,600	3.34	3,600 (8)	–
	200	96.44[c]	3,600	27.88	3,600	8.21	3,600 (25)	–
	400	92.69[d]	3,600	55.67	3,600	12.10	3,600 (75)	-

All models yield identically optimal objective function values. "GAP" denotes the average gap (provided by CPLEX) in % over all ten instances. CPU denotes the average CPU usage (by CPLEX) in seconds over all ten instances. In the column "CPU" of Model A with lower bound the values denote the CPU usage for step 4, while the numbers in brackets denote the time to derive the lower bound LB^* (steps 1–3). The maximum time has been set to 1 h Number of instances where an optimal solution is obtained: *a*: 9, *b*: 5, *c*: 3, *d*: 1

set is reported in the columns denoted by "GAP". For each problem set, the average computational effort (measured in CPU seconds) over ten instances is given in column "CPU". We consider Model A two times: (1) without lower bound and (2) with lower bound (see Sect. 2.2). The last column of Table 2 denotes the relative deviation of the lower bound LB^* from the optimal integer solution.

The results clearly show the advantage of Model A. This finding is generally confirmed by Haase and Müller (2014). In particular, we see that for mid-sized ($|J| = 20$) and large-sized ($|J| = 40$) problem sets the lower bound is of great advantage. For small-sized problem sets the quality of the lower bound is quite good (gap <5 %). The efficiency of the lower bound procedure as described in Sect. 2.2 is comparable to the results of the heuristics proposed by Zhang et al. (2012). In contrast to the heuristic approach we are able to evaluate the results in terms of effectiveness: The average gap for large-sized problem sets is somewhat more than 12 % (within 1 h computation time). Interestingly, we observed that for small problem sets most of the time needed is to prove optimality. That is, there is only little improvement of the integral solution but a lot of time is needed to improve the upper bound. Benati and Hansen (2002) observed similar patterns. If we assume that this pattern is true for larger problem sets as well, then the "true" gap would be smaller than 12 % found so far. We consider the first random instance for each of the following problem sets

1. $|J| = 40, |I| = 100$
   - $F^{A*} = 33.5555$; CPU = 6 h and 42 min
   - $F^A = 33.5555$; CPU seconds = 3,600; CPLEX-GAP = 4.22 %
   - $100 \left(F^{A*} - F^A\right) / F^{A*} = 0$ %
2. $|J| = 40, |I| = 200$
   - $F^{A*} = 35.7445$; CPU = 2 days and 14 h

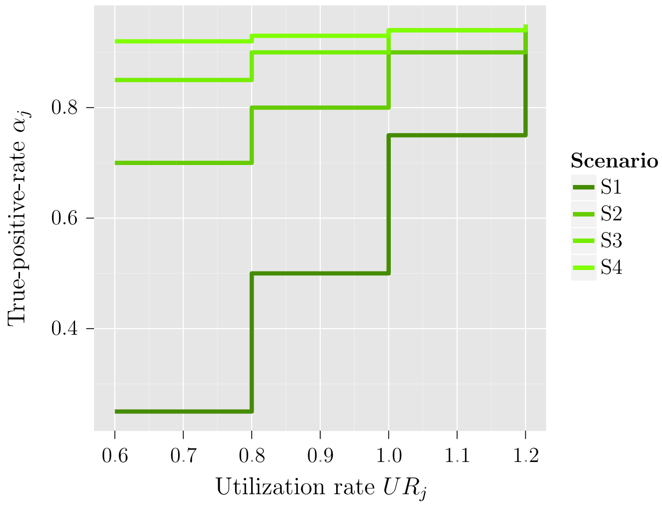


Fig. 2 True-positive-rate α_j. The values given are the true-positive-rates in a scenario respective to a given utilization rate. For example, for a facility located at j with an utilization rate UR_j of 0.9, the true-positive-rate α_j in scenario S1 is 0.5

- $F^A = 35.4280$; CPU seconds = 3,600; CPLEX-GAP = 7.02 %
- $100\left(F^{A*} - F^A\right)/F^{A*} = 0.89$ %

3. $|J| = 40$, $|I| = 400$
   - $F^{A*} = 37.0269$; CPU = 2 weeks and 5 days
   - $F^A = 36.4848$; CPU seconds = 3600; CPLEX-GAP = 8.50 %
   - $100\left(F^{A*} - F^A\right)/F^{A*} = 1.46$ %

These samples show that the true gap is lower than the one reported by CPLEX after one hour of computational time. So we are confident of the quality of the lower bound. In a second experiment we extend the small numerical example of Sect. 2.3 to investigate the impact of the minimum workload requirement of (9) on the participation and the detection rate of carcinoma. Therefore, we first define the utilization rate as

$$UR_j = \sum_{i \in I} g_i x_{ij} \quad \forall\, j \in J, \tag{26}$$

We consider four different scenarios of the coherence between UR_j and true-positive-rate α_j as given in Fig. 2. Of course, scenario S1 is unrealistic: a true-positive-rate of <50 % would make the whole diagnostic procedure questionable. However, we consider this scenario for comparison reasons. The most realistic scenarios would be somewhere between scenarios S2 and S4.

We define

$$DP = \sum_{i \in I} g_i \sum_{j \in J} \alpha_j x_{ij} \tag{27}$$

as the total detection potential. Given a preventive health care facility network, DP denotes the number of clients for which a detection of a carcinoma is possible. We consider $|J| = 20$, $|I| = 100$, and $B = 1.2 \cdot 1.5 = 1.8$. All other parameters

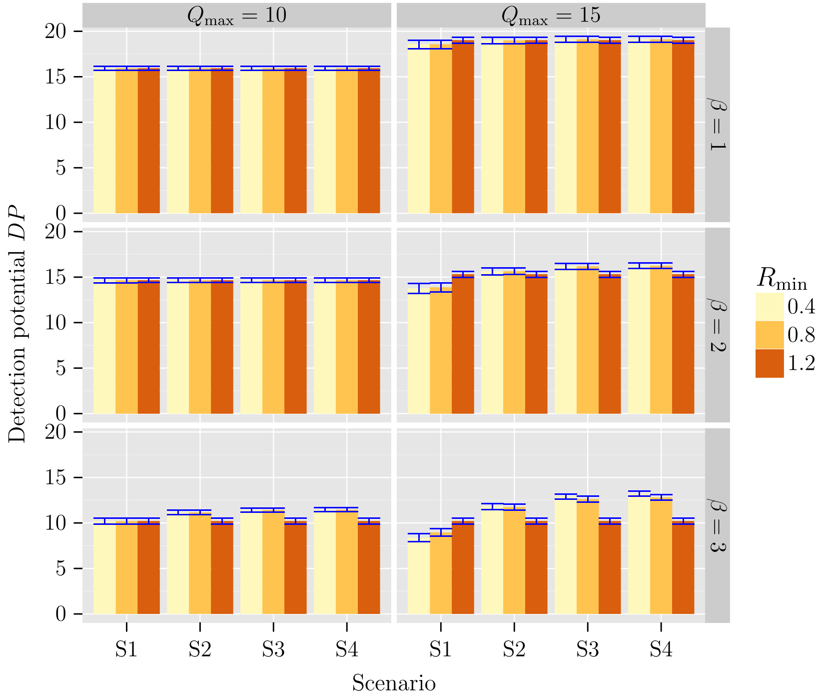


Fig. 3 Detection potential (DP) for scenarios with respect to minimum workload requirement ($R_{\min}$), maximum number of feasible servers ($Q_{\max}$), and travel-time sensitivity (β). Settings: $|I| = 100$, $|J| = 20$, and $B = 1.2 \cdot 1.5 = 1.8$. Values given are the averages over 10 random instances

are identical to the first computational study. We compute 10 random instances for $R_{\min} = \{1.2, 0.8, 0.4\}$. The results are given in Figs. 3 and 4. The findings can be summarized as follows:

1. Participation (F^{A*}) and detection potential (DP) decrease as the travel-time sensitivity (β) increases. This is reasonable since the more sensitive clients react on travel-time the less attractive is participation, i.e. to patronize a facility.
2. In general, the more servers are available ($Q_{\max}$) the higher is participation (F^{A*}) and the detection potential (DP).
3. If resources are scarce (i.e., $Q_{\max} = 10$), then there is no difference between scenarios or levels of minimum workload ($R_{\min}$) in terms of participation (F^{A*}) and detection potential (DP), because all located facilities exhibit utilization (UR_j) larger than 1.2 (yielding high values of α_j).
4. If high travel-time sensitivity (i.e., $\beta = 3$) meets scarce resources (i.e., $Q_{\max} = 10$) then we observe a negative impact of large values of minimum workload (i.e., $R_{\min}$ $= 1.2$) on participation (F^{A*}) and detection potential (DP). Compared to lower values of minimum workload, a large value of minimum workload implies low numbers of located facilities, and hence lower values of participation and detection potential.
5. If resources are not scarce (i.e., $Q_{\max} = 15$), then participation (F^*) and detection potential (DP) decline in minimum workload ($R_{\min}$) — given that the true-positive-rate (α_j) does not decline fast with decreasing utilization (UR_j), i.e. scenarios

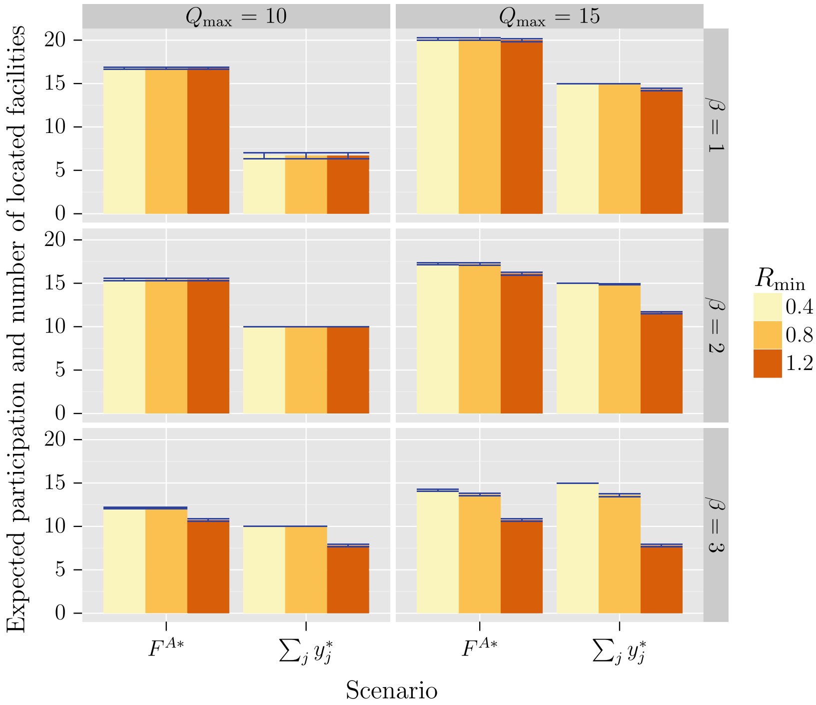


Fig. 4 Expected participation (F^{A*}) and number of located facilities ($\sum_j y_j^*$) with respect to minimum workload requirement (R_{min}), maximum number of feasible servers (Q_{max}), and travel-time sensitivity (β). Values given are the averages over ten random instances. Settings: see Fig. 3

S2, S3, and S4. This pattern becomes more explicit if travel-time sensitivity (β) increases. Of course, more facilities are located if minimum workload is low and hence participation and detection potential are higher compared to high values of minimum workload.

4 Conclusion

As our computational investigation shows, heuristics—as proposed by Zhang et al. (2012)—are not compulsory to solve the preventive health care facility location problem, if one uses a tight model formulation as proposed in this paper. In particular, our approach to derive a lower bound seems to be very promising, because in contrast to the findings of Zhang et al. (2012) we are able to solve problem sets with 20 potential locations and up to 400 demand points to (or at least close to) optimality within one hour using CPLEX. We find that the lower bound—computed in a few seconds—deviates at most 5 % from the optimal solution. For sets with 40 potential locations we are able to improve the gap by nearly 70 % within one hour of computational time using the lower bound procedure. The resulting gap (reported by CPLEX) is between 3 and 12 %. The computational study shows that the our model is advantageous in terms of solvability compared to the model proposed by Zhang et al. (2012). Our experiments indicate that the optimal solution is found quite rapidly and most of the computational time is needed to prove optimality. According to this, the development of more efficient and effective approaches to derive better bounds might be an objective. At least our approach might be used to analyze the quality of heuristics.

Another important insight into the preventive health care facility location problem is provided by our discussion of the impact of the minimum workload requirement on the expected number of detected carcinomas ("detection potential"). If the goal is to maximize this objective by a maximization of the participation (rate) subject to the minimum workload requirement as a proxy for service quality our computational study reveals that minimum workload requirement

- is compulsory, if the true-positive-rate declines very fast with decreasing utilization of facilities,
- might be useful, if travel-time sensitivity is low and/or resources (i.e. number of servers) are scarce,
- should be rather low or avoided if travel-time sensitivity is high and/or resources are not scarce.

If we reconsider the discussion in Sect. 2.3—in particular Fig. 1—in the light of the findings of Sect. 3, we can say that the decrease of participation in minimum workload requirement yields a decline in detection rate which is larger than the increase of detection rate in minimum workload requirement if clients react sensitive on travel-time. This pattern is more visible, if the number of available servers is large. Based on these findings, the explicit maximization of the detection potential using a true-positive-rate as a function of utilization would be an interesting future research objective.

We agree with Zhang et al. (2012) that a thorough empirical investigation of the clients (by discrete choice analysis) to understand their choice behavior is necessary. In our paper we present a general definition of clients' utility to use the results of such an investigation in a mathematical program. Moreover, the incorporation of expected waiting times (Marianov et al. 2008) and quality of care (true-positive rate, for example) in clients' utility function would be an interesting future research objective as proposed by Zhang et al. (2012).

Another interesting issue for future research is the incorporation of the choice set generating process in mathematical models. One approach might be the use of a threshold value as proposed by Haase (2009). Another possible research question is the consideration of flexible substitution patterns of facilities (Müller et al. 2009). The MNL provides constant substitution patterns by construction. However, it is well known from empirical studies on locational choice behavior that constant substitution patterns are unlikely to exist (Sener et al. 2011).

Acknowledgments The very helpful comments of two anonymous reviewers and an associate editor are gratefully acknowledged.

Appendix

Model proposed by Zhang et al. (2012)

Additionally to the definitions of Sect. 2.1 we denote the parameters

h_i fraction of clients at node i,
λ expected number of clients per period over the entire area (Poisson rate); the Poisson rate of node i is $\lambda \cdot h_i$,
$\overline{\lambda}_k$ maximum participation rate at a facility with k servers; $\overline{\lambda}_0 = 0$,
$\Delta\lambda_k = \overline{\lambda}_k - \overline{\lambda}_{k-1}$,
M_1 big number; = 1 for example Zhang et al. (2012),
M_2 big number; = 1 for example Zhang et al. (2012),

as well as

z_{ijo} artificial continuous variable for avoiding non-linearity with $o \in J$; corresponds to the result of $x_{ij} w_{o1}$.

and the mathematical model (Model Z) corresponding to the original contribution of Zhang et al. (2012)

$$\max F^Z = \lambda \sum_{i \in I} h_i \sum_{j \in J} x_{ij} \tag{28}$$

subject to

$$x_{ij} + \sum_{o \in J} e^{-\beta t_{io}} z_{ijo} = e^{-\beta t_{ij}} w_{j1} \qquad i \in I, j \in J \tag{29}$$

$$z_{ijo} \leq x_{ij} \qquad i \in I, j, o \in J \tag{30}$$

$$z_{ijo} \leq M_1 w_{p1} \qquad i \in I, j, o \in J \tag{31}$$

$$z_{ijo} \geq x_{ij} - M_2(1 - w_{p1}) \qquad i \in I, j, o \in J \tag{32}$$

$$\sum_{j \in J} \sum_{k=1}^{K} w_{jk} \leq Q_{\max} \tag{33}$$

$$w_{jk+1} \leq w_{jk} \qquad j \in J, k = 1, 2, \ldots, K-1 \tag{34}$$

$$\lambda \sum_{i \in I} h_i x_{ij} \geq R_{\min} w_{j1} \qquad j \in J \tag{35}$$

$$\lambda \sum_{i \in I} h_i x_{ij} \leq \sum_{k=1}^{K} \nabla\lambda_k w_{jk} \qquad j \in J \tag{36}$$

$$x_{ij} \geq 0 \qquad i \in I, j \in J \tag{37}$$

$$z_{ijo} \geq 0 \qquad i \in I, j, o \in J \tag{38}$$

$$w_{jk} \in \{0, 1\} \qquad j \in J, k = 1, 2, \ldots, K \tag{39}$$

Basics of discrete choice analysis

The MNL is well known for analyzing discrete choice decisions of individuals (McFadden 1973, 2001). Let N be the set of individuals (customers, clients etc.), M the choice set (set of alternatives the individual chooses from), and L the set of attributes or characteristics (attractiveness determinants). The choice set M must be exhaustive and the alternatives have to be mutually exclusive. Roughly speaking, all alternatives the individuals face have to be included in the choice set. Individual $n \in N$ chooses exactly one alternative from choice set M. In the discrete choice modeling literature it is assumed that an individual $n \in N$ chooses alternative $j \in M$ that maximizes utility (see Train 2003, for example). That is, n chooses j, iff

$$u_{nj} > u_{nm} \quad \forall\, m \in M, m \neq j. \tag{40}$$

The utility u_{nj} of alternative j for individual n consists of a deterministic component v_{nj} and a stochastic component ϵ_{nj}, i.e.

$$u_{nj} = v_{nj} + \epsilon_{nj}. \tag{41}$$

Usually, the deterministic component is modeled as a linear function:

$$v_{nj} = \sum_{l \in L} \beta_{jl} c_{njl}, \tag{42}$$

where c_{njl} is the value of attribute l concerning individual n and alternative j, and the coefficient β_{jl} is the utility contribution per unit of attribute l related to alternative j (Ben-Akiva and Lerman 1985). In applications, β_{jl} have to be estimated (by maximum-likelihood) using choice data from empirical studies (Anderson et al. 1992; Ben-Akiva and Lerman 1985; Louviere et al. 2000; Street and Burgess 2007; Müller et al. 2008; and Train 2003). Since u_{nj} of (41) is stochastic we can only make probabilistic statements about (40):

$$p_{nj} = \text{Prob}\left(u_{nj} > u_{nm}\ \forall\, m \in M, m \neq j\right). \tag{43}$$

Assuming that the stochastic component ϵ_{nj} is independent, identically extreme value distributed, the probability (43) that individual n chooses alternative j is determined by

$$p_{nj} = \frac{e^{v_{nj}}}{\sum_{m \in M} e^{v_{nm}}}, \tag{44}$$

which is the well-known MNL (Ben-Akiva and Bierlaire, 2003). Having said this, it is obvious that the MNL of (44) exhibits utility maximization behavior of the choice makers. In other words, using MNL means to assume that clients choose the facility that maximizes their utility (i.e., clients choose the most attractive—"the optimal"—facility). Hence, the underlying choice rule is utility maximization. Note, if $\epsilon_{nj} = 0\,\forall\, n, j$, then the choice problem of (40) becomes deterministic. Consider, for example, $u_{nj} = v_{nj} = -t_{nj}$ with t_{nj} as the travel time of individual n to location j. This means that clients wish to obtain services from the facility with the shortest travel time. Obviously, such a choice model is characterized to be deterministic. If we assume that all clients located in $i \in I$ exhibit the same observable characteristics, then (1) results from (42) and (2) results from (44).

References

Achabal D, Gorr W, Mahajan V (1982) Multiloc: a multiple store location decision model. J Retail 58:5–25

Anderson S, de Palma A, Thisse JF (1992) Discrete choice theory of product differentiation. MIT Press, Cambridge

Aros-Vera F, Marianov V, Mitchell JE (2013) p-hub approach for the optimal park-and-ride facility location problem. Eur J Oper Res 226(2):277–285

Ben-Akiva M, Bierlaire M (2003) Discrete choice models with applications to departure time and route choice. In: Hall R (ed) Handbook of transportation science, International series in operations research and management science, Kluwer, pp 7–38

Ben-Akiva M, Lerman S (1985) Discrete choice analysis, theory and application to travel demand. MIT Press, Cambridge

Benati S, Hansen P (2002) The maximum capture problem with random utilities: problem formulation and algorithms. Eur J Oper Res 143:518–530

Berman O, Krass D (2002) Locating multiple competitive facilities: spatial interaction models with variable expenditures. Ann Oper Res 111:197–225

Bierlaire M, Lotan T, Toint P (1997) On the overspecification of multinomial and nested logit models due to alternatives specific constants. Transp Sci 31:363–371

Canadian Institute for Health Information (2011) Wait times in Canada—a comparison by Province, 2011. Technical report, Canadian Institute for Health Information

Carling K, Haansson J (2013) A compelling argument for the gravity p-median model. Eur J Oper Res 226(3):658–660

CPLEX, IBM ILOG (2009) V12. 1: users manual for cplex. Int Bus Mach Corp 46(53):157

Drezner T (1994) Optimal continuous location of a retail facility, facility attractiveness, and market share: an interactive model. J Retail 70:49–64

Haase K (2009) Discrete location planning. Technical Report WP-09-07, Institute for Transport and Logistics Studies, University of Sydney

Haase K, Müller S (2014) A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. Eur J Oper Res 232(3):689–691

Haase K, Müller S (2013) Management of school locations allowing for free school choice. Omega 41(5):847–855

Health Council of Canada (2007) Wading through wait times. what do meaningful reductions and guarantees mean? Published online, http://www.healthcouncilcanada.ca/rpt_det.php?id=127. Accessed 12 Feb 2014

Louviere J, Hensher D, Swait J (2000) Stated choice methods: analysis and applications. Cambridge University Press, Cambridge

Marianov V, Rfos M, Icaza MJ (2008) Facility location for market capture when users rank facilities by shorter travel and waiting times. Eur J Oper Res 191(1):30–42

McCarl B, Meeraus A, van der Eijk P, Bussieck M, Dirkse S, Steacy P (2008) McCarl GAMS user guide. http://gams.com/docs/document.htm

McFadden D (1973) Conditional logit analysis of qualtitative choice behaviour. In: Zarembka P (ed) Frontiers of econometrics. Academic Press, New York, pp 105–142

McFadden D (2001) Economic choices. American economic review 91(3):351–378

Müller S, Haase K (2014) Customer segmentation in retail facility location planning. Business research. Accepted for publication

Müller S, Haase K, Kless S (2009) A multi-period school location planning approach with free school choice. Environ Plan A 41(12):2929–2945

Müller S, Haase K, Seidel F (2012) Exposing unobserved spatial similarity: evidence from German school choice data. Geogr Anal 44:65–86

Müller S, Tscharaktschiew S, Haase K (2008) Travel-to-school mode choice modelling and patterns of school choice in urban areas. J Transp Geogr 16(5):342–357

Perry N, Broeders M, de Wolf C, Toernberg S, Holland R, von Karsa L (2006) European guidelines for quality assurance in breast cancer screening and diagnosis. European communities. Health and consumer protection, 4th edn

Sener I, Pendyala R, Bhat C (2011) Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. J Transp Geogr 19(2):294–303

Street D, Burgess L (2007) The construction of optimal stated choice experiments. Wiley, New York

Tejada JJ, Ivy JS, King RE, Wilson JR, Ballan MJ, Kay MG, Diehl KM, Yankaskas BC (2014) Combined des/sd model of breast cancer screening for older women, ii: screening-and-treatment simulation. IIE Trans 46(7):707–727

Train K (2003) Discrete choice methods with simulation. MIT Press, Cambridge

Wait Time Alliance (2012) Shedding light on Canadians' total wait for care—report card on wait times in Canada. Technical report, Wait Time Alliance

Zhang Y, Berman O, Verter V (2012) The impact of client choice on preventive healthcare facility network design. OR Spectr 34:349–370

Business Research
DOI 10.1007/s40685-014-0008-6

ORIGINAL RESEARCH

Customer segmentation in retail facility location planning

Sven Müller · Knut Haase

AQ1 **Abstract** In this contribution, we discuss a facility location model to maximize firms' patronage, while demand is determined by a multinomial logit model (MNL). We account for customer segmentation based on customer characteristics. Hence, we are able to reduce the bias to the objective, which is due to constant substitution patterns of the MNL. Numerical studies show that averaging customer characteristics yield a bias of more than 15 % of the objective function value compared to segmentation. Using GAMS/CPLEX, we are able to solve problem sets with 2 segments, 500 demand points and 10 potential locations to optimality in 1 h computation time. If we consider 50 potential locations, the gap reported by CPLEX is <8 % in 1 h. We present an illustrative case example of a furniture store company in Germany (data are available as electronic supplementary material to this article). The corresponding problem is solved to optimality in a few minutes.

Keywords Multinomial logit model · Facility location · Maximum capture problem · Customer choice · Heterogeneous demand · Substitution patterns

JEL Classification C61 · C35 · C44 · R32

Electronic supplementary material The online version of this article (doi:10.1007/s40685-014-0008-6) contains supplementary material, which is available to authorized users.

Responsible editor: Sönke Albers.

S. Müller (✉) · K. Haase
Institute for Transport Economics, Universität Hamburg, Von-Melle-Park 5,
20146 Hamburg, Germany
e-mail: sven.mueller@uni-hamburg.de

K. Haase
e-mail: knut.haase@uni-hamburg.de

1 Introduction

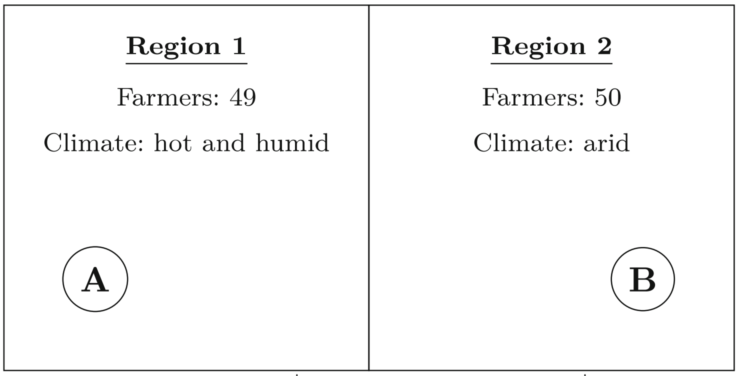
In this paper, we consider a situation where companies (retail store chains, for example) compete for their market share. Suppose for example that a firm wants to locate new shops in a geographical market. The decision variable under control is only where to locate the new facilities. The way customers make their choices is to be taken into account, too (Serra and Colome 2001). The reaction of possible competitors (price, locations) is not considered here.

We discuss a model—based on the maximum capture problem—for the optimal location of K facilities. Customers' choices are modeled according to a specific discrete choice model, namely the multinomial logit model (MNL). Other demand models (the Huff-model, for example) might be used instead of the MNL. Our approach is valid for such kind of models as well. However, we do not consider this here. In general, discrete choice models are the workhorse for the analysis of individual choice behavior (McFadden 1973, 2001). In literature, we find several applications of discrete choice models for spatial choice situations (Timmermans et al. 1992; Dellaert et al. 1998). Inspite of their long-term and widespread use, we find only few references in the operations research literature on facility location that account for discrete choice models. One reason may be the mathematical sophistication of the choice models. For example, de Palma et al. (1989), Benati (1999) and Marianov et al. (2008) discuss non-linear model formulations for discrete locational decisions. To the best of our knowledge, Benati and Hansen (2002) are the first who proposed a linear reformulation of the non-linear MNL. Their approach results in a hyperbolic sum integer problem. Haase (2009) uses constant substitution patterns of the MNL to find a linear integer reformulation. Aros-Vera et al. (2013) apply this approach to the planning of park-and-ride facilities. Finally, Zhang et al. (2012) propose an alternative approach similar to Benati and Hansen (2002). Haase and Müller (2014) show that a variant of the model of Haase (2009) seems to be superior to the formulations of Benati and Hansen (2002) and Zhang et al. (2012).

The MNL exhibits the well-known independence from irrelevant alternatives property (IIA). Roughly speaking, this property implies that each choice alternative (facility location) is an equal substitute to every other alternative. Unfortunately, it is empirically evidenced that this core property is unlikely to hold in spatial choice context (Bhat and Guo 2004; Hunt et al. 2004). The linear reformulations of the MNL already introduced in the literature are all based on the assumption that customers of a given demand point are homogenous in their observable characteristics (age and income, for example). In this contribution, we show that, if customers of a given demand point are portioned into homogenous subgroups according to their characteristics, the predictive bias due to the IIA might be reduced (Sect. 2). Of course, simply considering average characteristics are not sufficient as the following illustrative example shows (see Fig. 1).

Consider a country with only two regions (1 and 2) and a firm selling rice seeds to farmers. Farmers are assumed to bunker seeds at a facility of the firm. There are two potential facility locations A and B (there are no competitors). Region 1 contains location A and region 2 contains location B. Farmers located in region 1

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | ☐ LE | ☐ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK



Buying rice seeds	Region 1	Region 2	Total
Yes	49	0	49
No	0	50	50
Total	49	50	99

Fig. 1 The rice seed example

buy rice seeds only in A, while those of region 2 buy only in B. Region 1 contains 49 farmers and region 2 contains 50 farmers. Now assume that the climate in region 1 is hot and humid, while the climate of region 2 is arid (both regions might be separated by mountains). Since we expect that all of the farmers of region 1 buy rice seeds, but none of the farmers of region 2 would do so, we end up with a choice probability of buying rice seeds of 0.495 if we consider the population average. Now, assume the task of the firm is to select the facility location that maximizes the expected rice seed customers. Of course, we would select location B (in region 2), because $0.495.50 > 0.495.49$. However, the true sales are 0, because none of the farmers located in region 2 buys rice seeds, while the farmers of region 1 would only patronize a facility located in A. If the firm considers segment-specific choice probabilities instead (1 for farmers of region 1 and 0 for farmers of region 2), the optimal solution would be facility location A with an expected number of 49 rice seed customers. As a result, the expected bias, i.e., the relative deviation between the two solutions, is 100 %. Now, we learn from this example that simply considering average customer characteristics (instead of proper segmentation) may yield remarkably biased predictive outcomes. In other words, if customer characteristics are considered, it is advisable to employ segmentation instead of the averages of customer characteristics.

In this paper, we present an elucidating model formulation to account for customer segmentation within a mixed-integer program that enables to consider customer choice behavior by an MNL that accounts for customer characteristics (Sect. 3.1). Moreover, we present a simple lower bound and objective cuts for our problem (Sect. 3.2). We demonstrate the usefulness of our approach in extensive numerical studies (Appendix). Finally, we present an illustrative case example to show how our approach might be applied to support decision making for the management of a globally operating furniture store retail chain (Sect. 4).

97 2 A probabilistic choice model

98 Let us consider the following problem statement:

99 *Find K facility locations from all potential locations J such that the total*
100 *patronage for the K facilities is maximized.*

101 First, we define the sets

102 I demand nodes representing zones, like census blocks etc., that contain the customers,

103 M_i locations (existing and potential ones) from which the customers located in $i \in I$ choose exactly one location. M_i may include a no-choice-alternative, indicating that customers might not occupy any facility. Hence, the no-choice alternative (a dummy facility, for example) reflects the proportion of customers who do not consume (services or products) at any facility. We might consider a special case such that $M_i = M \; \forall \; i \in I$.

104 J potential locations for the facilities a decision maker (a firm, for example) has to decide on: $J \subseteq \bigcup_{i \in I} M_i$. Note $M_i \setminus J$ may include facility locations of competitors and/or the no-choice-alternative. That is, $\{M_i \setminus J\}$ comprises locations that are not influenceable by the decision maker. Further, $J_i = M_i \cap J$.

105 R_i is a set of choice alternatives faced by the customers of $i \in I$ that denotes the number, type, and/or the amount of purchases conducted by the customers. Hence, the choice set faced by customers located in $i \in I$ is $\{M_i \times R_i\}$. Consider exemplarily a customer located in a given demand node $i = 1$ who chooses to make a purchase of €10, €20, or €30 at any opened facility within a given time period. So $R_1 = \{10, 20, 30\}$. Let us further assume there are only two facilities, i.e., $M_1 = \{A, B\}$, then the choice set is $\{(A, 10), \ldots, (B, 10), \ldots, (B, 30)\}$. A choice of $(A, 20)$ means that the customer chooses to make a purchase of €20 at facility A. Note, the choice set must be exhaustive and the choice alternatives have to be mutually exclusive. Roughly speaking, all alternatives the customers actually face have to be included in the choice set. The generation of $\{M_i \times R_i\}$ is a sophisticated issue. We refer to Swait (2001) for further details.

106
107 We consider the parameters

108 h_i number of customers located in node $i \in I$, and

109 v_{ijr} as the deterministic utility of customers located in $i \in I$ patronizing $j \in M_i$ making a purchase denoted by $r \in R_i$. This could be a measure of generalized cost etc.

110 K number of facilities to be located, with $0 < K < |J|$.

111
112 Further, we define the binary decision variable

113 y_j = 1, if location $j \in J$ provides a facility (0, otherwise), and

114
115 the non-negative variable

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | □ LE | □ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK

x_{ijr} as the choice probability of customers of node $i \in I$ who makes a purchase denoted by $r \in R_i$ at a facility located at $j \in J_i$. If we assume that the choice probability is given by the MNL, x_{ijr} is defined as

$$x_{ijr} = \frac{e^{v_{ijr}} y_j}{\sum_{o\in R_i}\left(\sum_{m\in M_i\setminus J} e^{v_{imo}} + \sum_{m\in J_i} e^{v_{imo}} y_m\right)} \quad \forall\, i \in I, j \in J_i, r \in R_i. \tag{1}$$

Note, if $M_i \setminus J \neq \emptyset$, then $\sum_{j\in J_i}\sum_{r\in R_i} x_{ijr} < 1$ for all $i \in I$. Now the problem can be modeled as a mixed-integer non-linear program:

$$\text{Maximize} \sum_{i\in I}\sum_{r\in R_i}\sum_{j\in J_i} f(i,r,j)x_{ijr} \tag{2}$$

subject to (1) and

$$\sum_{j\in J} y_j = K \tag{3}$$

$$y_j \in \{0,1\} \quad \forall\, j \in J. \tag{4}$$

Demand is determined by $f(i,r,j)x_{ijr}$ with $f(i,r,j)$ as a function denoting the consumption. We denote F as the objective function value of (2). In literature, we find exact linear reformulations of (1) such that (2)–(4) can be modeled as a mixed-integer program: Haase (2009) and Aros-Vera et al. (2013) employ specific properties of the MNL, while Zhang et al. (2012) propose an approach based on variable substitution similar to Benati and Hansen (2002). In Sect. 3, we present a modified reformulation of Haase (2009). At first, we focus on important properties of (1) in the following subsequent sections.

We assume in the following that $|R_i| = 1\ \forall\, i \in I$ simplifying v_{ijr}, (1), and (2) for convenience reasons. Of course, all formulations of the subsequent sections are valid for $|R_i| > 1\ \forall\, i \in I$ as well.

2.1 The independence from irrelevant alternatives property

The IIA property is well known in discrete (locational) choice literature (Ray 1973; Sheppard 1978; McFadden 2001; Sener et al. 2011). One outcome of the IIA is that the ratio of choice probabilities of two alternatives (i.e., facility locations) remains constant no matter whether other alternatives are available or not (*constant substitution pattern*). That is, the probability of patronizing a facility located in j relative to a facility located in m is independent of the existence and attributes of any other facility. Consider two arbitrary but existing facility locations $j, m \in M_i$ to be given. Then, according to (1), the ratio of the choice probabilities x_{ij} and x_{im} is

$$\frac{x_{ij}}{x_{im}} = \frac{\mathrm{e}^{\mathrm{v_{ij}}}}{\mathrm{e}^{\mathrm{v_{im}}}} = \mathrm{e}^{\mathrm{v_{ij}}-\mathrm{v_{im}}} \quad \forall\, \mathrm{i} \in \mathrm{I}. \tag{5}$$

The IIA property of (5) implies that a new facility or change in the attractiveness of an existing facility other than m or j will draw patronage from competing facilities in direct proportion to their choice probabilities. In contrast, in applications, it is extremely unlikely that this property holds (Haynes and Fotheringham 1990; Müller et al. 2012; Hunt et al. 2004). In situations when the IIA property is not valid we should consider discrete choice models other than MNL (mixed logit or nested logit, for example). See Train (2009) for further reading. Müller et al. (2009), Haase (2009) and Haase and Müller (2013b) propose approximate approaches that are able to incorporate a large class of discrete choice models into mathematical programs.

2.2 Aggregation issues

The MNL and hence (1) is based on the theory of utility maximization behavior of individuals. That is, each individual chooses the location that maximizes its utility. Given our problem statement of Sect. 2 and the corresponding model (2)–(4), we are interested in aggregate measures (market shares, total patronage etc.) instead of individual choice probabilities. Data on customer demand are usually given as an aggregate measure (number of customers, for example). Now, the question arises how we should compute the choice probability of all customers (individuals) located in a given demand point $i \in I$? The answer depends on the specification of the utility v_{ij}. If v_{ij} does not contain characteristics of the customers (age, income, and so forth) then the choice probability x_{ij} applies to all customers in $i \in I$ in the same way and thus, (2) is a proper formulation. In contrast, the incorporation of customer characteristics in v_{ij} will improve the accurateness of x_{ij} (Koppelman and Bhat 2006, pp 21–23 and pp 41–46). However, aggregation is more tedious in such a case.

Example 1 For simplicity reasons, we consider only one demand node $i = i'$. Consider $J = M_{i'} = \{A, B, C\}$. Further, we assume i' contains two customers $n \in \{1, 2\}$. Let the deterministic utility function for customer n be given as

$$v_{nj} = -g_{i'j}/q_n \quad \forall\, j \in J, \tag{6}$$

with $g_{i'j}$ as the cost for a trip from i' to j and q_n is the income of customer n. The higher the income the lesser the impact of travel cost (Casado and Ferrer 2013). Now, there are basically two ways of computing $x_{i'j}$:

1. we use the average income of $n = 1$ and $n = 2$ (i.e., the average income of demand node i') denoted by $\overline{q}_{i'} = (q_1 + q_2)/2$ to compute $\overline{v}_{i'j}$ and thus $\overline{x}_{i'j}$, or
2. we first compute the choice probabilities for each customer x_{nj} and then we determine the average choice probability of customers located in i' as $\tilde{x}_{i'j} = (x_{n=1,j} + x_{n=2,j})/2$.

In general, (1) is expected to be inaccurate compared to (2) because of the non-linear relationship between $x_{i'j}$ and $v_{i'j}$ in (1). Consider the values given in Table 1. As expected, $\overline{x}_{i'j}$ determined by (1) and $\tilde{x}_{i'j}$ determined by (2) are different. As

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | □ LE | □ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK

Table 1 Aggregation, choice probabilities and the IIA property

	$j = A$	$j = B$	$j = C$	$x_{i'A}/x_{i'C}$
$g_{i'j}$	1	2	4	
q_1	9	9	9	
q_2	1	1	1	
$\bar{q}_{i'}$	5	5	5	
Solution I: y_j	1	1	1	
$\bar{x}_{i'j}$ using $\bar{q}_{i'}$	0.422	0.346	0.232	**1.822**
$x_{n=1,j}$	0.383	0.343	0.274	1.396
$x_{n=2,j}$	0.705	0.259	0.035	20.085
$\tilde{x}_{i'j}$	0.544	0.301	0.155	**3.516**
Solution II: y_j	1	0	1	
AQ1 $\bar{x}_{i'j}$ using $\bar{q}_{i'}$	0.646	0	0.354	**1.822**
$x_{n=1,j}$	0.583	0	0.417	1.396
$x_{n=2,j}$	0.953	0	0.047	20.085
$\tilde{x}_{i'j}$	0.768	0	0.232	**3.302**

Of course, income q_n as a characteristic of the customer is constant over alternatives. The choice probabilities are computed using (1) and (6). The last column contains the ratio of choice probabilities of facility locations A and C according to (5). We consider two solutions (i.e., I and II) to problem (2)–(4)

shown by Train (2009), pp 29–32 (2) should be preferred. In addition, we observe an interesting pattern if we apply customer characteristics in an appropriate way: the ratio of the average choice probabilities $\tilde{x}_{i'A}/\tilde{x}_{i'C}$ depends on the existence of facility location B (non-constant substitution pattern). Although the IIA property does apply to each customer n, it does not apply to the population of i' as a whole. The key point is that there are two distinct segments of the population (high and low income) with different choice probabilities: We compare to different solutions to (2)–(4), namely solution I (all locations are selected) and solution II (location B is not selected). The customer with low income ($n = 2$) considers location A to be a better substitute to B than C. In contrast, for customer $n = 1$ (high income), locations A and C are more or less equal substitutes to location B. This pattern is due to the different evaluation of travel cost by the two segments (i.e., customers).

There are two lessons learned so far: First, the more customer characteristics are included in v_{ij} in an appropriate way, the better are the forecast properties of MNL, x_{ij}, respectively. Second, by applying segmentation to our model (2)–(4) as outlined in (2), we are able to reduce the bias of x_{ij} and F due to the IIA of (5) to some extent. In applications, one would be interested in how to classify customers, and how

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | □ LE | □ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK

many customer segments are appropriate for a given application. Of course, segmentation makes sense only if the deterministic part of utility contains factors that vary over choice makers. Usually, such factors are socio-economic factors like age, gender, income, occupation, car ownership, and so forth. In empirical studies, socio-economic factors that are continuous measures (age and income, for example) are usually considered as categorical measures. For example, a proband is asked whether his/her age is (a) below 20 years, (b) between 20 and 40 years, (c) between 40 and 60 years, or (d) older than 60 years. Now consider a deterministic utility function with only two socio-economic factors: gender and age. Gender, of course, consists of only two categories: female and male. So, we end up with eight customer segments: the four age levels for each of the two genders. Considering many socio-economic factors with many levels yields a large number of segments. How many segments are appropriate and tractable could not be said in the abstract. It rather depends on the application, in particular, the empirically specified choice model. See Ben-Akiva and Lerman (1985), pp 131–153 for a detailed discussion of aggregation and segmentation.

3 A probabilistic choice model with customer segmentation

In Sect. 2, we have demonstrated that the IIA may yield biased values of x_{ij} of (1) and hence a biased objective function value F of (2). Moreover, a partition of the population of a demand point $i \in I$ into homogenous sub-populations (i.e., segmentation) enables us to reduce the bias due to the IIA. In this section, we propose how to explicitly account for segments of customers (heterogeneous customer demand) in a linear mixed-integer model formulation of (2)–(4).

3.1 Mathematical formulation

In addition to the definitions of Sect. 2, we consider the set

S_i segments of the customers located in demand node $i \in I$; for example high and low income or male and female or a combination of income and gender.

Next, we denote the parameters

$\tilde{h}_{is}$ number of customers according to segment $s \in S_i$ located in node $i \in I$,

$\tilde{v}_{isj}$ as the deterministic utility of customers of segment $s \in S_i$ located in $i \in I$ patronizing $j \in M_i$,

π_{isj} choice probability of customers of segment $s \in S_i$ at node $i \in I$ who access service at a facility located at $j \in J_i$ given that all $m \in J$ are established, i.e., $\pi_{isj} = \mathrm{e}^{\tilde{v}_{isj}} / \sum_{m \in M_i} \mathrm{e}^{\tilde{v}_{ism}}$,

φ_{isj} choice probability of customers of segment $s \in S_i$ at node $i \in I$ who access service at a facility located at $j \in J_i$ given that $j \in J_i$ is the only facility location established, i.e., $\varphi_{isj} = \mathrm{e}^{\tilde{v}_{isj}} / (\mathrm{e}^{\tilde{v}_{isj}} + \sum_{m \in M_i \setminus J} \mathrm{e}^{\tilde{v}_{ism}})$, and

Journal : Small-ext 40685	Dispatch : 8-7-2014	Pages : 27
Article No. : 8	□ LE	□ TYPESET
MS Code : mueller_haase	☑ CP	☑ DISK

ζ_{is} cumulative choice probability of customers of segment $s \in S_i$ at node $i \in I$ who access service at competing facilities given that all potential facilities $j \in J$ are located, i.e., $\zeta_{is} = \sum_{l \in M_i \setminus J} (e^{\widetilde{v}_{isl}} / \sum_{m \in M_i} e^{\widetilde{v}_{ism}})$. Therefore,

$$\zeta_{is} + \sum_{j \in J_i} \pi_{isj} = 1 \quad \forall\, i \in I, s \in S_i.$$

Finally, we define the non-negative variables

$\widetilde{x}_{isj}$ as the MNL choice probability of customers of segment $s \in S_i$ at node $i \in I$ who access service at a facility located at $j \in J_i$, and

z_{is} as the cumulative choice probability of customers of segment $s \in S_i$ at node $i \in I$ who do not access any facility of the considered firm.

Then, our model according to the problem statement of Sect. 2 is

$$\text{maximize} \sum_{i \in I} \sum_{s \in S_i} \widetilde{h}_{is} \sum_{j \in J_i} \widetilde{x}_{isj} \tag{7}$$

subject to

$$z_{is} + \sum_{j \in J_i} \widetilde{x}_{isj} = 1 \quad \forall\, i \in I, s \in S_i \tag{8}$$

$$\widetilde{x}_{isj} - \varphi_{isj} y_j \leq 0 \quad \forall\, i \in I, s \in S_i, j \in J_i \tag{9}$$

$$\widetilde{x}_{isj} - \pi_{isj} y_j \geq 0 \quad \forall\, i \in I, s \in S_i, j \in J_i \tag{10}$$

$$\widetilde{x}_{isj} - \frac{\pi_{isj}}{\zeta_{is}} z_{is} \leq 0 \quad \forall\, i \in I, s \in S_i, j \in J_i \tag{11}$$

$$\sum_{j \in J} y_j = K \tag{12}$$

$$\widetilde{x}_{isj} \geq 0 \quad \forall\, i \in I, s \in S_i, j \in J_i \tag{13}$$

$$z_{is} \geq 0 \quad \forall i \in I, s \in S_i \tag{14}$$

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | □ LE | □ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK

$$y_j \in \{0,1\} \quad \forall\, j \in J. \tag{15}$$

We denote $\widetilde{F}$ as the objective function value of (7). Let be given a combination of $i \in I$, $s \in S_i$, and $j \in J_i$. For convenience reasons, we assume for a moment that $|M| = 2$ and $|J| = 1$ with $M = \{j,k\}$ and $J = \{j\}$, accordingly $M_i = M$ and $J_i = J$. Now, if $y_j = 0$, then $\widetilde{x}_{isj} = 0$ because of (9) and further $z_{is} = 1$ because of (8). If $y_j = 1$, then according to (11), $\widetilde{x}_{isj} = z_{is} \cdot \pi_{isj}/\zeta_{is}$, because of (7) and $z_{is} \cdot \pi_{isj}/\zeta_{is} \le \varphi_{isj}$. Due to (8) and substitution, we get the correct choice probabilities $\widetilde{x}_{isj} = \mathrm{e}^{\widetilde{V}_{isj}} / \left(\mathrm{e}^{\widetilde{V}_{isj}} + \mathrm{e}^{\widetilde{V}_{isk}}\right)$ with k indicating the facility location of the competitor. Of course, these coherences are valid for $|M| > 2$ and $|J| > 1$ as well. Therefore, constraints (8)–(11) together with (7) yield the MNL choice probabilities. For more details, we refer to Haase (2009) and Aros-Vera et al. (2013). Using φ_{isj} in (9) and π_{isj} in (10) yields bounds on $\widetilde{x}_{isj}$ that are tighter than simply using $0 \le \widetilde{x}_{isj} \le y_j$. In contrast to Aros-Vera et al. (2013), we do not consider redundant constraints in our model: Using (11) yields $|I| \cdot |S_i| \cdot |J_i|$ constraints instead of $|I| \cdot |S_i| \cdot |J_i|^2$ constraints.

3.2 Lower bound and objective cuts

To derive an intelligible lower bound for $\widetilde{F}$ of (7), we consider the binary variable w_{mj}. Further, we define the non-negative variable

$$Q = \sum_{i\in I} \sum_{s\in S_i} \sum_{m\in J} \sum_{j\in J\setminus\{m\}} \pi_{isj} w_{mj}. \tag{16}$$

If we minimize Q subject to (16) and

$$\sum_{j\in J\setminus\{m\}} w_{mj} = K-1 \quad \forall\, m \in J \tag{17}$$

$$w_{mj} \in \{0,1\} \quad \forall\, m,j \in J \tag{18}$$

the quantity

$$a_m = \sum_{i\in I} \sum_{s\in S_i} \widetilde{h}_{is} \frac{\pi_{ism}}{\sum_{j\in J} \pi_{isj} w^*_{mj}} \quad \forall\, m \in J, \tag{19}$$

denotes the maximum attractiveness of facility location $m \in J$ with w^*_{mj} indicating that j belongs to the $K-1$ most attractive facility locations compared to m. If we maximize Q subject to (16)–(18), then the quantity

$$b_m = \sum_{i\in I} \sum_{s\in S_i} \widetilde{h}_{is} \frac{\pi_{ism}}{\sum_{j\in J} \pi_{isj} w^*_{mj}} \quad \forall\, m \in J, \tag{20}$$

denotes the minimum attractiveness of facility location $m \in J$ with w^*_{mj} indicating that j belongs to the $K-1$ least attractive facility locations compared to m. To

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | □ LE | □ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK

derive a lower bound, we choose the K-largest $j \in J$ according to b_j. Denote this set as $\tilde{J}$. Accordingly, $\tilde{J}_i = \tilde{J} \cap M_i$. Now compute the lower bound as:

$$LB = \sum_{i \in I} \sum_{s \in S_i} \tilde{h}_{is} \sum_{j \in \tilde{J}_i} \frac{e^{\tilde{v}_{isj}}}{\sum_{m \in M_i \setminus J} e^{\tilde{v}_{ism}} + \sum_{m \in \tilde{J}_i} e^{\tilde{v}_{ism}}}. \tag{21}$$

Finally, we add

$$\tilde{F} \geq LB \tag{22}$$

$$\tilde{F} \leq LB - \sum_{j \in \tilde{J}} b_j (1 - y_j) + \sum_{j \in J \setminus \tilde{J}} a_j y_j \tag{23}$$

to our model (7)–(15) to account for a lower bound (LB) (22) and an objective cut OC1 (23). A lower bound for problems with capacities is presented in Haase and Müller (2013a). Now, we might define the quantities

$$\alpha_j = -LB + \sum_{i \in I} \sum_{s \in S_i} \tilde{h}_{is} \sum_{l \in \tilde{J}_i \cup \{j\}} \frac{e^{\tilde{v}_{isl}}}{\sum_{m \in M_i \setminus J} e^{\tilde{v}_{ism}} + \sum_{m \in \tilde{J}_i \cup \{j\}} e^{\tilde{v}_{ism}}} \quad \forall j \in J \setminus \tilde{J} \tag{24}$$

and

$$\gamma_j = \sum_{i \in I} \sum_{s \in S_i} \tilde{h}_{is} \left(\sum_{l \in J \setminus \{j\}} \frac{e^{\tilde{v}_{isl}}}{\sum_{m \in M_i \setminus J} e^{\tilde{v}_{ism}} + \sum_{m \in J \setminus \{j\}} e^{\tilde{v}_{ism}}} - \sum_{l \in J} \pi_{isj} \right) \quad \forall j \in \tilde{J}. \tag{25}$$

Based on Benati and Hansen (2002), we can define a second objective cut OC2 alternatively to (23)

$$\tilde{F} \leq LB + \sum_{j \in J \setminus \tilde{J}} \alpha_j y_j + \sum_{j \in \tilde{J}} \gamma_j (1 - y_j). \tag{26}$$

Note, γ_j in (25) is negative for all $j \in \tilde{J}$ by construction.

We are interested in the impact of the number of segments, the lower bound, the objective cuts, the number of competitors on the solution and the solvability of our approach. The corresponding numerical examples can be found in the Appendix. The major findings of these numerical examples are that (1) segmentation has significant impact on the computational effort, (2) the lower bound (22) provides a quite good solution (it deviates <1 % from the optimal solution), and (3) the use of the objective cut OC1 (23) is particularly appealing if we do not expect to find an optimal solution within a given time. Further, we solve problem sets with 2 segments, 500 demand points and 10 potential locations to optimality in 1 h computation time. If we consider 50 potential locations, the gap reported by CPLEX is <8 % in 1 h.

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | □ LE | □ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK

4 Illustrative case example: furniture store location in Germany

In this section, we apply our model of Sect. 3.1 to a hypothetical—but still realistic— branch-extension of a large furniture store company in Germany. Figure 2a shows the already existing facility locations and the potential facility locations of the considered firm, as well as the locations of the main competitors in the market. The firm already runs 46 stores in the year 2012 with a market share of 12.5 % and 46 million customers yielding 3.7 billion Euro revenue. The firm aims to massively expand in the market in the near future. It is intended to establish 5–15 new facilities until 2020. The task is to find out the optimal locations for a given number of new facilities (K^+) from 50 potential facility locations and the corresponding expected market share of the firm.

We consider the centroids of the 415 German "Kreise" (municipalities) as demand points. The locations of the facilities (existing, potential, and competitors) are given by longitude and latitude coordinates. The euclidean distance in kilometers between a demand point $i \in I$ and a facility location $j \in M$ is denoted by d_{ij}. The choice set for each demand node $i \in I$ is defined by

$$M_i = \left\{ j \in M \middle| d_{ij} \leq \delta \right\} \tag{27}$$

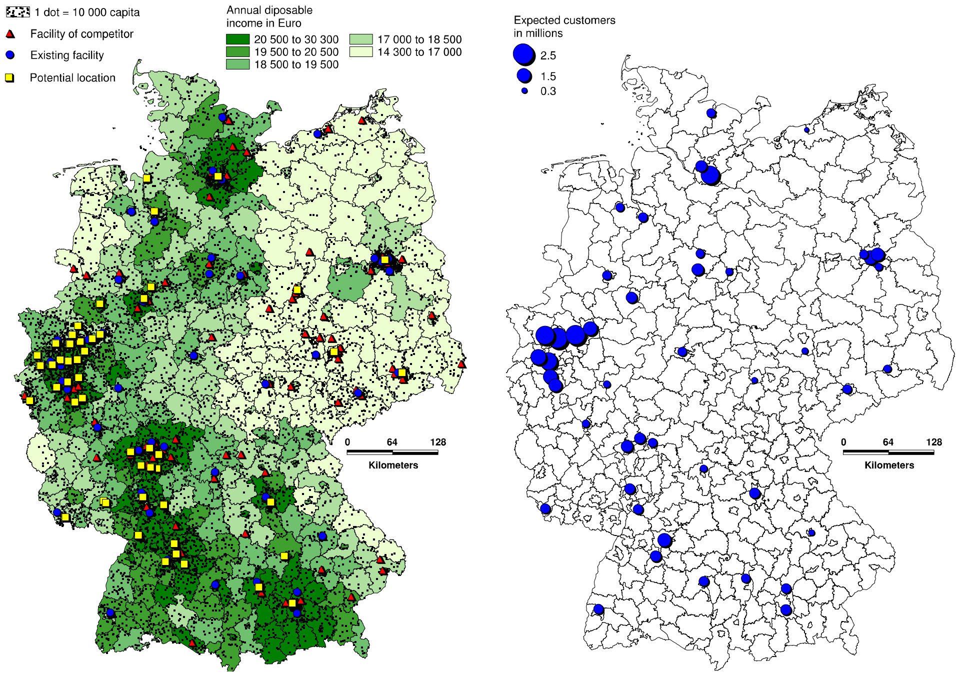
with M as the set of all facility locations and δ as a threshold distance. If $\pi_{isj} < 0.00001$ then we remove j from M_i. There exist 101 facility locations of the competitors. Thus, $|M| = 197$. Customers do not consider facilities located more distant than δ as a conceivable alternative. Since, the main customers of the firm are aged between 15 and 25, we consider two distinct segments of customers: $\widetilde{h}_{i,s=1}$ as the number of customers aged between 15 and 25 and $\widetilde{h}_{i,s=2}$ as the number of customers of all other ages. The deterministic part of utility (see Sects. 2, 3.1) is given as

$$\widetilde{v}_{is,j=0} = \beta^{\mathrm{inc}} \cdot \mathrm{INC_i} \quad \forall\, \mathrm{i} \in \mathrm{I}, \mathrm{s} \in \mathrm{S}, \tag{28}$$

$$\widetilde{v}_{isj} = \beta_s^{\mathrm{dist}} \cdot d_{ij} \quad \forall i \in I, s \in S, j \in M_i, j > 0, \tag{29}$$

with $\mathrm{INC_i}$ as the average annual disposable income of the population located in $i \in I$ in 1,000 Euro. Total population and $\mathrm{INC_i}$ are given in Fig. 2a. The ratio $\sum_i \widetilde{h}_{i,s=1} / \sum_i \widetilde{h}_{i,s=2} = 0.163$. Coefficients β^{inc} and β_s^{dist} are the utility contribution per unit of the corresponding attribute (distance and income). Equation 28 denotes the utility for not choosing any of the facility locations of the firm (potential and existing) or the competitors. Roughly speaking, $j = 0$ denotes a dummy facility absorbing all demand not satisfied by the facilities of the firm or the competitor. The dummy facility $j = 0$ comprises the utility of customers either to patronize a small, local furniture store or not to consume furniture anyway. Note that (28) and (29) are rather simplistic specifications of utility to make the application more comprehensible.

Journal : Small-ext 40685 Dispatch : 8-7-2014 Pages : 27
Article No. : 8 □ LE □ TYPESET
MS Code : mueller_haase ☑ CP ☑ DISK



(a) Data and Locations **(b)** Base Scenario

Fig. 2 Existing facilities and potential locations of the firm as well as facilities of the competitors (**a**). Expected customers in base scenario (**b**)

In a real-world application, the coefficients $\beta^{\text{inc}}, \beta_s^{\text{dist}}$ of (28) and (29) have to estimated using empirical choice data (i.e., discrete choice analysis). Large companies can easily afford a comprehensive empirical study to appropriately estimate the coefficients of the utility functions. Here, we cannot obtain such estimates, hence we rely on parameter estimates from other empirical studies. Suarez et al. (2004) provide coefficient estimates for a shopping center choice model. They distinguish between two different segments of customers (target group and others) and estimate coefficients of the distance between the customers location and the shopping center for both customer segments. Here, we employ these coefficient estimates, given as

$$\beta_{s=1}^{\text{dist}} = -0.078$$
$$\beta_{s=2}^{\text{dist}} = -0.088.$$

This indicates that the main customers ($s = 1$, population aged between 15 and 25) are less sensitive to distance than other customers. Goldman (1976) provides empirical evidence on the coherence between income and the propensity of shopping at a specific facility. Based on Fotheringham and Trew (1993), we might consider

$$\beta^{\text{inc}} = -0.015.$$

Now, we are able to compute the expected patronage for each existing facility using (21) and hence the total expected market share of the firm as

Springer

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | □ LE | □ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK

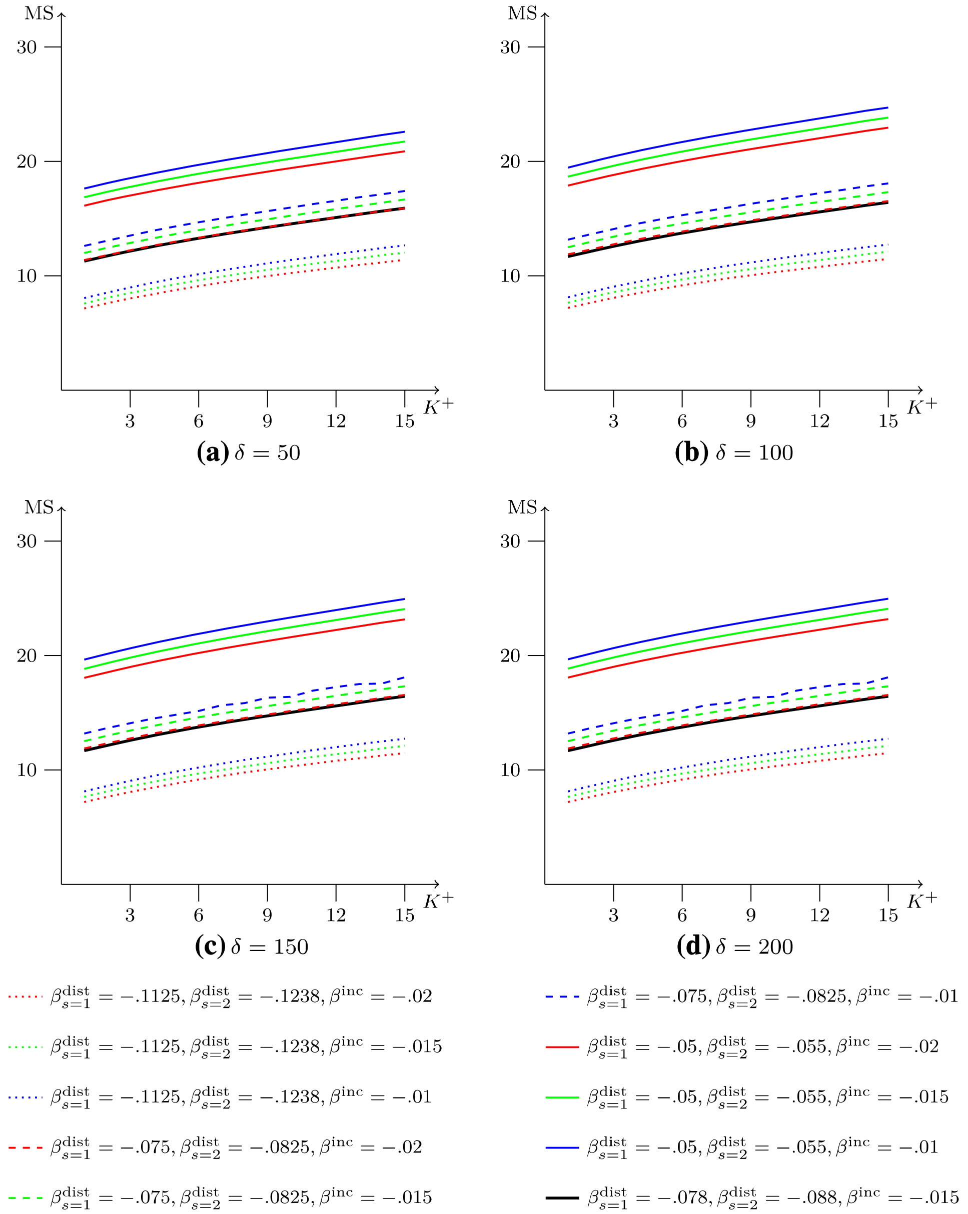
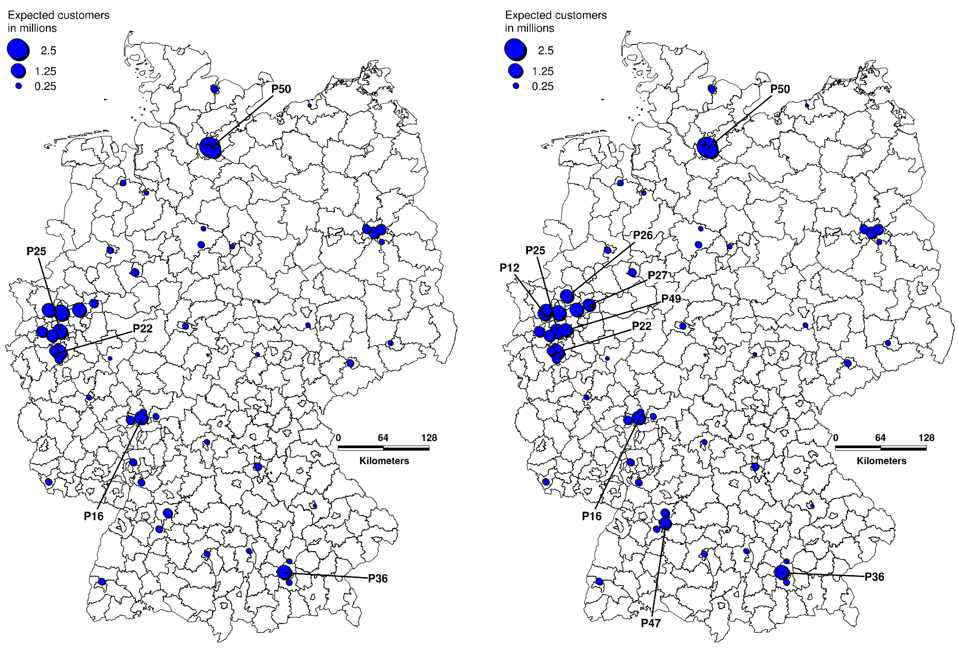


Fig. 3 Results of sensitivity analysis for $\delta, \beta^{\text{dist}}_s, \beta^{\text{inc}}$, and market share (MS). K of (12) is given by $46 + K^+$ (46 facilities are already in the market)

$$\text{MS} = \widetilde{\text{F}} / \sum_{i \in I} \sum_{s \in S_i} \widetilde{h}_{is}. \tag{30}$$

We consider this as the base scenario. Figure 2b displays the result. We know that, on average, a customer is assumed to make five shopping visits a year. This yields 41 million customers over all existing facilities and a total expected market share of 11.12 %. The expected market share is below the reference value of 12.5 %. This is



(a) $K = 51$, i.e. $K^{+} = 5$ **(b)** $K = 56$, i.e. $K^{+} = 10$

AQ1 **Fig. 4** Results for two possible scenarios. Parameter settings used: $\delta = 150$, $\beta^{\text{dist}}_{s=1} = -0.078$, $\beta^{\text{dist}}_{s=2} = -0.088$, and $\beta^{\text{inc}} = -0.015$. The newly located facilities are labeled, f.e. P16, P22, P25, P36, and P50 in **a**—where "P" denotes that the corresponding location is selected from the set of potential locations. Non-labeled locations are already existing facilities

reasonable, because we do not consider online purchases and there might be some inconsistencies close to the border of Germany due to transnational purchases of customers. On average, a customer spends 80 Euro per visit yielding an annual revenue of 3.28 billion Euro. This is close to the reference value of 3.7 billion Euro. We conclude that our demand model makes predictions fairly well.

Since our parameters do not stem from a unique study on furniture store customer behavior in Germany, we first investigate the sensitivity of the solution to parameter variations. The locational decision variables y_j are fixed to one for the already existing facility locations (i.e., $j < 47$). We solve our model of Sect. 3 for various parameter settings and for different distance thresholds δ of (27). We are interested in MS's dependence on K^{+}. We have implemented our model in GAMS 23.7 and we use CPLEX 12.2 on a 64-bit Windows Server 2008 with 4 Intel Xeon 2.4 GHz processors and 24 GB RAM for all studies. All problems considered in this section are solved to optimality within minutes. The results of Fig. 3 show a piecewise linear increase of the market share in K^{+}. The slope is nearly 0.35 indicating that with each additional facility, the total market share of the firm increases by 0.35 % points. Note, the underlying function is not necessarily concave. The sensitivity analysis indicates that the market share is independent from the distance threshold $\delta > 50$ and the weight of the income β^{inc}. In contrast, the scale of the market share

Journal : Small-ext 40685 Dispatch : 8-7-2014 Pages : 27
Article No. : 8 □ LE □ TYPESET
MS Code : mueller_haase ☑ CP ☑ DISK

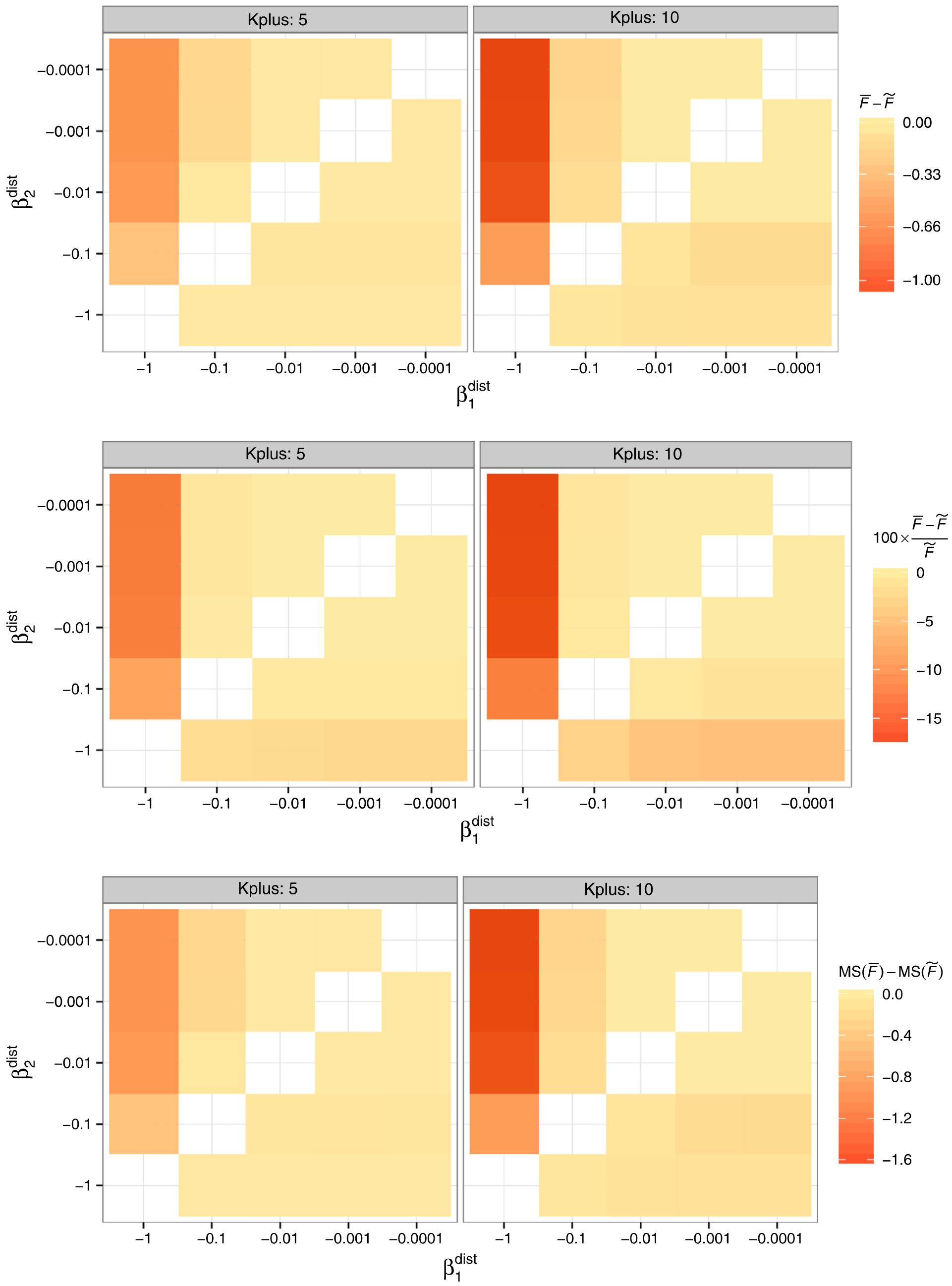


Fig. 5 Results of Example 2: bias of objective function due to neglected segmentation. "Kplus" corresponds to K^+. The values of $\overline{F} - \widetilde{F}$ are given in million customers. The numerical values corresponding to this figure are given in Appendix (see Table 2)

heavily depends on the distance parameters (β_s^{dist}). This finding stresses the need for firms to employ the estimates based on unique choice studies (Street and Burgess 2007; Müller et al. 2008; Louviere et al. 2000 for how to design studies and experiments for discrete choice analysis).

Table 2 Results of Example 4

K	$\lvert M \setminus J \rvert$	Market share	CPU
5	5	0.525	65.028
	10	0.353	40.164
	15	0.269	33.077
10	5	0.683	730.747
	10	0.520	190.007
	15	0.415	131.345
15	5	0.762	115.467
	10	0.606	88.412
	15	0.512	76.692

For each problem set, we have computed ten instances. The numbers given are the averages over ten instances. CPU denotes the time used by CPLEX. All instances are solved to optimality. $|I| = 100$ and $|J| = 20$

Based on the (linear) relationship between MS and K^+, the firm's management is enabled to identify a specific number of new facilities to be located. The optimal locations and the expected (annual) patronage of the new facilities can be displayed in maps and enhance the decision making of the firm's management. Figure 4 exemplifies a market expansion with 5 and 10 new facilities. In a real-world management application, one usually has to account for locally varying locational (and maybe operational) cost. In such a situation, one would be interested in the relationship between cost (or budget) and market share. The firm is further interested in the impact of segmentation of their customers (see Sect. 2.2). Therefore, we consider the following example that extends Example 1.

Example 2 We expect the more the two segments differ, the larger is the predictive bias of the MNL and thus the larger is the bias of the objective function value if segmentation is neglected. Due to the specification of the deterministic part of utility in (29), the difference in choice probabilities between the two segments corresponds to the difference between $\beta^{\text{dist}}_{s=1}$ and $\beta^{\text{dist}}_{s=2}$.To evaluate the impact of neglected segmentation, we first consider $\beta^{\text{dist}}_{s=1} = \beta^{\text{dist}}_{s=2} = \beta^{\text{dist}}$ with $\beta^{\text{dist}} = (\beta^{\text{dist}}_{s=1} + \beta^{\text{dist}}_{s=2})/2$ in (29). This corresponds to a simple average of utilities as described in (1) of Sect. 2.2. The corresponding solution in terms of selected locations is denoted by $\bar{J} = \left\{ j \in J \middle| y^*_j = 1 \right\}$. Based on $\bar{J}$, we compute the MNL choice probabilities using segmentation, i.e., we use $\beta^{\text{dist}}_{s=1}$ and $\beta^{\text{dist}}_{s=2}$ instead of β^{dist} in (29). The corresponding objective function value is denoted as $\bar{F}$ and the corresponding market share is given by $\text{MS}(\bar{F})$.

We consider $\beta^{\text{dist}}_s \in \{-1, -0.1, -0.01, -0.001, -0.0001\}$, $\beta^{\text{inc}} = -0.015$, and $\delta = 150$. Further, we consider two scenarios: $K^+ = 5$ and $K^+ = 10$. The results are given in Fig. 5. The patterns for the total deviation $\bar{F} - \tilde{F}$, relative deviation $100 \times (\bar{F} - \tilde{F})/\bar{F}$, and the deviation of the market shares $\text{MS}(\bar{\text{F}}) - \text{MS}(\tilde{\text{F}})$ are similar. The most eye-catching bias occurs if $\beta^{\text{dist}}_{s=1} = -1$. Consider exemplarily

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | □ LE | □ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK

$\beta^{\text{dist}}_{s=1} = -1$ and $\beta^{\text{dist}}_{s=2} = -0.1$, i.e., segment $s = 1$ evaluates each additional kilometer ten times as negative as segment $s = 2$ (i.e., $\beta^{\text{dist}}_{s=1}/\beta^{\text{dist}}_{s=2} = 10$). In case that segmentation is neglected, the corresponding distance-coefficient is $\beta^{\text{dist}} = -0.55$. As a consequence, a large part of customers (recall that, $\sum_i \tilde{h}_{i,s=1}/\sum_i \tilde{h}_{i,s=2} = 0.163$) evaluates distance more than five times as negative as this would be the case with segmentation. Of course, the corresponding deviation is remarkable (-8.9 % for $K^+ = 5$ and -12.5 % for $K^+ = 10$). The asymmetric pattern in Fig. 5 is due to the uneven distribution of population over the two segments (the population of segment 2 is larger than the population of segment 1): the more the true coefficient of the large part of the population (segment 2) deviates from the average coefficient the larger is the expected predictive error. In contrast, a large deviation of the true coefficient of segment 1 has impact only on a small part of the population and the corresponding expected predictive error is comparably small. Obviously, the extent of the error heavily depends on the scale of the coefficients. Consider, for example, $\beta^{\text{dist}}_{s=1} = -1$ and $\beta^{\text{dist}}_{s=2} = -0.1$. The corresponding ratio is 10 and the expected error for $K^+ = 5$ is -8.88 %. Now, for $\beta^{\text{dist}}_{s=1} = -0.1$ and $\beta^{\text{dist}}_{s=2} = -0.01$ the corresponding ratio is 10 again. However, the corresponding error is only -0.18 %. This pattern is due to the non-linear relationship between distance (deterministic utility) and the choice probabilities (i.e., a s-shaped probability function). As the coefficients (weighting of travel distance) get larger (i.e., approaching 0) the probabilities of choosing to patronize a facility approach the largest possible value. For these values of the deterministic utility the difference in the corresponding choice probabilities between the two segments become small.

The bias found in our study is comparable to those reported in studies on spatial aggregation (Andersson et al. 1998; Daskin et al. 1989; Current and Schilling 1987; Murray and Gottsegen 1997). In literature, ratios of segment-specific coefficients larger than 50 are reported (Müller et al. 2012; Koppelman and Bhat 2006, pp 133–134). However, the difference between segment-specific distance-coefficients used in our application is small. We have considered parameter settings that yield a ratio $\beta^{\text{dist}}_{s=1}/\beta^{\text{dist}}_{s=2} = 0.91$ (see Fig. 3). As a consequence, the expected bias is below 1 % if we neglect segmentation in our application. Nevertheless, the consideration of segments yields valuable insights, because the utility function (29) and the corresponding coefficients are arbitrarily chosen. As stated before, for a real application, the company is expected to specify utility functions and estimate the corresponding coefficients on unique choice data. The firm may use such a numerical study to make assumptions about worst-case scenarios.

5 Summary

By an intelligible example, we demonstrate that the independence from IIA of the MNL may yield false predictions. This finding is well founded on empirical studies. When the MNL is used in a mathematical program to incorporate customer choice

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | ☐ LE | ☐ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK

behavior, the model outcomes are very likely to be biased as well. Although the MNL is founded on individual choice behavior, in facility location planning we are interested in the share of customers of a demand point patronizing a certain facility. If we assume the customers of a demand point are homogenous, i.e., they exhibit the same observable characteristics, then there is no need for segmentation. If we assume the customers to be heterogeneous then segmentation of the customers according to their characteristics (income and age, for example) should be employed. By proper segmentation, we are able to reduce the predictive bias of the MNL in terms of market shares.

In this contribution, we present a model formulation for the maximum capture problem that explicitly allows for customer segmentation using the MNL to find optimal shopping facility locations. Moreover, we propose an intelligible approach to derive a lower bound for our model. Extensive computational studies show the impact of proper segmentation as well as the efficiency of our approach: using aggregate customer characteristics instead of proper segmentation may yield a predictive bias of the objective function value of more than 15 % deviation from the optimal objective function value. Our lower bound is found in <1 s and deviates <1 % from the optimal solution. Problems with 2 segments, 50 potential locations and 500 demand points can be solved to a gap <8 % within 1 h using GAMS/CPLEX. Based on our numerical studies concerning the quality of the lower bound, it is reasonable to assume that the true gap is remarkably smaller than 8 %. We apply our approach in an illustrative case example of a globally operating furniture store company that intends to increase its market share in Germany by branch expansion. This problem can be solved to optimality within few minutes. Our example shows how the novel approach can be used for management decision support.

Based on our findings, several possible directions of future research appear. It is of interest to find analytically bounds on the bias of the objective function value due to missing segmentation under various segmentation patterns and specifications of utility. Further, the explicit consideration of substitution patterns, i.e., correlation between facility locations, is a very important issue to be analyzed. Efficient solution methods are necessary to account for larger problem sets. Finally, our approach is useful to other areas of operations research; assortment optimization, for example Kök and Fisher (2007).

Acknowledgments The very helpful comments and suggestions of three anonymous reviewers are gratefully acknowledged. They made significant contributions in order to improve the paper. We further thank the editors for helpful suggestions. Finally, we thank Sonja Bröning for copy editing. However, the responsibility for any remaining error is with the authors.

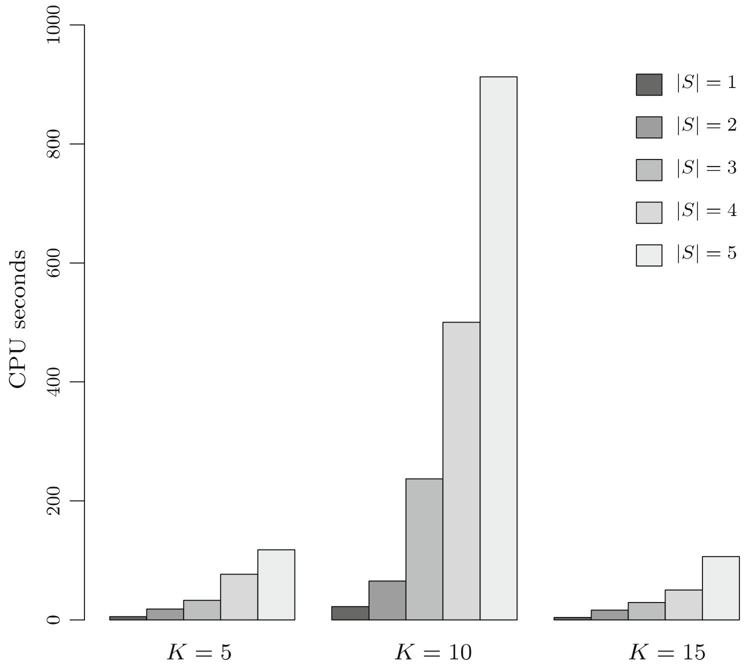


Fig. 6 Example 3: $|M \setminus J| = K$, $|I| = 50$ and $|J| = 20$. For each problem set, we consider ten randomly generated instances. The values are the averages over ten instances

Appendix

In this section, we provide numerical examples to validate and test the mathematical formulation of Sect. 3. We assume $M_i = M$, $S_i = S$ and $J_i = J\ \forall\ i \in I$. For given I, M, and J, we generate longitude and latitude coordinates using a random uniform distribution in the interval $[0, 100]$. We set the maximum computational time to 1 h if not stated otherwise. Further, we assume that demand is completely satisfied, i.e., a no-choice alternative does not exist. To generate the demand $\widetilde{h}_{is}$, we first generate a population $\mathrm{Pop_i}$ for each demand node $i \in I$ using a random uniform distribution in the interval $[0, 10]$ weighted by the ratio $|M|/|I|$. Further, we generate weights $\omega_{is}\ \forall\ i \in I, s \in S$ using a random uniform distribution in the interval $[0, 1]$. Then, $\widetilde{h}_{is} = \mathrm{Pop_i} \cdot \omega_{\mathrm{is}} / \sum_{\mathrm{s' \in S}} \omega_{\mathrm{is'}}$. Let the utility function be

$$\widetilde{v}_{isj} = \beta_s t_{ij} \quad \forall\ i \in I, s \in S, j \in M, \tag{31}$$

with t_{ij} as the travel-time between $i \in I$ and $j \in J$; computed as the rectangular distance between $i \in I$ and $j \in J$ divided by 60. All other parameters of Sect. 3 can be easily derived. In the following, we consider several numerical examples to test our mathematical formulation.

Example 3 In this study, we are interested in the additional burden due to the number of segments. We set $|I| = 50$, $|J| = 20$, $K = \{5, 10, 15\}$, $|S| = \{1, 2, \ldots, 5\}$, $|M \setminus J| = K$, and β_s from a uniform distribution in the interval $[-2, -0.5]$. Note, in

Table 3 Results of examining the effect of the lower bound and the objective cuts (Example 5)

$\|J\|$	K	Model 1		Model 2				Model 3		Model 4		Model 5	
		CPU	GAP	CPU	CPU-LB	GAP	GAP-LB	CPU	GAP	CPU	GAP	CPU	GAP
20	5	922.207	0.000	47.263	0.067	0	0.356	22.333	0.00	25.180	0.00	24.999	0.00
	10	3,203.147	7.178	124.311	0.051	0	0.012	185.112	0.00	167.089	0.00	193.737	0.00
30	5	1,836.594	0.000	44.848	0.031	0	0.120	109.710	0.00	122.300	0.00	130.276	0.00
	10	3,600.000	25.319	1,932.590	0.030	0	0.022	3,258.599	1.64	3,279.984	1.52	3,278.786	1.63

Numbers given are averages over ten randomly generated instances. CPU denotes the time in seconds used by CPLEX. CPU-LB denotes the time in seconds used by CPLEX to determine the lower bound. GAP is the gap reported by CPLEX in percent. GAP-LB denotes the deviation of the lower bound LB from the optimal solution in percent. We consider $|I| = 50$, $|M \setminus J| = K$. Over all instances, CPLEX found the optimal solution

Model 1 (7)–(15)

Model 2 (7)–(15), (22), i.e., incl. lower bound

Model 3 (7)–(15), (22), (23), i.e., incl. lower bound and objective cut OC1

Model 4 (7)–(15), (22), (26), i.e., incl. lower bound and objective cut OC2

Model 5 (7)–(15), (22), (23), (26), i.e., incl. lower bound and objective cuts OC1 and OC2

Journal : Small-ext 40685 Dispatch : 8-7-2014 Pages : 27
Article No. : 8 ☐ LE ☐ TYPESET
MS Code : mueller_haase ☑ CP ☑ DISK

applications the number of segments will be small due to data availability. See Ben-Akiva and Lerman (1985), pp 148–150 for an illustrative case study. For each problem set K and S, we compute ten randomly generated instances. Figure 6 displays the results. We observe that the computational effort increases with the number of segments. Seemingly, it depends on the ratio $K/|J|$ how fast the computational effort increases in $|S|$. If only a few locations have to be selected ($K = 5$) or many locations have to be selected ($K = 15$), the computational effort is small compared to the situation where 50 % of the potential locations have to be selected ($K = 10$).

Example 4 In this example, we investigate the impact of the number of competing facility locations $|M \setminus J|$ and the number of facilities to be located K. We consider $|I| = 100$, $|J| = 20$, $|S| = 2$, $\beta_{s=1} = -1$, and $\beta_{s=2} = -0.5$ for nine different problem sets with ten instances each. The results are given in Table 3. The market share of the considered firm declines in the number of competing facilities. The smaller K the more the decline of the market share in the number of competitors (nearly 50 % decline for $K = 5$ compared to somewhat more than 30 % for $K = 15$). If the number of established facilities and the number of competing facilities are equal, then market shares are nearly the same (especially, if many facilities are established). This study confirms the findings of Example 3 concerning the ratio $K/|J|$ and the corresponding computational effort. Further, the study shows an interesting pattern: there seems to be a positive relationship between the market share and the computational effort (the larger the market share the more CPU time is needed).

Example 5 Now we are interested in the efficiency of the lower bound described in Sect. 3.2. We consider four problem sets with $|J| = \{20, 30\}$ and $K = \{5, 10\}$. For each problem set, 10 randomly instances are generated. Further, we set $|I| = 50$, $|M \setminus J| = K$, $|S| = 2$, $\beta_{s=1} = -1$, and $\beta_{s=2} = -0.5$. For each instance, we solve our model with and without the lower bound (22) and with and without the OC1 (23) and OC2 (26). Table 4 displays the results. For all instances, CPLEX found the optimal solution within 1 h computational time. However, for larger problem sets ($K > 5$), we are able to prove optimality within 1 h only if we use the lower bound LB. We are able to decrease the computational effort remarkably (at least 20 times faster) using LB. The lower bound is found in <1 s and LB deviates <1 % from the optimal solution. In a small numerical example Benati and Hansen (2002) show that they find the optimal solution to their problem by variable neighborhood search in <1 s for problem sets up to $|J| = 50$ and $K < 10$. Concerning the objective cuts OC1 (23) and OC2 (26), we observe a benefit only for small problem sets ($|J| = 20$, $K = 5$). Unfortunately, for larger problem sets, the computational effort increases (up to 2.5 times slower). Possibly, this is due to a degeneration of the LP relaxation using the objective cuts. This finding is confirmed by the results of Benati and

Journal : Small-ext 40685 | Dispatch : 8-7-2014 | Pages : 27
Article No. : 8 | □ LE | □ TYPESET
MS Code : mueller_haase | ☑ CP | ☑ DISK

Table 4 Results of Example 6

$\lvert I \rvert$	$\lvert J \rvert$	LB and OC1 Equations (22) and (23) CPU	 GAP	LB Equation (22) CPU	 GAP
100	10	5.70	0.00	4.60	0.00
	25	2,931.03	0.74	2,377.71	0.19
	50	3,600.00	5.88	3,600.00	7.20
				7,200.00	7.10
250	10	27.84	0.00	20.89	0.00
	25	3,600.00	3.65	3,600.00	3.70
				7200.00	3.51
	50	3,600.00	7.34	3,600.00	10.78
				7,200.00	10.21
500	10	104.34	0.00	62.72	0.00
	25	3,600.00	5.96	3,600.00	6.41
				7,200.00	5.75
	50	3,600.00	7.61	3,600.00	13.23
				7,200.00	13.15

We consider model (7)–(15), (22). For each problem set, we have computed ten random instances. The numbers are the averages over ten instances. CPU denotes the time in seconds used by CPLEX (maximum computation time 1 or 2 h). GAP denotes the solution gap in percent provided by CPLEX. We consider $|S| = 2$, $K = |J|/2$ and $|M \setminus J| = K$

Hansen (2002). They report that their upper bound based on submodular maximization—which is comparable to our objective cuts—performs not as good as the bound provided by concave relaxation. In our study, we find no remarkable difference in performance between OC1 and OC2.

Example 6 The objective of this numerical example is to figure out up to what problem size we are able to solve our problem to (or close to) optimality. We consider $|I| \in \{100, 250, 500\}$, $|J| \in \{10, 25, 50\}$, $K = \lceil |J|/2 \rceil$, $|M \setminus J| = \lceil 2/3 \cdot |J| \rceil$, $\beta_{s=1} = -1$, and $\beta_{s=2} = -0.5$. For each of the nine problem sets, we solve ten AQ1 instances. The results are given in Table 5. Small-sized problem sets ($|J| = 10$) can be easily solved to optimality. Medium-sized problem sets ($|J| = 25$) can be solved up to a gap of <6 % in 1 h. For large problem sets ($|J| = 50$), the gap becomes disappointing if we only use the lower bound (22). In contrast, if we use the lower bound (22) and the OC1 (23), we are able to reduce the gap to somewhat more than 7 % within 1 h. Taking into account the good quality of the lower bound (see Example 5) and the observation that most of the time is needed to prove optimality, we may assume that the "true" gap is even smaller. Note, Benati and Hansen (2002) made the same observation.

Journal : Small-ext 40685 Dispatch : 8-7-2014 Pages : 27
Article No. : 8 □ LE □ TYPESET
MS Code : mueller_haase ☑ CP ☑ DISK

Table 5 Data of Example 2 in Fig. 5

In.	K^+	β_1^{dist}	β_2^{dist}	β^{dist}	$\overline{F}$	$\text{MS}(\overline{F})$	$\widetilde{F}$	$\text{MS}(\widetilde{F})$	Δ_1	Δ_2	Δ_3
1	5	−1	−0.1	−0.55	3.50	5.23	3.84	5.74	−0.34	−8.88	−0.51
2	10	−1	−0.1	−0.55	4.19	6.26	4.79	7.16	−0.60	−12.52	−0.90
3	5	−1	−0.01	−0.50	4.39	6.55	5.02	7.49	−0.63	−12.57	−0.94
4	10	−1	−0.01	−0.50	5.11	7.64	6.15	9.18	−1.03	−16.82	−1.54
5	5	−1	−0.001	−0.50	4.47	6.68	5.14	7.67	−0.66	−12.88	−0.99
6	10	−1	−0.001	−0.50	5.20	7.77	6.29	9.40	−1.09	−17.29	−1.63
7	5	−1	−0.0001	−0.50	4.48	6.70	5.15	7.69	−0.67	−12.94	−0.99
8	10	−1	−0.0001	−0.50	5.21	7.78	6.31	9.42	−1.09	−17.34	−1.64
9	5	−0.1	−1	−0.55	1.02	1.53	1.04	1.56	−0.02	−2.07	−0.03
10	10	−0.1	−1	−0.55	1.31	1.96	1.36	2.03	−0.05	−3.38	−0.07
11	5	−0.1	−0.01	−0.06	21.15	31.58	21.19	31.64	−0.04	−0.18	−0.06
12	10	−0.1	−0.01	−0.06	22.44	33.52	22.57	33.71	−0.13	−0.57	−0.19
13	5	−0.1	−0.001	−0.05	21.92	32.74	22.09	32.99	−0.16	−0.75	−0.25
14	10	−0.1	−0.001	−0.05	23.31	34.82	23.50	35.09	−0.18	−0.78	−0.27
15	5	−0.1	−0.0001	−0.05	21.98	32.82	22.15	33.09	−0.18	−0.81	−0.27
16	10	−0.1	−0.0001	−0.05	23.37	34.90	23.57	35.20	−0.20	−0.85	−0.30
17	5	−0.01	−1	−0.50	1.16	1.73	1.19	1.78	−0.03	−2.51	−0.05
18	10	−0.01	−1	−0.50	1.46	2.18	1.54	2.30	−0.08	−5.35	−0.12
19	5	−0.01	−0.1	−0.06	9.88	14.75	9.93	14.83	−0.05	−0.54	−0.08
20	10	−0.01	−0.1	−0.06	10.97	16.39	11.05	16.50	−0.07	−0.66	−0.11
21	5	−0.01	−0.001	−0.01	24.50	36.59	24.51	36.61	−0.01	−0.06	−0.02
22	10	−0.01	−0.001	−0.01	26.06	38.91	26.06	38.92	−0.00	−0.02	−0.01
23	5	−0.01	−0.0001	−0.01	24.53	36.64	24.55	36.66	−0.02	−0.07	−0.02
24	10	−0.01	−0.0001	−0.01	26.10	38.97	26.10	38.98	−0.00	−0.01	−0.01

Journal : Small-ext 40685 Dispatch : 8-7-2014 Pages : 27
Article No. : 8 □ LE □ TYPESET
MS Code : mueller_haase ☑ CP ☑ DISK

Table 5 continued

In.	K^+	β_1^{dist}	β_2^{dist}	β^{dist}	$\overline{F}$	$MS(\overline{F})$	$\widetilde{F}$	$MS(\widetilde{F})$	Δ_1	Δ_2	Δ_3
25	5	−0.001	−1	−0.50	1.18	1.75	1.21	1.80	−0.03	−2.72	−0.05
26	10	−0.001	−1	−0.50	1.47	2.20	1.56	2.33	−0.09	−5.54	−0.13
27	5	−0.001	−0.1	−0.05	10.01	14.94	10.06	15.03	−0.05	−0.54	−0.09
28	10	−0.001	−0.1	−0.05	11.04	16.49	11.18	16.70	−0.14	−1.29	−0.21
29	5	−0.001	−0.01	−0.01	23.94	35.75	23.95	35.77	−0.01	−0.04	−0.02
30	10	−0.001	−0.01	−0.01	25.47	38.04	25.48	38.05	−0.01	−0.03	−0.01
31	5	−0.001	−0.0001	−0.00	24.66	36.83	24.68	36.86	−0.02	−0.07	−0.03
32	10	−0.001	−0.0001	−0.00	26.23	39.17	26.24	39.18	−0.01	−0.03	−0.01
33	5	−0.0001	−1	−0.50	1.18	1.76	1.21	1.81	−0.03	−2.74	−0.05
34	10	−0.0001	−1	−0.50	1.47	2.20	1.56	2.33	−0.09	−5.56	−0.13
35	5	−0.0001	−0.1	−0.05	10.02	14.96	10.07	15.04	−0.05	−0.54	−0.08
36	10	−0.0001	−0.1	−0.05	11.05	16.50	11.19	16.72	−0.14	−1.29	−0.22
37	5	−0.0001	−0.01	−0.01	23.94	35.76	23.96	35.78	−0.01	−0.05	−0.02
38	10	−0.0001	−0.01	−0.01	25.48	38.05	25.48	38.05	−0.00	−0.01	0.00
39	5	−0.0001	−0.001	−0.00	24.63	36.79	24.65	36.81	−0.02	−0.07	−0.02
40	10	−0.0001	−0.001	−0.00	26.19	39.12	26.19	39.12	0.00	0.00	0.00

Deviations are denoted by $\Delta_1 = \overline{F} - \widetilde{F}, \Delta_2 = 100(\frac{\overline{F}-\widetilde{F}}{\widetilde{F}})$, and $\Delta_3 = MS(\overline{F}) - MS(\widetilde{F})$

In. denotes the instance

Journal : Small-ext 40685
Article No. : 8
MS Code : mueller_haase
Dispatch : 8-7-2014
□ LE
☑ CP
Pages : 27
□ TYPESET
☑ DISK

References

Andersson, G., R.L. Francis, T. Normark, and B.M. Rayco. 1998. Aggregation method experimentation for large-scale network location problems. *Location Science* 6(4): 25–39.

Aros-Vera, F., V. Marianov, and J.E. Mitchell. 2013. p-Hub approach for the optimal park-and-ride facility location problem. *European Journal of Operational Research* 226(2): 277–285.

Ben-Akiva, M., and S. Lerman. 1985. *Discrete choice analysis, theory and applications to travel demand*. Cambridge: MIT Press.

Benati, S. 1999. The maximum capture problem with heterogeneous customers. *Computers and Operations Research* 26(14): 1351–1367.

Benati, S., and P. Hansen. 2002. The maximum capture problem with random utilities: problem formulation and algorithms. *European Journal of Operational Research* 143(3): 518–530.

Bhat, C.R., and J. Guo. 2004. A mixed spatially correlated logit model: formulation and application to residential choice modeling. *Transportation Research Part B Methodological* 38(2): 147–168.

Casado, E., and J.C. Ferrer. 2013. Consumer price sensitivity in the retail industry: latitude of acceptance with heterogeneous demand. *European Journal of Operational Research* 228(2): 418–426.

Current, J.R., and D.A. Schilling. 1987. Elimination of source a and b errors in p-median location problems. *Geographical Analysis* 19(2): 95–110.

Daskin, M., A. Haghani, M. Khanal, and C. Malandraki. 1989. Aggregation effects in maximum covering models. *Annals of Operations Research* 18(1): 113–139.

Dellaert, B.G., T.A. Arentze, M. Bierlaire, A.W. Borgers, and H.J. Timmermans. 1998. Investigating consumers' tendency to combine multiple shopping purposes and destinations. *Journal of Marketing Research* 35(2): 177–188.

Fotheringham, S., and R. Trew. 1993. Chain image and store-choice modeling: the effects of income and race. *Environment and Planning A* 25(2): 179–196.

Goldman, A. 1976. Do lower-income consumers have a more restricted shopping scope? *Journal of Marketing* 40(1): 46–54.

Haase, K. 2009. *Discrete location planning*. Tech. Rep. WP-09-07. Institute for Transport and Logistics Studies, University of Sydney.

Haase, K., and S. Müller. 2013a. *Insights into clients' choice on preventive health care facility location planning*. Working paper, University of Hamburg.

Haase, K., and S. Müller. 2013b. Management of school locations allowing for free school choice. *Omega* 41(5): 847–855.

Haase, K., and S. Müller. 2014. A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. *European Journal of Operational Research* 232(3): 689–691.

Haynes, K., and S. Fotheringham. 1990. The impact of space on the application of discrete choice models. *The Review of Regional Studies* 20(1): 39–49.

Hunt, L., B. Boots, and P. Kanaroglou. 2004. Spatial choice modelling: new opportunities to incorporate space into substitution patterns. *Progress in Human Geography* 28(6): 746–766.

Kök, G., and M.L. Fisher. 2007. Demand estimation and assortment optimization under substitution: methodology and application. *Operations Research* 55(6): 1001–1021.

Koppelman, F.S., and C. Bhat. 2006. *A self instructing course in mode choice modeling: multinomial and nested logit models*. Prepared for US Department of Transportation Federal Transit Administration.

Louviere, J., D. Hensher, and J. Swait. 2000. *Stated choice methods : analysis and applications*. Cambridge: Cambridge University Press.

Marianov, V., M. Rfos, and M.J. Icaza. 2008. Facility location for market capture when users rank facilities by shorter travel and waiting times. *European Journal of Operational Research* 191(1): 30–42.

McFadden, D. 1973. Conditional logit analysis of qualtitative choice behaviour. In *Frontiers of econometrics*, ed. P. Zarembka, 105–142. New York: Academic Press.

McFadden, D. 2001. Economic choices. *American Economic Review* 91(3): 351–378.

Müller, S., S. Tscharaktschiew, and K. Haase. 2008. Travel-to-school mode choice modelling and patterns of school choice in urban areas. *Journal of Transport Geography* 16(5): 342–357.

Müller, S., K. Haase, and S. Kless. 2009. A multi-period school location planning approach with free school choice. *Environment and Planning A* 41(12): 2929–2945.

Journal : **Small-ext 40685**	Dispatch : **8-7-2014**	Pages : **27**
Article No. : **8**	□ LE	□ TYPESET
MS Code : **mueller_haase**	☑ CP	☑ DISK

Müller, S., K. Haase, and F. Seidel. 2012. Exposing unobserved spatial similarity: evidence from German school choice data. *Geographical Analysis* 44(1): 65–86.

Murray, A.T., and J.M. Gottsegen. 1997. The influence of data aggregation on the stability of p-median location model solutions. *Geographical Analysis* 29(3): 200–213.

de Palma, A., V. Ginsburgh, M. Labbe, and J.F. Thisse. 1989. Competitive location with random utilities. *Transportation Science* 23(4): 244–252.

Ray, P. 1973. Independence of irrelevant alternatives. *Econometrica* 41(5): 987–991.

Sener, I., R. Pendyala, and C. Bhat. 2011. Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. *Journal of Transport Geography* 19(2): 294–303.

Serra, D., and R. Colome. 2001. Consumer choice and optimal locations models: formulations and heuristics. *Papers in Regional Science* 80(4): 439–464.

Sheppard, E. 1978. Theoretical underpinnings of the gravity hypothesis. *Geographical Analysis* 10(4): 386–402.

Street, D., and L. Burgess. 2007. *The construction of optimal stated choice experiments*. New York: Wiley.

Suarez, A., I.R. del Bosque, J.M. Rodriguez-Poo, and I. Moral. 2004. Accounting for heterogeneity in shopping centre choice models. *Journal of Retailing and Consumer Services* 11(2): 119–129.

Swait, J. 2001. Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B Methodological* 35(7): 643–666.

Timmermans, H., A. Borgers, and P. van der Waerden. 1992. Mother logit analysis of substitution effects in consumer shopping destination choice. *Journal of Business Research* 23(2): 311–323.

Train, E. Kenneth. 2009. *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.

Zhang, Y., O. Berman, and V. Verter. 2012. The impact of client choice on preventive healthcare facility network design. *OR Spectrum* 34(2): 349–370.

Journal : **Small-ext 40685**	Dispatch : **8-7-2014**	Pages : **27**
Article No. : **8**	□ LE	□ TYPESET
MS Code : **mueller_haase**	☑ CP	☑ DISK

Omega 41 (2013) 847–855

Contents lists available at SciVerse ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Management of school locations allowing for free school choice

Knut Haase, Sven Müller*

Institute for Transport Economics, University of Hamburg, D-20146 Hamburg, Germany

A R T I C L E I N F O

Article history:
Received 11 May 2012
Accepted 22 October 2012
Processed by B. Lev
Available online 23 November 2012

Keywords:
Discrete location
Simulation-based optimization
Endogenous demand
Random utility model
Mixed logit
School network
0–1 program

A B S T R A C T

Nearly without exception, we find in literature (school) location models with exogenously given demand. Indeed, we know from a large number of empirical studies that this assumption is unrealistic. Therefore, we propose a discrete location model for school network planning with free school choice that is based on simulated utility values for a large average sample. The objective is to maximize the standardized expected utility of all students taking into account capacity constraints and a given budget for the school network. The utility values of each student for the schools are derived from a random utility model (RUM). The proposed approach is general in terms of the RUM used. Moreover, we do not have to make assumptions about the functional form of the demand function. Our approach, which combines econometric and mathematical methods, is a linear 0–1 program although we consider endogenous demand by a highly non-linear function. The proposed program enables practicing managers to consider student demand adequately within their decision making. By a numerical investigation we show that this approach enables us to solve instances of real size optimally – or at least close to optimality – within few minutes using GAMS/Cplex.

1. Introduction

German students are free to choose the secondary school to enroll at. That means enrollment is not mandatory as determined by the location of students. Although the literature mostly ignores free school choice [24], it is evident that there is an emerging need for free school choice in countries other than Germany [10,16]. The application of a qualified student might be refused only if the school capacity will be exceeded. In Germany this is rather not the case due to the decrease in the number of students in the last decade in many regions of Germany. In particular in eastern Germany, the demographic processes force the local authorities to close less demanded schools [30]. For an appropriate planning of school locations it is crucial to determine the students' demand very concisely. Therefore, approaches using proxies like distance or travel-time in order to determine demand are less reasonable (see [2–4,11,37] for example).

But how can we measure the demand for school locations more appropriately? For a given set of schools – we call this the *choice set* – the school choice probabilities of students can be easily computed using a random utility model [32]. Note that other approaches, such as the Huff-Model [17,18] and the Multiplicative Competitive Interaction Model [34], yield a probabilistic student allocation as well. However, a disaggregate approach, like a random utility model (RUM), has several advantages (in transferability and efficiency) towards an aggregated approach (gravity models). For more details on this issue see [8,21,42]. At first glance it seems that using RUM – although more appropriate – comes at heavy additional expenses. This is not true. School choice data (revealed preferences) should be easily available at authorities or the schools themselves. If not, one can conduct a stated choice experiment (SP) on a small sample of students in order to get a sufficiently rich data set. Additionally, SP can be used in order to estimate the demand of schools not established yet [41]. One has to be aware of scaling issues using SP data – particularly in combination with revealed preference data – though. Nowadays, there exist tailored methods to use both data in demand models. For more details we refer to [7,44].

Now, school location planners have appropriate school choice data at hand. How can they exploit this data in order to make better decisions about school locations? Subject of a location model is the selection of a subset of all locations under study (or rather potential locations) with respect to an optimization criterion. In general, the elements of the selected subset are the locations to be closed or opened, respectively. In school location models the selected subset – or rather the solution – depends heavily on the demand for the locations [38,1,28]. Today in location modeling literature, the demand for the locations is exogenously given. Though, this is an unrealistic assumption. Several empirical studies give evidence that the demand for a certain location might depend on the existence of other locations

* Corresponding author. Tel.: +49 40 42838 9021.
E-mail addresses: knut.haase@wiso.uni-hamburg.de (K. Haase), sven.mueller@wiso.uni-hamburg.de (S. Müller).

0305-0483/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.omega.2012.10.008

due to substitution effects [8,10,14,20,22,23,32]. Thus, the demand depends on the solution of the location model and demand is therefore endogenous. If we consider the demand as exogenously given, this might yield wrong locational decisions. To be able to use the results of a RUM analysis, Müller et al. [31] introduce a scenario-based multi-period school location model. For a certain period, a scenario defines which schools should be open and which should be closed. As for each scenario the choice set is given, it can be evaluated regarding cost and expected demand. This approach can be applied if the number of schools is small. The number of scenarios grows exponentially with the number of schools. Despite this reference there exists sparse literature on integrating RUM – even simple multinomial logit models – in mathematical programs for location planning (see [25,40] for example). However, these approaches employ difficult and complex formulations. Moreover, heuristics are needed to obtain a solution. In this paper, we propose an approach that combines econometric and mathematical methods in order to tackle these issues. Therefore, our approach takes forward the literature due to

1. a very general formulation allowing for an arbitrarily close approximation of any RUM to be taken into account - and no assumptions about the functional form of the demand function have to be made,
2. an intelligible model formulation which is (optimally) solvable within few minutes using a standard MIP-solver, and
3. a relaxation from geographical scaling issues[1] inherent to standard location models.

Since we do not expect all readers to be familiar with the theory on RUM, we give a brief introduction to RUM in the next section. In Section 3 we describe how the demand determined by a RUM is incorporated in a mathematical program. Therefore, we start with a very intuitive but highly non-linear formulation (non-linear 0–1 program). This is followed by a new linear 0–1 program. In Section 4 we describe the data for our numerical investigation (Section 4.3). We conclude with a summary in Section 5.

2. Demand determined by random utility models

In order to understand the way of model building in Section 3 it is necessary to get a basic understanding of how demand can be determined by RUM. Of course, space is limited and we refer to [5,21,43] for more details on RUM and discrete choice analysis. The following brief explanation is based on Train [43] if not mentioned otherwise.

2.1. Theory on random utility models

Let us consider a student i who obtains from alternative (school) s an utility u_{is}. Student i chooses school s from a finite and exhaustive choice set of schools $\mathcal{C}_i$ only if $u_{is} > u_{il}\ \forall l \neq s$. The alternatives $s \in \mathcal{C}_i$ have to be mutually exclusive. In order to operationalize a *representative utility* v_{is} we use n observable attributes of the schools as faced by the student and observable attributes of the students, labeled w_{isn}

$$v_{is} = \sum_n \beta_{sn} w_{isn}. \tag{1}$$

Obviously v_{is} depends on coefficients β_{sn} that are unknown to the researcher and therefore estimated statistically. However, this dependence is suppressed for the moment. Since there are aspects of utility that the researcher does not or cannot observe $v_{is} \neq u_{is}$. Therefore, utility is decomposed as

$$u_{is} = v_{is} + \epsilon_{is}, \tag{2}$$

where ϵ_{is} captures the factors that affect utility but are not included in v_{is}. This decomposition is fully general, since ϵ_{is} is defined as simply the difference between true utility u_{is} and the part of utility that the researcher captures in v_{is}. Given its definition, the characteristics of ϵ_{is}, such as its distribution, depend critically on the researcher's specification of v_{is}. In particular, ϵ_{is} is not defined for a choice situation per se. Rather, it is defined relative to a researcher's representation of that choice situation. This distinction becomes relevant when evaluating the appropriateness of various specific RUM. Since ϵ_{is} of (2) is random the probability that student i chooses school s is

$$P_i(s|\mathcal{C}_i) = \text{Prob}(u_{is} > u_{il} |\ \forall l \in \mathcal{C}_i). \tag{3}$$

McFadden [26] shows that if we assume that ϵ_{is} is independent and identically extreme value distributed (iid EV) over s, then we get by some algebraic transformation

$$P_i(s|\mathcal{C}_i) = \frac{e^{v_{is}}}{\sum_{l \in \mathcal{C}_i} e^{v_{il}}}. \tag{4}$$

Because of the closed-form choice probabilities the multinomial logit (MNL) model in (4) has been the workhorse for discrete choice analysis for decades. The coefficients β_{sn} of (1) are estimated by a maximum-likelihood procedure. Today, standard software packages like Biogeme [6] are available for estimation of coefficients and computation of choice probabilities. The critical part of the assumption is that the unobserved factors are uncorrelated over schools, as well as having the same variance for all schools. This leads to the prominent "independence of irrelevant alternatives" (IIA) property of the MNL. This property implies a constant ratio of choice probabilities of two alternatives (constant substitution). Though, the assumption of independence can be inappropriate in some situations. Unobserved factors related to one school might be similar to those related to another school. The IIA property ensures that the ratio of choice probabilities for any two alternatives (schools) is unaffected by the presence or change of any other alternative and its attributes. Therefore, a change in the probability of one alternative will lead to identical changes in relative choice probabilities for all other alternatives. While it is an empirical question and a matter of the specification of v_{is} whether IIA holds for a given data set, it is known that IIA is unlikely to hold in spatial choice applications. For example, Haynes and Fotheringham [15] note that size, aggregation, dimensionality, spatial continuity and variation and location characteristics of spatial choice data are likely to produce substitution patterns that violate IIA. For this reason other models like the nested logit or mixed multinomial logit (MMNL) model that rely on different assumptions about ϵ_{is} have been developed. The MMNL is particularly appealing due to its property to approximate *any* RUM arbitrarily close [27]. The MMNL can be operationalized by decomposing the error term ϵ_{is} of (2) into

$$\epsilon_{is} = \sum_q \zeta_q a_{isq} + o_{is}, \tag{5}$$

where o_{is} is still iid EV. $\zeta_q a_{isq}$ are $|Q|$ different error components with $q \in Q$ that, along with o_{is}, define the stochastic part ϵ_{is} of utility u_{is}. a_{isq} are operationalized by observed attributes relating to school s and student i. ζ_q are random terms with zero mean.

[1] Also known as the modifiable areal unit problem (MAUP) as defined by Openshaw [35].

So the MMNL utilities are

$$u_{is} = \sum_n \beta_{sn} w_{isn} + \sum_q \zeta_q a_{isq} + o_{is}. \tag{6}$$

Based on this the MMNL choice probabilities are

$$P_i(s|\mathcal{C}_i) = \int_\zeta \left(\frac{e^{\sum_n \beta_{sn} w_{isn} + \sum_q \zeta_q a_{isq}}}{\sum_{l \in \mathcal{C}_i} e^{\sum_n \beta_{ln} w_{iln} + \sum_q \zeta_q a_{ilq}}} \right) f(\zeta)\, d\zeta. \tag{7}$$

$f(\zeta)$ is a $|Q|$-dimensional density function. Roughly speaking, the MMNL probability of (7) is a weighted average of the MNL (4) evaluated at different values of the vector ζ, with the weight given by the density of $f(\zeta)$. The parameters characterizing f and coefficients β_{sn} are estimated using simulated maximum likelihood. If the parameters characterizing f are statistically significantly different from zero,[2] then the MMNL is the more appropriate model compared to the MNL [43, p. 143]. Note that there are no constraints in terms of the density function f. Any density function can be used. The MNL is a special case where the mixing distribution $f(\zeta)$ is degenerate at fixed parameters. Since (7) does not have a closed-form, the choice probabilities have to be simulated. In contrast to the MNL (4) the MMNL (7) is able to handle any substitution pattern desired [9].

In (7) the exponent of e embodies attributes of the schools and students (w_{isn}, a_{isq}) with fixed coefficients β_{sn} and random coefficients ζ_q. If w_{isn} and a_{isq} overlap (in the sense that some of the same attributes enter w_{isn} and a_{isq}), the coefficients of these attributes can be considered to vary randomly with mean β_{sn} and the same distribution as ζ_q around their means.

2.2. Multinomial logit vs. mixed multinomial logit

By a small example we show the difference between the choice probabilities determined by MNL and MMNL. Imagine a student i who chooses from a set of three schools A, B and C. For reasons of convenience we assume that there is just one observable attribute w_{is1}: the distance from the student's home to school. The locations of the schools and the student as well as the distances are given in Fig. 1.

Due to the proximity of school A to C both schools may share some unobserved attributes. For example, they use a common facility (sports ground, assembly hall). Also, we may assume that the student is carried to school by her commuting mother every morning. Schools A and C are both on the route of the trip to work. Hence, they share common unobserved attributes because of accessibility reasons. Consider the following utility functions of our small example:

$$u_{iA} = \beta w_{iA1} + \zeta_1 \cdot 1 + o_{iA},$$
$$u_{iB} = \beta w_{iB1} + o_{iB},$$
$$u_{iC} = \beta w_{iC1} + \zeta_1 \cdot 1 + o_{iC}.$$

Let us say the coefficient of w_{is1}, $\beta = -1$ for all schools $s \in \{A,B,C\}$. Following the notation of (6), we consider only one error component a_{is1} that equals one for schools A and C, zero otherwise. The corresponding coefficient $\zeta_1 \sim \log \mathcal{N}(\mu,\sigma^2)$ with parameters $\sigma = 1$ and $\mu = 0$. This reflects common unobserved (spatial) attributes of schools A and C which are positively evaluated by the student. Note that the coefficient β and the parameters μ and σ of the log-normal density function are set arbitrarily to simple values here in order to make the example comprehensible.[3] Now, we are able to compute the choice probabilities of the MNL (4) and the MMNL (7) that are both specific operationalizations of the RUM formulation in (3). In order to compute the MNL choice probabilities of student i we first compute $v_{is} = \beta w_{is1}$ for all schools A, B and C. Then we use v_{is} to compute (4). Since there are infinite values of ζ we have to simulate the integral of (7) in order to get the respective choice probabilities of the MMNL. See [43, p. 148] for how the simulation is performed. We are particularly interested in the choice probabilities of schools B and C if school A is closed for example (i.e. not available). The computed values are given in Table 1.

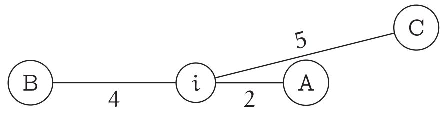


Fig. 1. Exemplary choice situation: a student i faces the choice between three schools. Schools A and C may share some unobserved attributes due to proximity.

Table 1
Exemplary choice probabilities of the student considered for schools A, B and C. The values of the MNL are obtained by (4) and the values of the MMNL are obtained by (7). The ratio of choice probabilities of schools B and C is constant for MNL while decreasing for MMNL (i.e., C is a better substitute for A than B).

School	A	B	C	B/C
All schools available				
MNL	0.844	0.114	0.042	2.718
MMNL	0.909	0.046	0.045	1.003
School A closed				
MNL	0.000	0.731	0.269	2.718
MMNL	0.000	0.437	0.563	0.775

Once school A is closed, the MNL overrates the choice probability of school B. This is due to the IIA property of the MNL (constant ratio of probabilities of B to C). In contrast, the MMNL does not exhibit the IIA and allows for flexible substitution patterns. Table 1 shows that school C is a better substitute for school A because of proximity (see Fig. 1). The different patterns of MNL and MMNL are not accidental, but rather due to the differences in variance and covariance of MNL and MMNL. For the MNL the variance is

$$\mathrm{var}(u_{is}) = \frac{\pi^2}{6}$$

and the covariance

$$\mathrm{cov}(u_{is},u_{il}) = 0.$$

In contrast the variance of the MMNL is

$$\mathrm{var}(u_{is}) = \frac{\pi^2}{6} + \sum_q a_{isq}^2 \sigma_q^2,$$

with σ_q obtained from the parameters characterizing f. The covariance of the MMNL is

$$\mathrm{cov}(u_{is},u_{il}) = \sum_q a_{isq} a_{ilq} \sigma_q^2.$$

To be more precise consider the following variance–covariance matrices G of our small example:

$$G^{\mathrm{MNL}} = \begin{pmatrix} \pi^2/6 & 0 & 0 \\ 0 & \pi^2/6 & 0 \\ 0 & 0 & \pi^2/6 \end{pmatrix},$$

$$G^{\mathrm{MMNL}} = \begin{pmatrix} \pi^2/6+1 & 0 & 1 \\ 0 & \pi^2/6 & 0 \\ 1 & 0 & \pi^2/6+1 \end{pmatrix}.$$

Here, we assume that the parameter σ_1 is statistically significant. Therefore, the MMNL is the better model. Note that if we

[2] See [5, p. 161] for more details about statistical significance.
[3] In an application β and σ (and optionally μ) have to be estimated.

are able to include all key attributes appropriately in v_{is} of (2), the choice process might be mapped by a MNL. Unfortunately, in a spatial choice situation this is seldom the case [19]. We incorporate the MMNL to our mathematical optimization problem of Section 3, because of its property of flexible substitution patterns and the potential to approximate any RUM arbitrarily close .

3. School location models with endogenous demand

The problem of designing a school network with free school choice can be stated as

> "From a set S of potential school locations find a subset $\overline{S}$ that maximizes students' overall or average (weighted) expected utility with respect to certain conditions like budget and capacity."

The utility values are given by (6). Admittedly, the absolute values of (6) are not useful here. Since students and schools are distributed unevenly, there exist students who have higher utilities for all schools available than other students. In order to avoid unfair evaluation of the utilities we have to consider standardized utility values of (6). The standardized utility $\hat{u}_{is}$ exhibits values between 0 and 1. Let u_{is} be given, then $\tilde{u}_{is} = u_{is} - u_i^{max}$, with $u_i^{max} = \max_{s \in S}\{u_{is}\}$ $\forall i$. Now $\acute{u}_{is} = \tilde{u}_{is} - \tilde{u}_i^{min}$, with $\tilde{u}_i^{min} = \min_{s \in S}\{\tilde{u}_{is}\}$ $\forall i$. Finally, $\hat{u}_{is} = \acute{u}_{is}/\acute{u}_i^{max}$ with $\acute{u}_i^{max} = \max_{s \in S}\{\acute{u}_{is}\}$ $\forall i$. Now the question arises: how do we incorporate utility and endogenous demand into a mathematical program corresponding to our problem statement?

3.1. An intuitive approach

In location models we usually have aggregated data at hand. That is, the demand is given as the number of students of a spatial unit[4] or a demand point $r \in R$ expected to enroll at school s. If we treat the students of r as homogeneous[5] we can interpret the choice probabilities of (7) as the proportion of students from r choosing school s. Moreover, we interpret the utilities of (6) as the utility that is perceived by every student in r. Based on this we formalize the problem of Section 3 in a very simple and intuitive way. Therefore, we have to define the sets

- S (potential) schools, indexed by s and l
- R demand points, indexed by r, and the parameters
- b total budget for the school network
- g_r number of students located in r
- k_s total cost of school s
- c_s capacity of school s measured in the number of students
- $\hat{u}_{rs}$ standardized utility of school s for students located in r

Finally, we consider

- Y_s as a binary variable indicating whether school s will be open ($Y_s = 1$) or not ($Y_s = 0$).

The objective of our problem is

$$\max \sum_{r,s} \hat{u}_{rs} g_r \left(\int_{\zeta} \left(\frac{Y_s e^{\sum_n \beta_{sn} w_{rsn} + \sum_q \zeta_q a_{rsq}}}{\sum_l Y_l e^{\sum_n \beta_{ln} w_{rln} + \sum_q \zeta_q a_{rlq}}} \right) f(\zeta)\, d\zeta \right). \tag{8}$$

The objective maximizes the overall weighted utilities perceived by the students located in r if they are enrolled at school s. The utilities are weighted by the number of students of r who are expected to choose school s: $g_r(\cdot)$. The term in brackets is the choice probability known from (7). Obviously, we consider the utilities only if a school s is available, i.e. $Y_s = 1$. Since we usually have to account for a total budget we write

$$\sum_s k_s Y_s \le b. \tag{9}$$

In order to make sure that every student is able to enroll at a school

$$\sum_s c_s Y_s \ge \sum_r gr \tag{10}$$

The domain of the decision variable is given by

$$Y_s \in \{0,1\} \quad \forall s. \tag{11}$$

Of course, we might incorporate school specific capacities, etc. But as we can easily imagine that the model is not solvable in a justifiable way because of the highly non-linear objective (8). Hence, we should focus on a much more simple reformulation of this non-linear 0–1 program at first (in terms of how demand can be considered).

3.2. A simulation-based approach

In order to find a suitable formulation (linear 0–1 program) of the problem stated in Section 3 we employ the approach of Haase [13]. The main idea is to consider students $i \in I$ explicitly, because the utilities of (6) are defined for each student i. Therefore, we do not need to consider the aggregated values g_r with $r \in R$ (spatial units) explicitly. This is very appealing, because our model becomes independent of the geographical scale to some extent. In contrast, the solution and the respective computational effort of the model in Section 3.1 relies on the geographical scale of R.

In most of the applications we might not obtain information about every single student. Thus, we simulate students from the available data. A small example should clarify how this could be done: Usually data is available in the form of g_r. Imagine $g_1 = 3.8$. In order to generate individual students from the total student number 3.8 we generate weights $\omega_i = 1$ for $i \in \{1,2,3\}$ and $\omega_4 = 0.8$ with $I = \{1,2,3,4\}$ for the district $r=1$. Hence, $\sum_{i=1}^{4} \omega_i = 3.8$. Moreover, attributes used in (6), like the travel distance, can be easily transferred from demand point r to individual i. Now, we are able to compute $\hat{u}_{is}$. This is done for all $r \in R$.

For our new mathematical program we additionally define the sets

- I simulated (artificial) students, indexed by i
- M capacity levels, indexed by m, and parameters
- k_{sm} total cost of school s with capacity level m
- c_{sm} capacity of school s given capacity level m measured in the number of *real* students
- $\hat{u}_{is}$ simulated standardized utility of school s for student i
- ω_i weights used to generate artificial students from aggregated data
- h number of *real* students in the considered region or study area
- b total budget for the school network

and

- X_{is} binary variable – which will be later relaxed – indicating whether student i will attend school s ($X_{is} = 1$) or not ($X_{is} = 0$)
- Y_{sm} binary variable indicating whether school s will be open ($Y_{sm} = 1$) with capacity level m or not ($Y_{sm} = 0$).

---

[4] A census block for example.

[5] Note that we might consider homogeneous subgroups of students of r in order to account for socio-demographic attributes like race and gender.

Table 2
School attributes: the majority of courses offered by a certain school specify the school's profile. Travel distances based on the road network are given as mean and standard deviation of all sampled students who are enrolled at the respective school [33]. Distances are computed by using a shortest path algorithm.

School *s*	Profile							School capacity*	Travel distance (m)		Cost in €
	a	*b*	*c*	*d*	*e*	*f*	*g*		Mean	SD	k_s
STBE						1		3	5508.938	2659.396	0
MC	1							4	5560.216	2641.011	487 499
WIES	1							3	5593.663	2883.41	611 974
ANNE	1							3	5997.945	2876.5	611 974
BB				1				4	5455.924	2811.178	491 196
DKS					1			3	6140.887	2564.16	709 093
WALD							1	2	6518.753	2541.077	0
RORO			1					3	5837.075	2576.963	545 867
PEST	1							3	7429.339	2776.758	502 898
KLOT**				1				3	10 383.761	3634.24	430 387
DKS2**					1			3	9176.732	3645.66	430 000
BUEH**					1			3	9176.732	3645.66	611 974
MAN		1						3	6124.141	3333.742	693 793
EVKZ						1		3	5904.646	3268.667	0
HOGA							1	1	5427.872	2919.612	0
JOHA					1			4	5427.872	2919.612	916 254
SEID	1							3	6396.568	3634.358	611 974
HE				1				5	5811.503	3343.249	754 801
GZW**	1							4	8874.82	4597.433	478 796
JAH				1				4	6732.897	3945.258	767 296
WUST**				1				3	7451.842	4222.193	611 974
FL				1				3	6146.491	3200.907	594 580
VITZ					1			4	6023.214	3276.623	507 460
PLAU				1				5	7085.786	3659.787	840 821
COTT**					1			5	7767.026	3726.206	665 470
JAS**	1							5	7986.274	3931.189	405 372

a: Math; *b*: Math core; *c*: Languages core; *d*: Math and Languages; *e*: Math and Music/Arts; *f*: Math, Languages and Music/Arts (Private); and *g*: No Profile (Private).

* Measured in the number of classes per grade.

** Outskirt school.

Using this notation we consider for our linear 0–1 program the objective function

$$\max F = \left(\sum_{i,s} \omega_i \hat{u}_{is} X_{is} \right) \bigg/ |I| \tag{12}$$

that maximizes the average expected weighted students' utility. By the equations

$$\sum_s X_{is} = 1 \quad \forall i, \tag{13}$$

we guarantee that a student attends exactly one school. In order to account for the total budget that is available for the school network we consider the constraints

$$\sum_{s,m} k_{sm} Y_{sm} \le b. \tag{14}$$

Of course, we have to make sure that the capacity of each open school is not exceeded. Therefore, we apply

$$\sum_i \omega_i \frac{h}{|I|} X_{is} \le \sum_m c_{sm} Y_{sm} \quad \forall s. \tag{15}$$

These constraints assure that students can only be assigned to an open school. To understand these constraints in the correct way, we have to remember (i) that we measure the school's capacity c_{sm} in the number of real students and (ii) that we assign simulated students $i \in I$ to school $s \in S$ (X_{is}). Because of that we need to adjust the number of simulated students assigned to school s, $\sum_i X_{is}$, in order to match the schools' capacity given in real student numbers. This is done by the factor $h/|I|$

$$\sum_m Y_{sm} \le 1 \quad \forall s \tag{16}$$

ensure that school s is established with at most one capacity level. Moreover, we may use

$$X_{is} \le \sum_m Y_{sm} \quad \forall i,s, \tag{17}$$

as a cut generating constraint. The domains of the variables are defined as follows:

$$X_{is} \in \{0,1\} \quad \forall i,s, \tag{18}$$

$$Y_{sm} \in \{0,1\} \quad \forall s,m. \tag{19}$$

Remarks.

- If we have data for every individual student available, then $\omega_i = 1\ \forall i$ and (12) becomes $\max F = \sum_{i,s} \hat{u}_{is} X_{is}$ and (15) becomes $\sum_i (h/|I|) X_{is} \le \sum_m c_{sm} Y_{sm}\ \forall s$.
- Let $|\overline{S}|$ be the number of established schools. Then, when relaxing the binary variables X_{is}, at most $2 \times |\overline{S}|$ non-integer variables X_{is} are in our solution. As $|\overline{S}|$ will be very small compared to $|I|$, we will derive a solution very close to the optimal integer solution. Therefore, we propose to relax X_{is}.
- The extension to a multi-period location problem as proposed by Müller [29] is straightforward.
- Instead of the utility values we can also try to maximize the number of first choice alternatives for the students as proposed by Church and Schoepfle [12]. Then the problem is reduced to a single Knapsack-problem. We do not apply this here.
- Of course, additional constraints like a pre-defined racial mix per school or a reasonable maximum travel-time for the commute-to-school could be easily added to our model.

Table 3
Variables used for MNL and MMNL and corresponding estimates of coefficients. Standard errors of the estimates are given in brackets. Dummy variables equal one if true, zero otherwise (if school s offers profile a, then $w_{is2}=1$ for example). Coefficient ζ_1 is a realization of a random variable with $\mathcal{N}(\mu_1,\sigma_1^2)$. Parameter σ_1 is the corresponding estimate. $\overline{\rho}^2$ is a relative measure of goodness-of-fit taking into account the degrees of freedom.

Variable	Coefficient	*MNL*	*MMNL*
w_{is1} : log distance from i to s	$\beta_{s1}=$	-1.750 (0.021)	-1.750 (0.023)
w_{is2}: Dummy profile a	$\beta_{s2}=$	0	0
w_{is3}: Dummy profile b	$\beta_{s3}=$	0.738 (0.094)	0.746 (0.089)
w_{is4}: Dummy profile c	$\beta_{s4}=$	1.410 (0.082)	1.420 (0.080)
w_{is5}: Dummy profile d	$\beta_{s5}=$	0.968 (0.065)	0.963 (0.063)
w_{is6}: Dummy profile e	$\beta_{s6}=$	0.896 (0.068)	0.886 (0.066)
w_{is7}: Dummy profile f	$\beta_{s7}=$	1.170 (0.075)	1.180 (0.071)
w_{is8}: Dummy profile g	$\beta_{s8}=$	$-$ 0.371 (0.152)	$-$ 0.363 (0.149)
a_{is1}: Dummy outskirt	$\zeta_1\sim\mathcal{N}(\mu_1,\sigma_1^2)$		$\mu_1=0$ σ1 = 0.441 (0.133)
$\bar{\rho}^2$		0.324	0.325

a: Math; b: Math Core; c: Languages Core; d: Math and Languages; e: Math and Music/Arts; f: Math, Languages and Music/Arts (Private); and g: No Profile (Private).

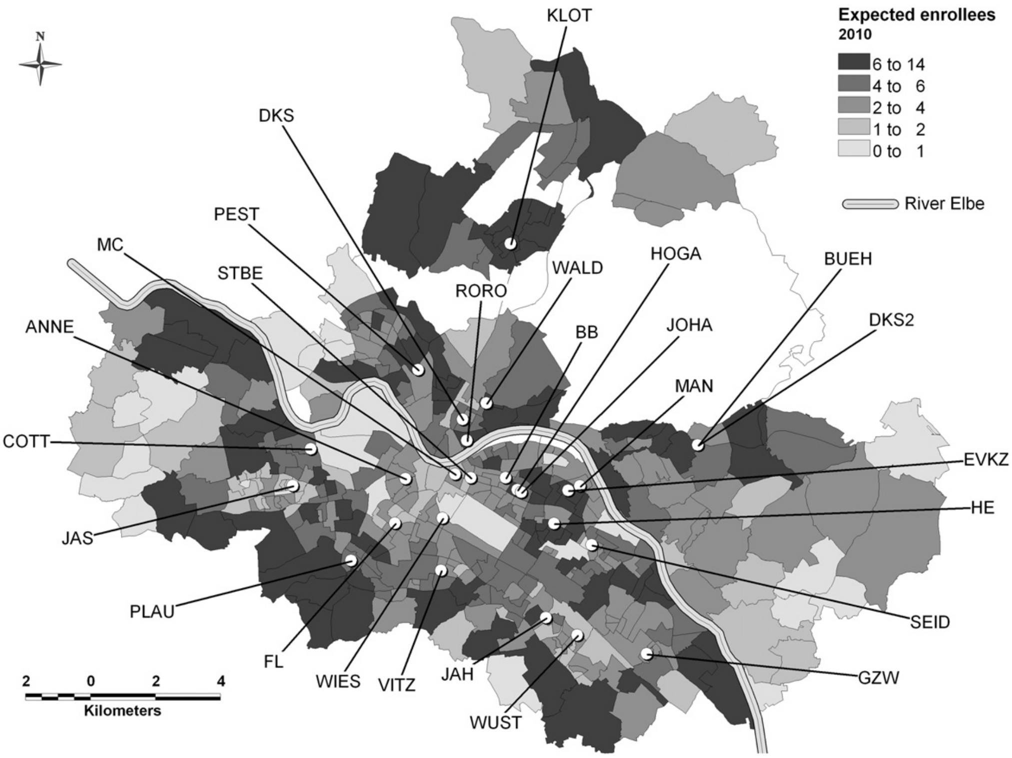


Fig. 2. Location of 26 Gymnasium-schools in the city of Dresden, Germany. The number of expected students who qualify for Gymnasium-school is given on the scale of statistical districts (392).

4. Numerical investigations on a real world application

In order to verify the applicability of the proposed model we employ our approach to a real world application. We consider the city of Dresden, Germany that is partitioned into 392 administrative units (statistical districts). The data of the potential 26 Gymnasium-schools (equivalent to high-school)[6] in our example is given in Table 2. Our goal is to determine a public school network for the year 2010. Therefore, we first specify and estimate a MNL and a MMNL using the school choice data of Müller et al. [32]. Then we discuss data related issues related to the mathematical program followed by the computational results.

4.1. Specification of random utility models

The specification of the RUM (MNL and MMNL) considered here can be easily verified using Tables 2 and 3 and the corresponding Eqs. (1), (2), and (6). The coefficients of Table 3 are estimated using Biogeme 1.8.[7] MMNL exhibits one error

[6] See [31,32] for more details.

[7] http://www.biogeme.epfl.ch

component for convenience reasons. Note that the error component specification of MMNL is quite challenging. For more details on this issue see Walker [44]. We interpret the error component in the way that all schools located at the outskirts share some common unobserved attributes. These attributes might be related to accessibility and perceived neighborhood characteristics. Since we do not know whether the students evaluate the unobserved attributes positively or negatively we assume a normal distribution with zero mean (i.e. $\mu_1 = 0$). We use the log of the distance traveled to school (w_{is1}) because we assume a non-linear evaluation of the distance. See Tables 2 and 3 for specification. The coefficient β_{s2} is set to zero for identification purposes. For more details on the estimation and specification of discrete choice models see [5,21,43].

4.2. Data issues

The cost and capacity per school are taken from Table 2. For each statistical district,[8] the number of expected newly enrolled students who qualify for Gymnasium-school is given. Based on this we compute the weights ω_i as described in Section 3.2. In total 1343 enrollees who qualify for Gymnasium-school were expected in the year 2010 (Fig. 2). Note that in the year 1995 this number was 2504. This demonstrates the considerable demographic change, in particular, in eastern Germany. Moreover, it justifies school consolidation. Capacity c_s is given in number of classes times number of students per class. We postulate a class size of 25 students. For reasons of convenience we do not consider capacity levels m. We assume that every enrolled student is allowed to remain with her school until the final examination, i.e. no change of schools is considered. Therefore, it is not necessary to consider separate student cohorts per school [31]. The annual fixed cost is given by k_s. The fixed costs of the four private schools are set to zero. The local education authority is not in charge for the funding of private schools.

In our computational investigation, all problems are implemented in GAMS 23.5 and solved with Cplex 12.0 on a Dell Precision M4500 with an Intel(R) Core(TM) i5 CPU with a 2.53 GHz processor and 4 GB 1067 MHz DDR3 memory using the operating system Windows 7 (64 bit).

4.3. Computational results

First of all we consider Table 4 in order to examine the influence of the stochastic component of utility (see (2) and (6)). Obviously, we obtain different solutions for MNL and MMNL in terms of the objective and open schools. Particularly, outskirt schools are affected as expected. This is because we account for correlation among outskirt schools in MMNL. We assume that MNL is less suitable, because the error component of MMNL is statistically significant. Now we might ask, what happens if we choose the wrong (random utility) model to describe the demand of the students for schools? For this reason, we choose MNL (as the "wrong" model) and solve our school location problem. Then, we fix the decision variables Y_s indicating the locational decisions. Then, we solve the school location problem again but with fixed Y_s and MMNL. We do this for different cardinal numbers of I as shown in Table 5. The locational decisions are based on the utilities $\hat{u}_{is}$ determined by MNL. In order to get an idea of "how wrong" the deployment of MNL might be, we have to compare the corresponding objective with the objective of the school location problem based on $\hat{u}_{is}$ determined by MMNL and fixed Y_s (see Table 5). At a first glance we might argue that the gap between the objectives is not remarkable, but one has to interpret the results relative to the sampled student numbers. That is, in the first row of Table 5, 1343 students are affected by a gap Δ of 0.5. In the last row 11 501 students are affected by a gap Δ of -0.05. Indeed, the empirical differences between MNL and MMNL are small as well (in terms of coefficient magnitude and error component structure). We expect an increase in the gap as the empirical differences increase as well (i.e. more complex error component structure). Table 6 is of a particular interest because we want to know about the influence of the weighting factor $h/|I|$. We obtain a robust solution using $h/|I| \le 1/8$. The true optimal objective function value seems to be between 92.9 and 93.0. Even problems with large $|I|$ are solvable within less than 20 min. With respect to the relaxation of X_{is}, all instances can be solved optimally or at least with a maximum relative gap of 2%. Table 6 verifies the second remark in Section 3.2.

Table 4
Sensitivity of the solution to the specification of the random component of utility. The budget $b = 6$ million €.

Model	I	$F^* \times 100$	Schools open
MNL	1343	92.91	STBE, BB, WALD, RORO, KLOT, DKS2, EVKZ, HOGA, HE, WUST, FL, VITZ, **PLAU**, COTT
MMNL	1343	92.89	STBE, BB, WALD, RORO, KLOT, **BUEH**, EVKZ, HOGA, HE, **JAH**, WUST, FL, VITZ, COTT

Table 5
Validation of the results based on MNL by the "true" model MMNL: We solve the school location problem with $\hat{u}_{is}$ determined by MNL. Then we fix Y_s to the solution and solve the problem again but using $\hat{u}_{is}$ determined by MMNL. The corresponding objectives are shown in the third column. The budget $b = 6$ million €.

I	$F^* \times 100$ of solution obtained and evaluated by MNL	$F^* \times 100$ of solution obtained by MNL and evaluated by MMNL	Δ
1343	92.91	92.41	0.5
2459	92.86	92.60	0.26
3598	92.87	92.87	0
4720	93.04	93.09	−0.05
5854	93.06	92.92	0.14
6994	92.73	92.98	−0.25
8131	92.91	92.86	0.05
9257	92.88	92.63	0.25
10 403	92.84	92.82	0.02
11 501	92.76	92.81	−0.05

5. Summary

We introduce an approach for formalizing the school network design problem where free school choice by students can be modeled by random utility models. Therefore, we are able to account for any substitution pattern of schools—that is, some schools might be better substitutes for each other than others. A mathematical optimization model is presented that explicitly accounts for simulated utilities of a large number of students. The model is independent of the number of spatial administrative units (demand points). For its solution we apply GAMS/Cplex. By a computational investigation we show that real problems can be solved optimally – or at least close to optimality – within a few minutes. For our distinct study we find a sample factor of at least eight times the actual student numbers to produce a robust

[8] Using smaller administrative units, like blocks, does not have any impact on the mathematical model formulation since we consider students i instead of demand points r (see Section 3.2).

Table 6
Computational study of decreasing sample factor h/I. We use MMNL specified in Table 3. The budget $b=6$ million €. CPU-time is given in seconds. Note $|S|=26$. Let t be the number of non-integer X_{is} in the seventh column, then a simple lower bound is given by $F^{*}-t\cdot 100/|I|$.

h/I	I	$F^{*}\times 100$	cpu-time	Schools open	Total nr. of X_{is}	Nr. of non-integer X_{is}
1/1	1343	92.89	16.92	STBE, BB, WALD, RORO, KLOT, BUEH, EVKZ, HOGA, HE, JAH, WUST, FL, VITZ, COTT	34 918	24
1/2	2459	92.79	67.61	See above	63 934	24
1/3	3598	92.84	105.85	See above	93 548	24
1/4	4720	92.81	333.61	STBE, BB, WALD, RORO, KLOT, ~~BUEH~~ **DKS2**, EVKZ, HOGA, HE, ~~JAH~~, WUST, FL, VITZ, **PLAU**, COTT	122 720	20
1/5	5854	92.70	249.71	STBE, BB, **DKS**, WALD, RORO, KLOT, DKS2, EVKZ, HOGA, HE, WUST, ~~FL~~ VITZ, PLAU, COTT	152 204	20
1/6	6994	92.84	381.39	STBE, BB, ~~DKS~~ WALD, RORO, KLOT, DKS2, EVKZ, HOGA, HE, WUST, **FL**, VITZ, PLAU, COTT	181 844	20
1/7	8131	93.01	517.31	STBE, BB, **DKS**, WALD, RORO, KLOT, DKS2, EVKZ, HOGA, HE, WUST, ~~FL~~ VITZ, PLAU, COTT	211 406	18
1/8	9257	92.61	729.49	STBE, BB, ~~DKS~~ WALD, RORO, KLOT, DKS2, EVKZ, HOGA, HE, WUST, **FL**, VITZ, PLAU, COTT	240 682	18
1/9	10 403	92.92	1088.24	See above	270 478	20
1/10	11 501	92.98	983.49	See above	299 026	16
1/15	17 189	92.89	2720.17	See above	446 914	18
1/25	28 478	92.94	6968.61	See above	740 428	16
1/50	56 761	92.98	33 745.88	See above	1 475 786	8
1/100	113 341	92.90	157 880.29	See above	2 946 866	0

solution. Moreover, we see that the use of a specific random utility model has to be justified empirically.

Our approach overcomes the problems arising with unrealistic proxies for demand. Now, practitioners are enabled to use results of state-of-the-art location choice models (here school choice) within a mathematical program in order to design a school network. Further, they are enabled to evaluate decisions on the characteristics of schools (the profile offered, for example) in terms of expected demand and optimal location. This might be of a particular interest to managers of private schools. Future tasks might comprise extensions and reformulations in order to model multi-period problems and non-public sector problems, like competitive facility location [39,36].

Acknowledgments

We like to thank two anonymous reviewers and the editor for helpful comments on an earlier version of the paper.

References

[1] Albareda-Sambola M, Fernandez E, da Gama FS. The facility location problem with Bernoulli demands. Omega 2011;39(3):335–45.
[2] Antunes A, Gomes C, Ribeiro A. An accessibility-maximization approach to public facility planning under distance-sensitive demands. Studies in Regional and Urban Planning 2004;10:109–17.
[3] Antunes A, Peeters D. A dynamic optimization model for school network planning. Socio-Economic Planning Sciences 2000;34:101–20.
[4] Bahrenberg G. Providing an adequate social infrastructure in rural areas: an application of a maximal supply dispersion model to elementary school planning in Rotenburg/Wümme (FRG). Environment and Planning A 1981;13:1515–27.
[5] Ben-Akiva M, Lerman S. Discrete choice analysis, theory and applications to travel demand. Cambridge, MA: MIT Press; 1985.
[6] Bierlaire M. BIOGEME: a free package for the estimation of discrete choice models. In: The third Swiss transport research conference; 2003.
[7] Bliemer MC, Rose JM. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transportation Research Part B: Methodological 2010;44(6):720–34.
[8] Borgers A, Oppewal H, Ponjé M, Timmermans H. Assessing the impact of school marketing: conjoint choice experiments incorporating availability and substitution effects. Environment and Planning A 1999;31:1949–64.
[9] Brownstone D, Train KE. Forecasting new product penetration with flexible substitution patterns. Journal of Econometrics 1999;89:109–29.
[10] Burgess S, Briggs A. School assignment, school choice and social mobility. Working Paper 06/157, Center for Market and Public Organization; 2006.
[11] Church RL, Murray AT. Modeling school utilization and consolidation. Journal of Urban Planning and Development 1993;119(1):23–38.
[12] Church RL, Schoepfle OB. The choice alternative to school assignment. Environment and Planning B 1993;20:447–57.
[13] Haase K. Discrete location planning. Working Paper ITLS, WP-09-07; 2009. p. 1–13.
[14] Hastings J, Kane T, Staiger D. Parental preferences and school competition, evidence from a public school choice program. Yale Economic Applications and Policy Discussion Paper 2005;11805:1–50.
[15] Haynes K, Fotheringham A. The impact of space on the application of discrete choice models. Review of Regional Studies 1990;20:39–49.
[16] Hoxby C, editor. The economics of school choice. Chicago: University of Chicago Press; 2003.
[17] Huff D. A probabilistic analysis of shopping centre trading areas. Land Economics 1963;39(1):81–90.
[18] Huff D. Defining and estimating a trading area. Journal of Marketing 1964;28(3):34–8.
[19] Hunt L, Boots B, Kanaroglou P. Spatial choice modelling: new opportunities to incorporate space into substitution patterns. Progress in Human Geography 2004;28:746–66.
[20] Kane J, Spizman L. Race, financial aid awards and college attendance, parents and geography matter. American Journal of Economics and Sociology 1994;53(1):85–95.
[21] Koppelman FS, Bhat C. A self instructing course in mode choice modeling: multinomial and nested logit models. Technical Report, U.S. Department of Transportation, Federal Transit Administration; 2006.
[22] Lankford H, Lee E, Wyckoff J. An analysis of elementary and secondary school choice. Journal of Urban Economics 1995;38:236–51.
[23] Lankford H, Wyckoff J. The effect of school choice and residential location on the racial segregation of students. Advances in Applied Microeconomics, vol. 14, Chapter 10. Elsevier; 2006. p. 185–239.
[24] Lemberg DS, Church RL. The school boundary stability problem over time. Socio-Economic Planning Sciences 2000;34:159–76.
[25] Marianov V, Ríos M, Icaza MJ. Facility location for market capture when users rank facilities by shorter travel and waiting times. European Journal of Operational Research 2008;191(1):30–42.
[26] McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P, editor. Frontiers of econometrics. New York: Academic Press; 1973. p. 105–42.
[27] McFadden D, Train K. Mixed MNL models for discrete choice response. Journal of Applied Econometrics 2000;15(5):447–70.
[28] Min H. A multiobjective retail service location model for fastfood restaurants. Omega 1987;15(5):429–41.
[29] Müller S. Dynamic school network planning in urban areas, Urban and Regional Planning, vol. 5. LIT Verlag, Berlin; 2008.
[30] Müller S. Bildungsstandorte und demographischer Wandel, Langfristig optimale Schulstandorte unter Berücksichtigung schwankender Schülerzahlen. Standort—Zeitschrift für angewandte Geographie 2010;34(1):6–10.
[31] Müller S, Haase K, Kless S. A multi-period school location planning approach with free school choice. Environment and Planning A 2009;41(12): 2929–45.
[32] Müller S, Haase K, Seidel F. Exposing unobserved spatial similarity: evidence from German school choice data. Geographical Analysis 2012;44: 65–86.

[33] Müller S, Tscharaktschiew S, Haase K. Travel-to-school mode choice modelling and patterns of school choice in urban areas. Journal of Transport Geography 2008;16(5):342–57.
[34] Nakanishi M, Cooper L. Parameter estimation for a multiplicative competitive interaction model—a least square approach. Journal of Marketing Research 1974;11:303–11.
[35] Openshaw S. The modifiable areal unit problem. Concepts and Techniques in Modern Geography 1984;38:41.
[36] Pelegrin B, Fernandez P, Perez MDG, Hernandez SC. On the location of new facilities for chain expansion under delivered pricing. Omega 2012;40(2):149–58.
[37] Pizzolato ND. A heuristic for large-size p-median locations with application to school location. Annals of Operations Research 1994;50:473–85.
[38] Redondo JL, Fernandez J, Arrondo AG, Garcia I, Ortigosa PM. Fixed or variable demand? Does it matter when locating a facility? Omega 2012;40(1):9–20.
[39] ReVelle C, Murray AT, Serra D. Location models for ceding market share and shrinking services. Omega 2007;35(5):533–40.
[40] Serra D, Colomé R. Consumer choice and optimal locations models: formulations and heuristics. Papers in Regional Science 2001;80:439–64.
[41] Street D, Burgess L. The construction of optimal stated choice experiments. Wiley; 2007.
[42] Timmermans H, Borgers A, van der Waerden P. Mother logit analysis of substitution effects in consumer shopping destination choice. Journal of Business Research 1991;23:311–23.
[43] Train KE. Discrete choice methods with simulation. Cambridge University Press; 2003.
[44] Walker J. Extended discrete choice models: integrated framework, flexible error structures and latent variables. PhD thesis, Massachusetts Institute of Technology; 2001.

2.2.2 Districting Problems

1. Haase, K., Müller, S. (2014): **Upper and lower bounds for the sales force deployment problem with explicit contiguity constraints**. *European Journal of Operational Research* **237**(2), 677-689.

2. Müller, S. (2011): **Assessment of school closures in urban areas by simple accessibility measures**. *Erdkunde* **65**(4), 401-414.

European Journal of Operational Research 237 (2014) 677–689

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Upper and lower bounds for the sales force deployment problem with explicit contiguity constraints

Knut Haase, Sven Müller *

Institute for Transport Economics, University of Hamburg, D-20146 Hamburg, Germany

ARTICLE INFO

Article history:
Received 18 August 2012
Accepted 27 January 2014
Available online 18 February 2014

Keywords:
Large scale optimization
Sales force deployment
Semi-infinite programming
Column generation
Branch-and-Price
Contiguity

ABSTRACT

The sales force deployment problem arises in many selling organizations. This complex planning problem involves the concurrent resolution of four interrelated subproblems: sizing of the sales force, sales representatives locations, sales territory alignment, and sales resource allocation. The objective is to maximize the total profit. For this, a well-known and accepted concave sales response function is used. Unfortunately, literature is lacking approaches that provide valid upper bounds. Therefore, we propose a model formulation with an infinite number of binary variables. The linear relaxation is solved by column generation where the variables with maximum reduced costs are obtained analytically. For the optimal objective function value of the linear relaxation an upper bound is provided. To obtain a very tight gap for the objective function value of the optimal integer solution we introduce a Branch-and-Price approach. Moreover, we propose explicit contiguity constraints based on flow variables. In a series of computational studies we consider instances which may occur in the pharmaceutical industry. The largest instance comprises 50 potential locations and more than 500 sales coverage units. We are able to solve this instance in 1273 seconds with a gap of less than 0.01%. A comparison with Drexl and Haase (1999) shows that we are able to halve the solution gap due to tight upper bounds provided by the column generation procedure.

1. Introduction

In many selling organizations, sales force deployment is an important tool by which sales management improves profit remarkably (Cravens & LaForge, 1983; Lodish, Curtis, Ness, & Simpson, 1988; Capron & Hulland, 1999 & Zoltners & Lorimer, 2000). It involves the concurrent resolution of four interrelated subproblems:

- sizing of the sales force,
- locations of the sales representatives,
- sales territory alignment, and
- sales resource allocation.

All these subproblems have to be solved simultaneously so that the total profit contribution will be maximized (Drexl & Haase, 1999). The subproblems can be briefly described as follows: Consider a large geographical market area that is partitioned into a set of so-called sales coverage units (SCUs). *Sizing of the sales force* implies selecting the appropriate number of sales representatives required to penetrate the market area. A sales territory and a location are unequivocally assigned to each sales representative. Thus, by sales force sizing we also decide on the number of sales territories and locations. The *salesperson location* aspect of the problem involves determining a SCU for a sales representative to be located in. *Sales territory alignment* may be viewed as the partitioning problem of grouping the SCUs around the locations of the sales representatives into larger geographic areas called "sales territories". *Sales resource allocation* refers to the problem of allocating a sales representative's time to the assigned SCUs. Note, sales force deployment is an aggregate planning problem. The usual planning horizon is one year. So, decisions as how often (e.g., every week, once a month), which days and at what time a customer is to be visited (and the sequence of customers) are the subject of subsequent planning stages (Golalikhani & Karwan, 2013).

Comprehensive reviews of related contributions are given by Drexl and Haase (1999) and Howick and Pidd (1990). Although there is a long tradition of sales force deployment modeling (particularly sales territory alignment) as shown by Lodish (1975), Zoltners (1976), Rangaswamy, Sinah, and Zoltners (1990) and Albers (1996) for example, we see that this field of research is still very popular (Segura-Ramiro, Ríos-Mercado, Álvarez Socarrás, &

* Corresponding author. Tel.: +49 40428389021.
E-mail addresses: knut.haase@wiso.uni-hamburg.de (K. Haase), sven.mueller@wiso.uni-hamburg.de (S. Müller).

http://dx.doi.org/10.1016/j.ejor.2014.01.061
0377-2217/

de Alba Romenus, 2007; Mantrala et al., 2010 & Lee & Yang, 2013) – especially if we consider the practical implementation of approaches (Zolterns & Sinah, 2005). To our novel approach the following work is of particular interest and is therefore discussed in more detail. Skiera and Albers (1998) formulate a model that addresses both the sales territory alignment and the sales resource allocation problems. They propose sales response functions of any given concave form at the level of SCUs. Sales are modeled as a function of selling time, that includes calling time as well as travel time, assuming a constant ratio of travel time to calling time. They consider a resource allocation model and a territory alignment model simultaneously. For the solution a so-called backward deletion procedure is considered. If desired, contiguous sales territories can be constructed by the heuristic. However, explicit contiguity constraints – that is, the constraints used in the model guarantee contiguous sales territories – are not considered. In Drexl and Haase (1999) all four subproblems are covered in order to maximize profit. For the solution of the model they present approximation methods capable of solving large-scale, real-world instances. The methods that provide lower bounds for the optimal objective function value are compared to upper bounds. On average the solution gap, i.e., the difference between upper and lower bounds with respect to the upper bound, is about 3%. Furthermore, they show how the methods can be used to analyze various problem settings of practical relevance.

We present a novel approach to compute tight upper and lower bounds for the sales force deployment problem based on an innovative mathematical program that explicitly accounts for contiguous sales territories. Although contiguous sales territories are compulsory in applications, explicit contiguity constraints are rare so far. Zoltners and Sinah (1983) propose a mixed integer formulation to account for contiguous sales territories. However, this approach is not explicit in terms of contiguity and as a consequence contiguity is not guaranteed (Shirabe, 2005). The contiguity constraints in Drexl and Haase (1999) grow exponentially with the number of spatial units (King, Jacobson, Sewell, & Cho, 2012). In contrast, our approach explicitly accounts for contiguity due to a set of contiguity constraints based on continuous flow variables. By this, we are able to significantly reduce the number of constraints. Further, our model is characterized by an infinite number of binary variables where each variable is related to a point of selling time in a time interval (selling time variables). For its solution we propose a Branch-and-Price approach, i.e., at each node of a Branch-and-Bound tree we solve a master problem by column generation. The corresponding restricted master problem consists of the linear relaxation of an integer formulation with a reduced number of variables. Analytical solutions are introduced for the subproblem. We derive an upper bound for the optimal objective function value of the linear relaxation using dual information. Branching is done on two categories of binary variables. At first, we branch on the variables indicating where locations for sales representatives should be setup. Afterward, we consider assignment variables indicating to which location a sales coverage unit should be assigned. Moreover, we present an intelligible approach to determine an initial lower bound.

Managers might use mathematical sales force deployment approaches to generate good initial solutions and then "fine-tune" the solution by individual expertise (Zolterns & Sinah, 2005). Usually, "what-if" scenarios are employed to improve overall sales performance. Therefore, fast and valid solutions of the mathematical model are needed in order to perform such a "management-heuristic" efficiently (Pinals, 2001). Our computational studies show that especially for large practical instances the new approach provides an outstanding solution quality in reasonable time.

The organization of this paper is as follows: Section 2 describes the problem setting. In particular, we define important terms and explain assumptions that are implicitly involved in these definitions. Section 3 introduces a new mathematical model for sales force deployment. Section 4 provides a Branch-and-Price approach to this formulation to derive tight lower and upper bounds. Section 5 covers the results of computational studies. To stress the practical relevance, we consider a large application-oriented example within the health care and pharmaceutical industry in Germany. The summary and conclusions are provided in Section 6.

2. Problem description

Subsequently, we explain and discuss terms that are important for understanding the addressed problem. Additionally, we provide a problem statement.

- **Account**: A customer (company, self-employed persons, etc.) who is expected to buy products of the company that employs the sales representative.
- **Sales Coverage Unit (SCU)**: A sales coverage unit is considered here to be a relatively small geographical area. The choice and thus the size of the SCUs depend upon the specific application and on whether the required data can be obtained (at a reasonable cost). Counties, zip codes, and company trading areas are some examples of SCUs (see for instance Zoltners & Sinah (1983) and Churchill, Ford, & Walker (1993)). A SCU might contain several accounts.
- **Sales territory**: A sales territory is a geographic area that consists of a set of SCUs with a responsible sales representative. To be responsible means that the sales representative has to provide service for all (potential) accounts located in the corresponding sales territory (Darmon, 2002). For example, a sales representative who is working for a company that provides materials concerning dental surgery will be responsible for all dentists practicing in her sales territory. She is only allowed to sell products in SCUs that belong to her sales territory. Contiguity – that is, the SCUs of a sales territory are connected (Church & Murray, 2009) – is usually demanded by the management of a company for some organizational reasons.
- **Travel time, calling time, selling time**: In order to sell a product a sales representative has to do some time consuming activities. The time required to travel from a location to an account, from an account to another account, and back to the location is called travel time. Although travel time might be considered as routing time, we neglect this fact here and assume travel time is only affected by the distance between SCU and the location of the sales representative. This is reasonable, because sales force deployment is a strategic management issue while routing is more related to operational management. Presenting a product and performing contract negotiations are examples of calling activities. The associated time is denoted as calling time. Now, selling time is defined as the sum of calling time and travel time. We have no specific information about the number of calls to individual accounts because our sales response functions are based on the SCU level.[1] Thus, we assume that the sales response function represents sales with selling time optimally allocated across accounts. For each sales representative the total selling time is restricted. We assume that the calling time is a constant fraction of the selling time. At a first glance, this appears to be a more or less rough average consideration. However, Skiera and Albers (1998) show that this assumption holds if mild conditions are given. For further details see Skiera and Albers (2008).

[1] It is easy to consider sales response function on the account level by generating a (artificial) SCU for each account.

- **Sales**: The (expected) sales in a specific SCU depend on the calling time a sales representative allocates to the SCU. Now, as the calling time is a constant fraction of the selling time and a restriction is defined on the selling time, we are going to consider the relationship between sales and selling time instead of sales and calling time. The relation between selling time and sales can be described by a sales response function (Skiera & Albers, 1998). In applications such a function should be specified by an econometric analysis. Examples are provided in Haase, Lange, and Missong (1999).
- **Fixed location costs, travel costs**: Fixed costs incur for each potential location. These costs arise due to the fixed salary of a sales representative, the rent of an office, car insurance, and so on. Some of these fixed costs may depend on the SCU of the realized location. This may be due to different rents or site costs across SCUs. We assume that travel costs are proportional to travel time.
- **Profit contribution, objective**: Each unit of sales provides a certain amount of profit contribution. The profit contribution obtained by one sales representative is the sum of the profit contributions resulting from the realized sales across the SCUs of the respective sales territory minus the incurred costs. Generally, we consider travel costs and fixed location costs. The sum of profit contributions obtained by the sales force has to be maximized. It may be noteworthy to mention that, as in applications the realized sales are stochastic so the expected profit contribution will be maximized. Typically, in order to describe the relation between selling time and sales, a concave sales response function is considered (Hruschka, 2006). Due to the concave sales response function, more selling time is necessary for the last sold product unit than for the first one. Now, more selling time indicates more travel time and thus more travel costs. So the travel costs incurred for the last sold product are larger than that incurred for the first one. Thus, in general the average profit contribution per sales unit in which travel costs are taken into account decreases as the number of sold products increases. Therefore, to accurately maximize the profit contribution obtained by a sales representative a (travel) cost function has to be taken explicitly into account. Note, in the applications of Drexl and Haase (1999) and Skiera and Albers (1998) travel costs are not significant.
- **Problem statement**: Determine the appropriate sales force size and respective sales representative locations, construct a contiguous sales territory for each sales representative, and allocate the total available selling time of each sales representative over her sales territory so that the sum of profit contributions obtained by the sales representatives is maximized.

3. Mathematical formulation

First, we formalize the profit contribution function in order to calculate the objective function coefficients (Skiera & Albers, 2008). Then we present a new and innovative mathematical model for the sales force deployment problem.

3.1. Profit contribution function

Let be

α	per unit profit contribution of sales, and
$t \in [0, \mathfrak{T}]$	selling time the sales representative located in SCU i allocates to SCU j. With $\mathfrak{T}$ as the total available selling time of a sales representative per period.

Then the profit contribution obtained by the sales representative is derived from the profit contribution function

$$p_{ij}(t) = \alpha \cdot s_{ij}(t) - k_{ij}(t) \quad (1)$$

where

$s_{ij}(t)$	is the sales response function and
$k_{ij}(t)$	is the selling costs function

of the sales representative, respectively. In the literature solely concave sales response functions are considered (see Skiera & Albers, 1998 & Mesak & Ellis, 2009).

Exemplary function type. To formulate an exemplary profit contribution function we introduce some more symbols:

b	calling time elasticity ($0 < b < 1$)
n_j	indicator for the "calling time profitability" of SCU j (e.g. number of (potential) accounts in SCU j)
β_{ij}	travel time fraction of the selling time allocated to SCU j by the sales representative located in SCU i ($0 < \beta_{ij} < 1$)
μ	scaling parameter ($\mu > 0$)
h	cost per travel time unit ($h \geqslant 0$)

Now, by

$$s_{ij}(t) = \mu \cdot n_j((1 - \beta_{ij})t)^b \quad (2)$$

we define a sales response function and by

$$k_{ij}(t) = h \cdot \beta_{ij} \cdot t \quad (3)$$

a cost function and then we derive the profit contribution function

$$p_{ij}(t) = c_{ij} \cdot t^b - o_{ij} \cdot t \quad (4)$$

where $c_{ij} = \alpha \cdot \mu \cdot n_j(1 - \beta_{ij})^b$ and $o_{ij} = h \cdot \beta_{ij}$.

As stated before and to be shown explicitly in the two empirical examples below the profit contribution function is assumed to be concave and selling costs are neglected (see Fig. 1). It should be noted here, that the solution approach of Section 4 is also suited for considering a calling time elasticity, say b_{ij}, depending both on a sales representative i and the SCU j. This might be indicated, for example, if we observe a heterogeneous spatial distribution of the customer demand as reported in Müller, Wilhelm, and Haase (2013).

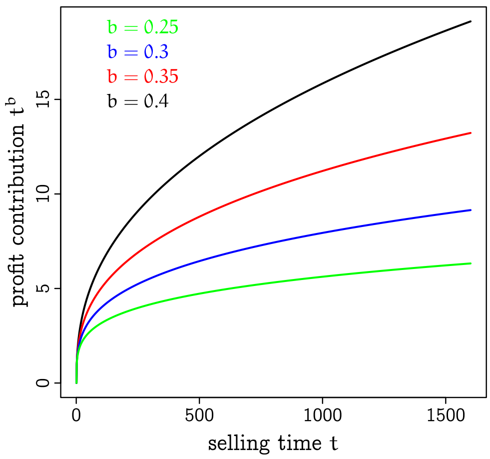


Fig. 1. Profit contribution dependent on selling time t with respect to different elasticities b. $c_{ij} = 1$ and $o_{ij} = 0$.

Example 1. Sales Territory Alignment. In Skiera and Albers (1998) the results of an application to a mid-sized German company are presented. The company had hired 10 sales representatives and adopted the 95 two-digit postal areas of Germany as SCUs. They consider the following sales response function corresponding to (2):

$$s_{ij}(t) = 1350\mathrm{POT}_j^{0.625}(1+\delta_{ij})^{-0.375}t^{0.375} \tag{5}$$

where

POT_j is the number of potential accounts in SCU j and
δ_{ij} is the ratio of travel time to calling time regarding sales representative i in SCU j.

Now, defining

$$c_{ij} = 1350\mathrm{POT}_j^{0.625}(1+\delta_{ij})^{-0.375} \tag{6}$$

we derive

$$s_{ij}(t) = c_{ij}\cdot t^{0.375} \tag{7}$$

The total available selling time is $\mathfrak{T} = 1350$ time units for each sales representative. Fixed costs have not to be taken into account as the sales representatives are already located. Moreover, as no crucial travel costs are considered, $o_{ij} = 0$.

Example 2. Sales Force Deployment. The following sales response function is considered by Drexl and Haase (1999):

$$s_{ij}(t) = 0.205H_j(4.6+d_{ij})^{-0.285}t^{0.285} \tag{8}$$

where

H_j is the number of residents of SCU j and
d_{ij} is the time to drive from SCU i to SCU j.

Setting $c_{ij} = 0.205H_j(4.6+d_{ij})^{-0.285}$ and $o_{ij} = 0$, again, we obtain the profit function as defined in (4).

3.2. Model

We define additionally the sets and parameters

- J set of SCUs, indexed by j and v,
- I set of SCUs ($I \subseteq J$) that can be used to locate a sales representative, indexed by i,
- J_i set of SCUs that can be assigned to a sales representative located in SCU i ($J_i \subseteq J$),
- A_j set of SCUs adjacent to SCU j, and
- f_i fixed costs per period for locating a sales representative in SCU i ($f_i \geqslant 0$).

Now we consider the decision variables

- y_i =1, if a location is setup in SCU i ($y_i = 0$, otherwise), and
- x_{ijt} =1, if a sales representative located in SCU i is allocating a selling time of t to SCU j ($x_{ijt} = 0$, otherwise),
- w_{ij} =1, if SCU j is assigned to sales representative location i ($w_{ij} = 0$, otherwise),

and the artificial variable

- q_{ivj} quantity of flow from v to j with origin in location i,

and then formulate an optimization model for sales force sizing, location of the sales representative, sales territory alignment, and sales resource allocation as follows:

$$\text{Maximize} \quad F = \sum_{i\in I}\sum_{j\in J_i}\sum_{t\in[0,\mathfrak{T}]} p_{ij}(t)x_{ijt} - \sum_{i\in I} f_i y_i \tag{9}$$

$$\text{subject to} \quad \mathfrak{T}\cdot y_i - \sum_{j\in J_i}\sum_{t\in[0,\mathfrak{T}]} t\cdot x_{ijt} \geqslant 0 \qquad i\in I \tag{10}$$

$$w_{ij} - \sum_{t\in[0,\mathfrak{T}]} x_{ijt} \geqslant 0 \qquad i\in I,\quad j\in J_i \tag{11}$$

$$w_{ij} - y_i \leqslant 0 \qquad i\in I,\quad j\in J_i \tag{12}$$

$$\sum_{i\in I | j\in J_i} w_{ij} \leqslant 1 \qquad j\in J \tag{13}$$

$$w_{ij} + \sum_{v\in A_j\cap J_i}(q_{ijv} - q_{ivj}) - \sum_{\substack{v\in J_i \\ i=j}} w_{iv} = 0 \quad i\in I,\ j\in J_i \tag{14}$$

$$|J_i|\cdot w_{ij} - \sum_{v\in A_j\cap J_i} q_{ivj} \geqslant 0 \qquad i\in I,\quad j\in J_i \tag{15}$$

$$x_{ijt} \geqslant 0 \qquad i\in I,\quad j\in J_i,\quad t\in[0,\mathfrak{T}] \tag{16}$$

$$q_{ijv} \geqslant 0 \qquad i\in I,\quad j\in J_i,\quad v\in A_j \tag{17}$$

$$w_{ij}\in\{0,1\} \qquad i\in I,\quad j\in J_i, \tag{18}$$

$$y_i\in\{0,1\} \qquad i\in I \tag{19}$$

Recall that $t\in[0,\mathfrak{T}]$ is the selling time the sales representative located in SCU i allocates to SCU j. The objective (9) maximizes the total profit contribution of all sales representatives. By (10) selling time of a sales representative is allocated to the assigned SCUs. (11) derive the assignment variables w_{ij}. (12) are cuts to force the location variables y_i towards integer. Since the profit contribution function $p_{ij}(t)$ is strictly concave in t as outlined in Section 3.1 the variables x_{ijt} are forced to take integer values (i.e., 0 or 1) due to (11) and the maximization of (9). Eq. (13) assign each SCU to one (or none) sales representative. (14) and (15) are the contiguity constraints: flow constraints (14) ensure a contiguous path from a location i to each assigned SCU j. (15) guarantee that there is no flow outside the sales territory of i which is originated from i. (16)–(19) define the domains of the variables.

Remarks.

(a) By definition of the binary assignment variables $w_{ij}\in\{0,1\}$, accounts (SCUs) are exclusively assigned to one individual sales representative. This is an assumption in marketing science and marketing management (Lodish, 1976 & Zoltners et al., Zoltners, Sinah, & Lorimer, 2009).

(b) The problem formulation contains an infinite number of variables. If we relax the integrality conditions an upper bound for the optimal objective function value of (9) can be computed by column generation. Therefore, we replace the interval $[0,\mathfrak{T}]$ by a countable set, denoted by T_{ij}. That is, for each sales representative located in SCU i we use a definite set of selling times to be allocated to SCU j.

(c) The contiguity constraints (14) and (15) are in the fashion of Shirabe (2009). Hence, our model regards an explicit contiguity formulation. In contrast to the model presented by Drexl and Haase (1999), here the use of continuous flow variables avoids an exponential increase of the number of constraints in the number of SCUs (King et al., 2012). In conjunction with the objective function (9) we expect the sales territories to be fairly compact as well.

(d) In certain companies accounts may be managed by a team of sales representatives. In order to keep the fashion of our model formulation (9)–(19) we might define $i\in I$ as the potential location for teams of sales representatives. To account for sales team characteristics (size, makeup etc.) we consider a set M and define y_{im} with $\sum_{m \in M} y_{im} = 1 \; \forall \, i \in I$. Further, we consider t_m. Of course, the model formulation (9)–(19) and the solution process (Section 4) have to be adapted accordingly.

4. Branch-and-Price approach

In the following we introduce a Branch-and-Price approach for solving our original problem (9)–(19) of Section 3. At first, we define the master problem and the subproblem of the column generation procedure. Then, we describe how upper bounds for the linear relaxation can be obtained. This is followed by the description of an approach that determines an initial lower bound for the integer problem. Finally, branching and bounding strategies are described.

4.1. Master problem

Master problem. The master problem (MP) is the linear relaxation of (9)–(19), i.e. (18) and (19) are replaced by

$$0 \leqslant w_{ij} \leqslant 1 \qquad i \in I, \quad j \in J_i, \quad \text{and} \tag{20}$$
$$0 \leqslant y_i \leqslant 1 \qquad i \in I \tag{21}$$

respectively. That is, we obtain a linear program (LP) from our original program (9)–(19). The corresponding objective function value is named F^{LP}.

Restricted master problem. Let $T_{ij} \subset \{t | t \in [0, \mathfrak{T}]\}$, then the restricted master problem (RMP) is given by MP using $t \in T_{ij}$ instead of $t \in [0, \mathfrak{T}]$ as outlined in remark (b) in Section 3.2. The corresponding objective function value is denoted by z.

4.2. Subproblem

When applying column generation for a maximization problem we have to identify variables with (maximum) positive reduced cost. In our application the reduced cost of a variable x_{ijt} is calculated by a function.

Reduced cost function. Let

σ_i be the dual variable associated with (10) of the RMP, and
γ_{ij} be the dual variable associated with (11) of the RMP for i and j with $i \neq j$.

Then the reduced cost function $\bar{p}_{ij}(t)$ associated with the sales representative located in SCU i who allocates selling time t to SCU $j \in J_i$ is defined by

$$\bar{p}_{ij}(t) = p_{ij}(t) + t \cdot \sigma_i + \gamma_{ij} \tag{22}$$

The reduced cost function $\bar{p}_{ij}(t)$ computes the reduced cost of the variable x_{ijt}.

Subproblem. A solution of the RMP provides an upper bound for (9)–(19), if

$$\forall \; i \in I, \quad j \in J_i, \quad t \in [0, \mathfrak{T}] : \bar{p}_{ij}(t) \leqslant 0 \tag{23}$$

is satisfied. Checking this condition is denoted as the subproblem.

Solution of the subproblem. Recall that $t \in [0, \mathfrak{T}]$ is the selling time the sales representative located in SCU i allocates to SCU j. Let t' be the selling time that maximizes $\bar{p}_{ij}(t)$. Consider $t \in [0, \mathfrak{T}]$. Since $p_{ij}(t)$ is assumed to be strictly concave (see Section 3.1) we are able to determine the optimal feasible selling time by

$$t^* = \begin{cases} t' & 0 \leqslant t' \leqslant \mathfrak{T} \\ \mathfrak{T} & t' > \mathfrak{T} \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

Now, if

$$\forall i \in I, \quad j \in J_i : \bar{p}_{ij}(t^*) \leqslant 0 \tag{25}$$

then (23) is satisfied.

Example: Consider the profit contribution function (4) with the sales response function (2) and the cost function (3). Then we have to maximize

$$\bar{p}_{ij}(t) = c_{ij} \cdot t^b - (o_{ij} - \sigma_i)t + \gamma_{ij} \tag{26}$$

for all $i \in I$ and $j \in J_i$. Differentiating (26) with respect to t and setting the derivative equal to zero leads to:

$$b \cdot c_{ij} \cdot t^{b-1} - o_{ij} + \sigma_i = 0 \tag{27}$$

Rearranging and applying (24) for given $i \in I$ and $j \in J_i$ leads to:

$$t^* = \begin{cases} \min\left(\left(\frac{o_{ij} - \sigma_i}{b \cdot c_{ij}}\right)^{\frac{1}{(b-1)}}, \mathfrak{T}\right) & , \quad o_{ij} - \sigma_i > 0 \\ \mathfrak{T} & , \quad \text{otherwise} \end{cases} \tag{28}$$

Thus, to check whether a basic solution of the RMP provides an upper bound for the original problem (9)–(19) we have to compute (28) and then $\bar{p}_{ij}(t^*)$ for each combination of $i \in I$ and $j \in J_i$. Now, if for at least one combination $i \in I$ and $j \in J_i : \bar{p}_{ij}(t^*) > 0$ then we might improve the actual solution of the RMP if we take into account variable x_{ijt^*}.

4.3. Upper bound for the linear relaxation

The objective function of the RMP provides a lower bound $\underline{F}^{\mathrm{LP}} = z$ for the linear relaxation of (9)–(19) as long as (25) is not satisfied; that is, not all variables are priced out (i.e., the reduced cost of all variables is less than or equal zero). Usually, the lower bound $\underline{F}^{\mathrm{LP}}$ is very close to the optimal objective function value of the corresponding linear relaxation after a relatively small number of column generation iterations, but many additional iterations may be required to show optimality (du Merle, Villeneuve, Desrosiers, & Hansen, 1999; Lübbecke & Desrosiers, 2005 & Desaulniers et al., Desaulniers, Desrosiers, & Solomon, 2005). To save computational time we terminate the column generation procedure after a certain number of iterations or alternatively when we can show that the current objective function value is very close to the optimal objective function value of the linear relaxation of (9)–(19). To evaluate a feasible integer solution we need a valid upper bound. Now, the maximum improvement of the objective function value by assigning SCU j to location i is indicated by $\bar{p}_{ij}(t^*)$. We take into account that the feasible maximum value of x_{ijt^*} is one. Due to (13) of the MP for each SCU j the improvement $\bar{p}_{ij}(t^*)$ can be realized at most for one location i. Thus

$$\bar{p}^*_j = \max\{(\bar{p}_{ij}(t^*) | i \in I), 0\} \tag{29}$$

indicates an upper bound for the improvement of the objective function value by (re-)assigning SCU j to a location i (with a modified selling time). Since $z\,(z^*)$ is the (optimal) objective function value of the RMP we know that $\overline{F} = z^* + \sum_j \bar{p}^*_j$ is the desired upper bound for the linear relaxation of (9)–(19); also known as the Dantzig–Wolfe bound (Lasdon, 2002, p. 163f). As stated in Section 4.2 the linear relaxation of (9)–(19) provides an upper bound on (9)–(19) and hence $\overline{F}$ is an upper bound on (9)–(19) as well.

4.4. Initial lower bound for the integer problem

We expect enormous computational effort due to the consideration of the contiguity constraints (14) and (15). Therefore, we propose an approach to find a (very good) initial lower bound $\underline{F}^0$ for the original problem (9)–(19). In order to obtain $\underline{F}^0$ we initialize the set of available selling time T_{ij}^0 with sufficient elements. "Sufficient" has to be interpreted in terms of the cardinality of T_{ij}^0 as well as the coverage of the time interval $[0, \mathfrak{T}]$. Given T_{ij}^0 we solve (9)–(19). Based on the respective solution - that is w_{ij}^0 and y_i^0 - we derive the optimal time values for each $i \in I$ and $j \in J_i$ as

$$t_{ij}^0 = \frac{c_{ij}^{\frac{1}{1-b}} \cdot w_{ij}^0}{\sum_{v \in J_i} c_{iv}^{\frac{1}{1-b}} \cdot w_{iv}^0} \cdot \mathfrak{T}. \tag{30}$$

The properties and the derivation of (30) are shown by Beckmann and Golob (1972). Since $T_{ij}^0 \subset [0, \mathfrak{T}]$ we are able to determine the initial lower bound as

$$\underline{F}^0 = \sum_{i \in I} \sum_{j \in J_i} c_{ij} \cdot \left(t_{ij}^0\right)^b \cdot w_{ij}^0 - \sum_{i \in I} f_i y_i^0. \tag{31}$$

We call this procedure an initial *MIP-LB*.

4.5. Column generation procedure

The column generation approach to compute an upper bound $\overline{F}$ of the addressed problem (9)–(19) is outlined in Table 1. In *step 1* and *step 2* the allowed solution gap ϵ for the optimal objective function value of the linear relaxation and η, the allowed gap for the integral solution, have to be set by the user. The same has to be done regarding the maximum number of column generation iterations G to solve an LP (cf. *step 3*). *Steps 4* to *6* are the initialization of the column generation process: We have to guarantee, that at the start of the column generation process the set T_{ij} contains elements so that for each selected (or rather located) sales representative the base (or home) SCU can be assigned. By *step 7* we update the iteration counter g. In *step 8* we solve the current RMP (by the simplex algorithm). From the optimal basic solution we derive the associated dual variables in *step 9*. Computing (24) we derive the optimal selling times in *step 10*. In *step 13* we extend the RMP only by variables that have positive reduced cost (multiple pricing). That is, we consider only those x_{ijt} for which $\bar{p}_{ij}(t_{ij}^*) > 0$. In *step 14* we compute the upper bound $\overline{F}$. Having at hand the upper and lower bound we are able to evaluate the actual solution. This process is repeated as long as the condition in *step 15* is fulfilled. For a finite set of variables the convergence properties of column generation are well known (Desaulniers et al., 2005). In presence of an infinite number of variables the convergence towards the optimal solution has been shown by Charnsethikul (2011). Note, $\underline{F}$ denotes the best known integral lower bound, where $\underline{F} = 0$ is always a feasible lower bound.

Table 1
Column generation procedure to solve the linear relaxation of (9)–(19). Recall that t is the selling time the sales representative located in SCU i allocates to SCU j.

Step 1	ϵ := user defined (allowed LP-gap);
Step 2	η := user defined (allowed gap for the integral solution);
Step 3	G := user defined (maximum number of iterations to solve LP);
Step 4	initialize integral lower bound $\underline{F} := \begin{cases} \underline{F}^0, & \text{if initial MIP-LB is used} \\ 0, & \text{otherwise} \end{cases}$;
Step 5	$\forall\, i,j\ T_{ij} := \begin{cases} t = t_{ij}^0, & \text{if initial MIP-LB is used} \\ t = \text{uniform}(0, 0.01) \cdot \mathfrak{T}, & \text{otherwise} \end{cases}$;
Step 6	$g = 0$;
Step 7	$g = g + 1$;
Step 8	solve restricted master problem (RMP);
Step 9	derive dual variables σ_i and γ_{ij};
Step 10	$\forall i, j \in J_i$: determine t^*;
Step 11	$\forall i, j \in J_i$: compute $\bar{p}(t^*)$;
Step 12	$\forall j$: compute $\bar{p}_j^*$;
Step 13	$\forall i, j \in J_i$: if $(\bar{p}_{ij}(t^*) > 0)$ then $T_{ij} := T_{ij} \cup \{t^*\}$;
Step 14	compute upper bound $\overline{F} = z^* + \sum_j \bar{p}_j^*$;
Step 15	if $\frac{z^*}{\overline{F}} < 1 - \epsilon$ and $g < G$ and $\overline{F} > \underline{F} \cdot (1 + \eta)$ go to *step 7*;

4.6. Branch-and-Bound process

To derive an integer solution we define a standard Branch-and-Bound algorithm based on the location variable y_i and the assignment variable w_{ij}. In each Branch-and-Bound node we solve the linear relaxation of (9)–(19) using the column generation procedure outlined in Section 4.5. Note, we have to consider the fixed variables of the preceding nodes. We branch at first on the location variables y_i. If all location variables are integer then we branch on w_{ij}.

Branching on location variables y_i**.** Now, let y_l be the non-integer variable with the smallest deviation to the next integer value. If $y_l < 0.5$ then we remove l from the set of potential locations I. Otherwise, we set the lower bound of y_l to be one.

Branching on assignment variables w_{ij}**.** If no non-integer variable y_l exists, we branch on w_{uv} with $u \in I$ and $v \in J_u$. We consider three strategies:

1. w_{uv} is the non-integer assignment variable for which
$$\sum_{t \in T_{uv}} t \cdot x_{uvt}^* \geqslant \sum_{t \in T_{ij}} t \cdot x_{ijt}^* \qquad \forall i \in I, \quad j \in J_i | w_{ij} \in (0, 1).$$
Where x_{ijt}^* are the optimal values for x_{ijt} obtained from *step 8* of the column generation procedure described in Section 4.5. That is, we branch on the non-integer assignment variable with the highest selling time. Note, this corresponds to branch preferentially on variables with a high objective function coefficient.
2. w_{uv} is the smallest non-integer assignment variable
3. w_{uv} is the largest non-integer assignment variable

If $w_{uv} < 0.5$ then we remove v from J_u. Otherwise, we remove v from all J_i with $i \neq u$. Thus, in the next LP-solution derived by our column generation procedure y_l, respectively, w_{uv} are forced to be integer.

Lower bound on integral solution. A lower bound, denoted $\underline{F}$, is derived when no non-integer variables y_l and w_{uv} exist (integral solution). Obviously, $y_i = 0$ for all $i \in I$ and $w_{ij} = 0$ for all $i \in I$ and $j \in J_i$ leads to $\underline{F} = 0$. As stated in Section 4.4 we may initialize $\underline{F} = \underline{F}^0$. We store the largest lower bound value and the corresponding solution.

Upper bound. Usually, the first node of a Branch-and-Bound tree takes the number zero. Therefore, by $\overline{F}_0$ we denote the smallest upper bound computed by our column generation procedure in the first node of the Branch-and-Bound tree.

Back tracking. Each node in our Branch-and-Bound tree will be considered at most two times. For example, let us assume that we branch at a node on y_l with $y_l = 0$. Then, if we consider the node again, we branch on y_l with $y_l = 1$. Of course, this will be done for w_{uv} as well where we branch at first towards $w_{uv} = 0$ or towards $w_{uv} = 1$ and then, at second, towards $w_{uv} = 1$ or towards $w_{uv} = 0$, respectively. In fact, we generate two child nodes for any non-integer variable. Back tracking is done when

(a) a first or an improved lower bound $\underline{F}$ has been obtained and/or
(b) the upper bound $\overline{F}$ at a node in our Branch-and-Bound tree is smaller or equal $\underline{F} \cdot (1 + \eta)$ and/or
(c) a node has been considered two times.

Termination. The Branch-and-Price approach terminates when

(a) all nodes are fathomed and/or
(b) a maximum computation time has been reached. The maximum computation time has to be defined by the user.

Table 2
Computational study on the influence of the branching strategy (*BS*) and the MIP-LB in order to determine an initial lower bound $\underline{F}^0$ (cf. Section 4.4) on the performance of the Branch-and-Price approach. For all instances complete implicit enumeration has been achieved (i.e. *St* = 1).

BS	$\lvert T_{ij}\rvert$	*1st LB*	$\underline{F}$	$\overline{F}_0$	ε	$\mathfrak{h}^{\overline{F}_0}$	$\mathfrak{h}^1$	$\mathfrak{h}$	*CG*	*BB*
MIP-LB **not** *applied*										
1	–	7988534.89	7988534.89	7989017.68	0.006043	4	20	32	34	28
2	–	7988519.95	7988519.95	7989017.68	0.006230	4	30	47	47	42
3	–	7988520.50	7988520.50	7989017.68	0.006223	4	49	84	47	42
MIP-LB applied										
1	16	6608532.61	7988534.89	7989017.68	0.006043	4	3	34	34	28
1	160	7982601.98	7982601.98	7990175.34	0.094783	5	9	14	6	1
2	16	6608532.61	7988519.95	7989017.68	0.006230	4	3	50	47	42
2	160	7982601.98	7982601.98	7990175.34	0.094783	5	9	13	6	1
3	16	6608532.61	7988520.50	7989017.68	0.006223	4	7	98	47	42
3	160	7982601.98	7982601.98	7990175.34	0.094783	5	9	14	6	1

R	200
$\lvert I\rvert$	10
f_i	100000
b	0.3
G	5
ϵ	0.01%
η	0.1%
BS	Branching strategy
1st LB	First found lower bound. If MIP-LB is used then the value corresponds to $\underline{F}^0$
$\underline{F}$	Lower bound
$\overline{F}_0$	Upper bound in Branch-and-Bound node 0
$\mathfrak{h}$	CPU-time in seconds
$\mathfrak{h}^1$	CPU-time in seconds needed to find a first lower bound (LB)
$\mathfrak{h}^{\overline{F}_0}$	CPU-time in seconds needed to compute upper bound in the root node $\overline{F}_0$
CG	Total number of column generation iterations
BB	Total number of iterations of the Branch-and-Bound algorithm

The maximum gap allowed for Branch-and-Bound in GAMS/CPLEX to solve the MIP is 1%

5. Computational experimentation

First, we discuss some general issues concerning our computational studies. Then we analyze the branching strategies and the MIP-LB (initial lower bound) with respect to the solution gap and the performance of our approach. The performance is measured mainly by the computational effort – $\mathfrak{h}$ as CPU-time in seconds – and the realized solution gap

$$\varepsilon = \frac{\overline{F}_0 - \underline{F}}{\overline{F}_0} \cdot 100$$

Moreover, we consider the total number of column generation iterations (*CG*) and the total number of Branch-and-Bound iterations (*BB*) and whether a complete implicit enumeration could be achieved (*St* = 1). From this we continue with a study on the coherence of the problem size and the computational effort. The third study examines the influence of the calling time elasticity b on the performance of our proposed approach. In order to check whether our approach is stable towards parameter variation we randomize the calling time profitability n_j. Then, we discuss the solution of a selected instance in terms of contiguous sales territory design. Finally, we show how our proposed approach can be employed in the decision making process and we discuss some managerial insights.

5.1. Notes, assumptions and definitions

If not stated otherwise, we implement the approach in the algebraic modeling language GAMS[2] 23.5.2 and solve all problems with Cplex 12.2 on a 64-bit Windows Server 2008 (VMWare) with Intel Xeon with 2.13 gigahertz processor and 8 gigabytes RAM. As sales coverage units we use the 532 NUTS-3 statistical regions of Germany. The maximum distance from the north of Germany to the south of Germany is roughly 1000 kilometers. From east to west the maximum distance is 800 kilometers (nearly 360,000 square kilometer). As an application we imagine a company that is selling medical and pharmaceutical products to practicing resident physicians and medical practices.

Specification of $p_{ij}(t)$. For each region j the number of medical doctors n_j, is taken from the *INFAS Geodaten KGS22, 2009, Firmenzähler* data set.[3] Denoting by e_{ij} the euclidean distance in kilometers between the centroids of the regions i and j then the fraction of travel time is computed as follows:

$$\beta_{ij} = 0.1 + \min\{0.9, 2 \cdot e_{ij}/800\}.$$

That is, we assume travel time is proportional to distance. Of course, we might use street network distances. This would enhance the accuracy. However, in order to show the suitability of our new approach we are convinced the crow-fly distances are suitable. We assume an average speed of 100 kilometer per hour by car in order to make the computation of the corresponding parameters easily comprehensible. Obviously, if we assume 8 hours of work a day, there is a maximum of 800 kilometer a sales representative could travel. Thus, if $i = j$ then 90% of the selling time is calling time. Moreover, if the SCU j is 360 kilometer or more away from the location in SCU i there is no calling time left. We arbitrarily choose $\alpha \cdot \mu = 10$. Travel costs are ignored ($o_{ij} = 0$).

The total available selling time $\mathfrak{T}$ is assumed to be 1600 hours per year (based on the assumption of 250 work days a year with 50 days have to be spent in the office). Moreover, we define a radius R (proportional to the limited travel-times) that is crucial in

[2] General Algebraic Modeling System: http://www.gams.com.

[3] http://www.infas-geodaten.de.

Table 3
Computational study on the coherences between problem size $(R, |I|)$ and performance $(\eta, \varepsilon, \mathfrak{h})$. We employ our MIP-LB with $|T_{ij}| = 160$ in order to determine an initial lower bound $\underline{F}^0$. The maximum computational time allowed is 21,600 seconds if η = 0.1% and 36,000 seconds if $\eta = 0.01\%$. The maximum gap allowed for Branch-and-Bound in GAMS/CPLEX to solve the MIP is 0.1%

R	$\|I\|$	η	$\mathfrak{h}$	$\underline{F}$	ε	St	$\sum_{i\in I} y_i$
100	10	1	4	6894109.87	0.400746	1	10
100	10	0.1	5	6894109.87	0.028040	1	10
100	10	0.01	8	6894109.87	0.001335	1	10
100	25	1	60	8571699.62	0.520645	1	25
100	25	0.1	80	8571699.62	0.037768	1	25
100	25	0.01	36000	8571699.62	0.126137	–	25
100	50	1	113	9338385.31	0.837946	1	34
100	50	0.1	315	9338385.31	0.562094	1	34
100	50	0.01	313	9338385.31	0.562094	1	34
200	10	1	9	7982601.98	0.826769	1	10
200	10	0.1	11	7982601.98	0.094783	1	10
200	10	0.01	46	7988507.10	0.008615	1	10
200	25	1	48	9236932.00	0.435091	1	25
200	25	0.1	52	9236932.00	0.088472	1	25
200	25	0.01	36,000	9236932.00	0.088472	–	25
200	50	1	420	9607868.99	0.616720	1	35
200	50	0.1	880	9607868.99	0.542271	1	35
200	50	0.01	965	9607868.99	0.542271	1	35
300	10	1	17	7985624.81	0.293639	1	10
300	10	0.1	19	7985624.81	0.097235	1	10
300	10	0.01	89	7990450.49	0.003900	1	10
300	25	1	62	9235345.81	0.739930	1	25
300	25	0.1	21,600	9235345.81	0.138291	–	25
300	25	0.01	36,000	9235345.81	0.138291	–	25
300	50	1	577	9617093.71	0.605881	1	35
300	50	0.1	1261	9617093.71	0.474749	1	35
300	50	0.01	1273	9617093.71	0.474749	1	35

f_i	100,000
b	0.3
G	10
ϵ	0.01%
BS	1
η	Allowed gap in % (cf. Section 4.6)
$\mathfrak{h}$	CPU-time in seconds
$\underline{F}$	Lower bound
St	Status = 1: complete implicit enumeration

order to generate the set J_i of SCUs assignable to a sales representative in SCU i:

$$j \in J_i | e_{ij} \leqslant R.$$

Cuts and unstable column generation. Pre-tests have shown that for the original MP the column generation procedure is unstable for certain instances in terms of pricing out. If we remove (12) from the MP than this problem does not occur. We assume that there are many restrictions due to (12) that are not tight. These might cause the instabilities. Therefore, we remove (12) from the model.

5.2. Results

Branching strategy and initial lower bound. The main concern of this study is to analyze the branching strategy (BS) and the MIP-LB to determine an initial lower bound. We are interested in the effect on the performance of our approach. The parameter settings and results are shown in Table 2. For $|T_{ij}| = 16$ the elements of T_{ij} are $\{100, 200, \ldots, 1600\}$ and for $|T_{ij}| = 160$ the elements of T_{ij} are $\{10, 20, \ldots, 1600\}$. Obviously, if we employ the initial MIP-LB we are able to determine a solution very quickly - particularly if the cardinality of T_{ij} is large (i.e., 160): We see that $\underline{F}^0$ satisfies tolerances (here η = 0.1%) if $|T_{ij}| = 160$. This is not true, if $|T_{ij}| = 16$. Therefore, the number of CG and BB is smaller for $|T_{ij}| = 160$ and hence the total computing time $\mathfrak{h}$ is smaller as well. However, the solution quality (in terms of $\underline{F}$ and ε) is better if we do not apply the MIP-LB. That is, because of the first lower bound (*1st LB*) found by the approach without MIP-LB (i) satisfies the quality tolerance and (ii) is larger than *1st LB* found by using the MIP-LB. This finding corresponds to the result that $\underline{F}$ is larger for $|T_{ij}| = 16$ than for $|T_{ij}| = 160$: The Branch-and-Price procedure improves the initial lower bound found by the MIP-LB if $|T_{ij}| = 16$. Concerning the branching strategy (cf. Section 4.6) we observe that BS = 1 seems to be most promising.

Problem size and solution quality. In this study we analyze the influence of the problem size – measured by R and $|I|$ – and the solution quality tolerance η (cf. Section 4.6) on the performance of our approach. Here we measure the performance by $\underline{F}, \mathfrak{h}, \varepsilon$ and St. Based on the results of the first study (cf. Table 2) we choose branching strategy BS = 1 and our MIP-LB in order to determine an initial lower bound $\underline{F}^0$. Parameter settings and results can be found in Table 3. Obviously, $\underline{F}$ grows with increasing R and $|I|$. Generally, we can say the larger the problem, the higher is the computational effort ($\mathfrak{h}$). However, for $|I| = 25$ we observe difficulties to achieve complete implicit enumeration (St = 1). We assume, that the contiguity constraints (14) and (15) might cause these problems. Note that if complete implicit enumeration is achieved (St = 1), the realized gap is the minimum of η and ε. For example consider the last three instances of Table 3. The first gap is 0.606 while the second and the third gap is 0.1 and 0.01 respectively. However, if we do not manage to achieve St = 1 within the maximum allowed time, than the gap to be considered is ε (cf. Section 4). Only for two instances the decline in η yields an improved lower bound $\underline{F}$.

Comparison with Drexl and Haase (1999). In order to compare our approach with the approach presented in Drexl and Haase

Table 4
Comparison to Drexl and Haase (1999). The computational effort for the heuristic of Drexl and Haase (1999) is below one second for all instances. GAP denotes the minimum of $\eta = 0.1\%$ and ε if complete implicit enumeration is achieved (see Section 5.2, computational study on problem size). The results given are the averages over ten randomly generated instances.

$\|J\|$	$\|I\|$	$\underline{F}$	$\overline{F}_0$	ε	GAP	$\mathfrak{h}$	Drexl and Haase (1999)		
							lower bound	ε	DEV
100	10	24875.54	25047.15	0.68	0.48	1451.3	24747.03	1.20	0.52
	25	27132.65	27430.12	1.08	0.10	27.2	26708.50	2.63	1.53
	50	28279.19	28680.06	1.39	0.10	266.5	26884.10	6.26	5.07
250	10	55007.89	55088.69	0.15	0.15	1320.7	55035.26	0.10	-0.05
	25	64223.32	64464.95	0.37	0.35	4336.1	63994.86	0.73	0.35
	50	68090.11	68519.40	0.62	0.29	1656.8	67152.42	2.00	1.38
500	10	95998.22	96456.23	0.47	0.47	5793.0	96317.75	0.14	-0.33
	25	120767.04	121393.71	0.51	0.51	5771.6	120855.82	0.44	-0.07
	50	133534.17	134414.86	0.65	0.65	6562.9	133221.89	0.89	0.23

$\text{DEV} = 100 \times (\underline{F} - \text{lower bound}) / \overline{F}_0$.

(1999) we use the data generating process outlined in Drexl and Haase (1999). We consider problem sets of different size, i.e. $|J| \in \{100, 250, 500\}$ and $|I| \in \{10, 25, 50\}$. Further, we consider $b = 0.3$, $\mathfrak{T} = 1300$, $\epsilon = 0.01$ %, $f_i \in [750, 1250]$, and $\eta = 0.1\%$. For each problem set we consider ten randomly generated instances. The maximum computation time is set to two hours. The Branch-and-Bound procedure is implemented in GAMS/CPLEX while the heuristic procedure of Drexl and Haase (1999) is implemented in the programming language C. We use an iMac with 3.4 gigahertz Intel Core i7 and 16 gigabytes RAM with Mac OS X. The results are given in Table 4. As expected, our approach generally performs better than the heuristic by Drexl and Haase (1999). However, we find that the heuristic approach performs quite good: the deviation from the best lower bound is at most 6%. Actually, it seems that for

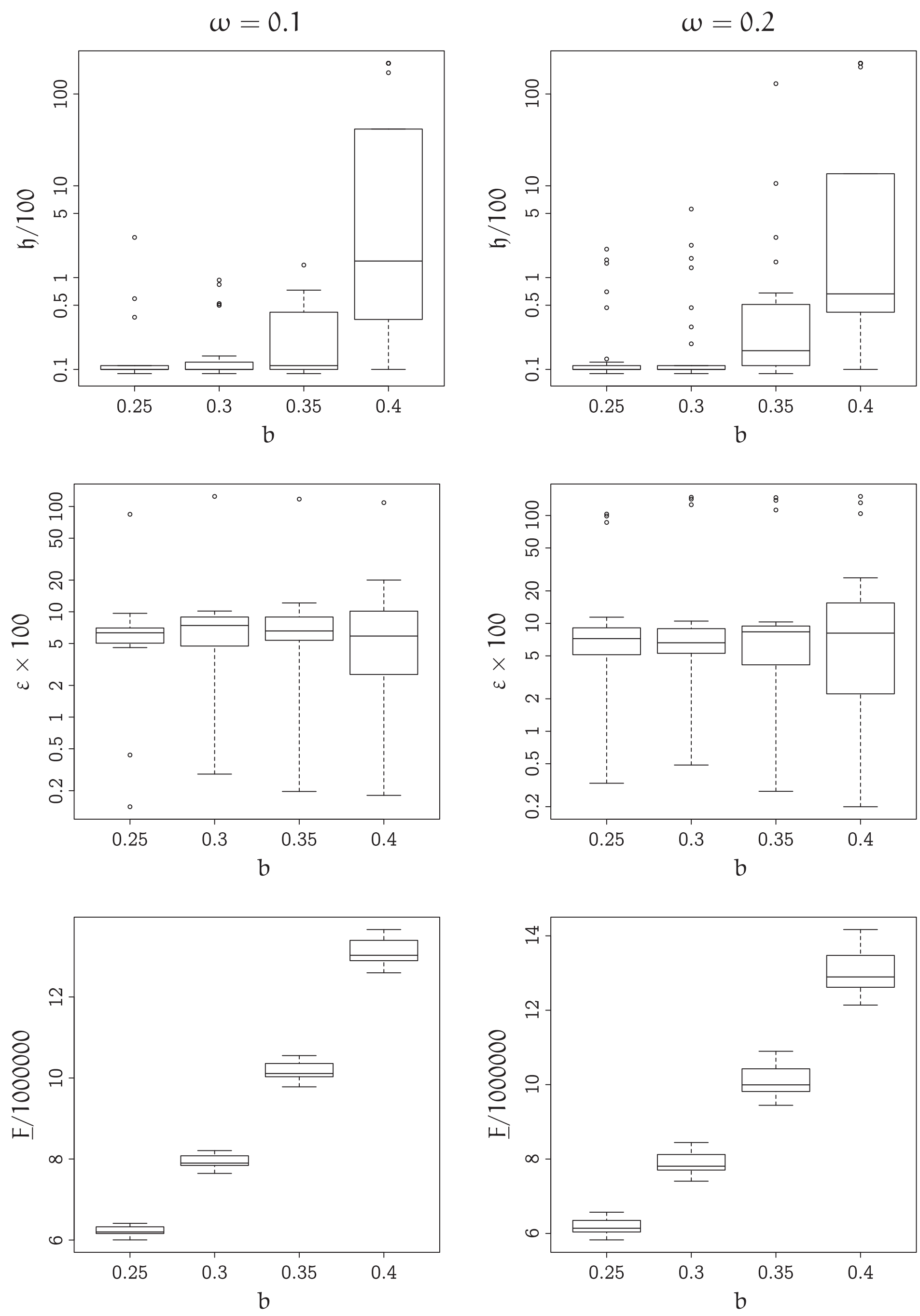
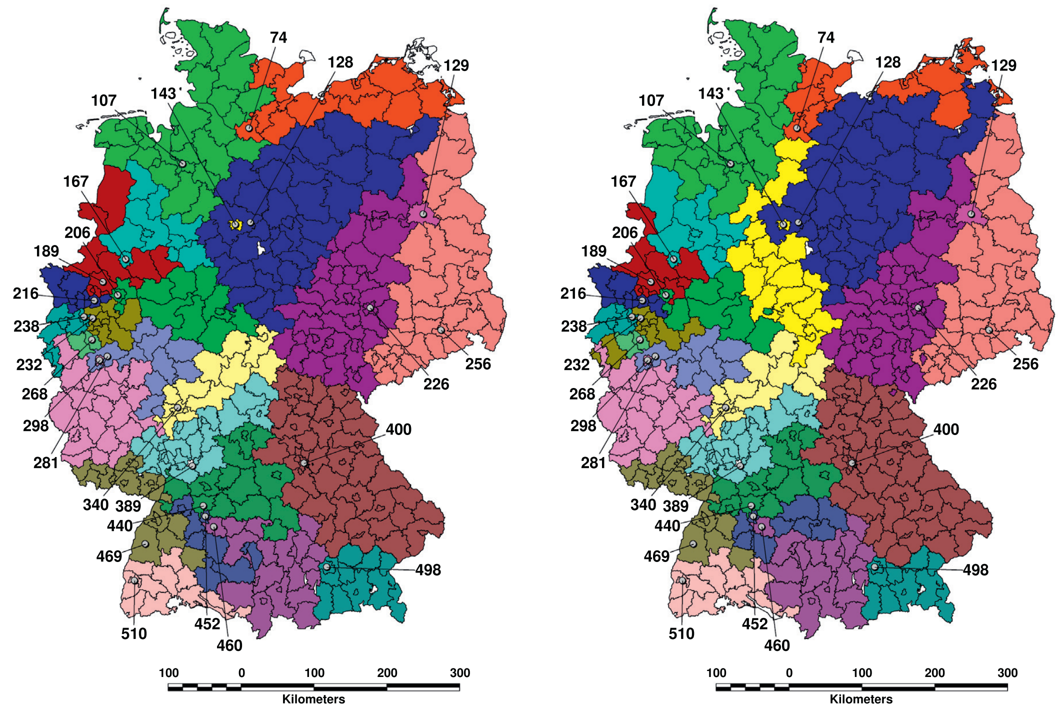


Fig. 2. Influence of calling time elasticity b on performance under randomized calling time profitability $\tilde{n}_j$. The box-plots represent the results of 30 instances of $\tilde{n}_j$. Parameters given: $R = 200$, $|I| = 10$, $f_i = 100{,}000$, $\epsilon = 0.01\%$, η=0.1%, $G = 10$, BS=1. MIP-LB used with $|T_{ij}| = 160$. Maximum gap allowed for MIP: 0.1%. Maximum computational time: 21,600 seconds. The ordinate of $\mathfrak{h}$ and ε are log-scaled.



	contiguity constraints	no contiguity constraints
$\underline{F}$	9235345.81	9245634.20
$\mathfrak{h}$	21600	50
St	-	1

Fig. 3. Sales territory design: The effect of the contiguity constraints (14) and (15) on resulting sales territories and selected measures. Parameter settings: $R = 300$, $|I| = 25$, $f_i = 100{,}000$, $b = 0.3$, BS = 1, $G = 10$, $\epsilon = 0.01\%$, $\eta = 0.1\%$. The maximum computational time allowed is 21,600 seconds. We employ the MIP-LB to determine an initial lower bound using $|T_{ij}| = 160$ and a maximum gap allowed for Branch-and-Bound in GAMS/CPLEX to solve the MIP of 0.1%.

problem sets with many SCUs (i.e., large $|J|$) the heuristic approach is advantageous as long as we do not consider many potential locations (i.e. large $|I|$). The comparison with Drexl and Haase (1999) illuminates another compelling argument for our approach: the procedure outlined in Section 4 yields tight upper bounds ($\overline{F}_0$). On average, the solution gap ε achieved by the procedure of Drexl and Haase (1999) in Table 4 is 1.6% (next to the last column). In contrast, Drexl and Haase (1999) report a gap of 3%. The improvement of the gap is due to better upper bounds provided by our procedure. The tight upper bounds arise from the column generation procedure and the explicit consideration of contiguity constraints (both, not considered in the approximation methods by Drexl &

Table 5
Computational study on work force size. We use MIP-LB of Section 4.4 only. *GAP* denotes the gap in percent reported by GAMS/CPLEX. *LOST* denotes the potential demand n_j not covered. Settings: $R = 300$, $|I| = 50$, maximum computational time 7200 seconds.

$\|J\|$	$\mathfrak{S}$	$\|T_{ij}\|$	$\underline{F}$	$\mathfrak{h}$	*GAP*	*LOST*
535	10	16	5929607.80	40.03	0.00	109691.90
		160	7539647.13	1126.27	0.00	627.11
	15	16	6843522.00	103.18	0.00	74735.73
		160	8000814.44	1741.06	0.00	415.67
	20	16	7378841.21	104.29	0.00	48176.27
		160	8169390.72	6442.28	0.00	292.41
1070	10	16	8744529.20	194.44	0.00	320682.68
		160	13130724.93	2531.02	0.00	5103.37
	15	16	10647548.52	972.42	0.00	267033.71
		160	14469879.91	7200.00	0.06	824.54
	20	16	12027431.73	1608.87	0.00	220991.89
		160	15257921.55	7200.00	0.05	1192.42
2140	10	16	12637926.29	7200.00	0.06	793901.19
		160	21300122.06	7200.00	0.05	132378.08
	15	16	15369905.36	7200.00	0.10	712236.17
		160	24386187.64	7200.00	0.50	42609.91
	20	16	17592718.90	7200.00	0.03	639969.51
		160	25553094.04	7200.00	4.04	39320.67

Table 6
Influence of locational costs on results. Parameter settings: $R = 300$, $|I| = 50$, $b = 0.3$, $G = 10, \epsilon = 0.01\%$, $\eta = 0.1\%$. The maximum computational time allowed is 21,600 seconds. We employ the MIP-LB to determine an initial lower bound using $|T_{ij}| = 160$ and a maximum gap allowed for Branch-and-Bound in GAMS/CPLEX to solve the MIP of 0.1%.

f_i	100,000	150,000
$\sum_i y_i$	35	24
$\overline{F}^0$	9662968.57	8327799.27
$\underline{F}$	9617093.71	8200707.72
Gap in %	0.1	0.1
$\mathfrak{h}$	1288	1861
St	1	1

$f_i = 100000$	$f_i = 150000$

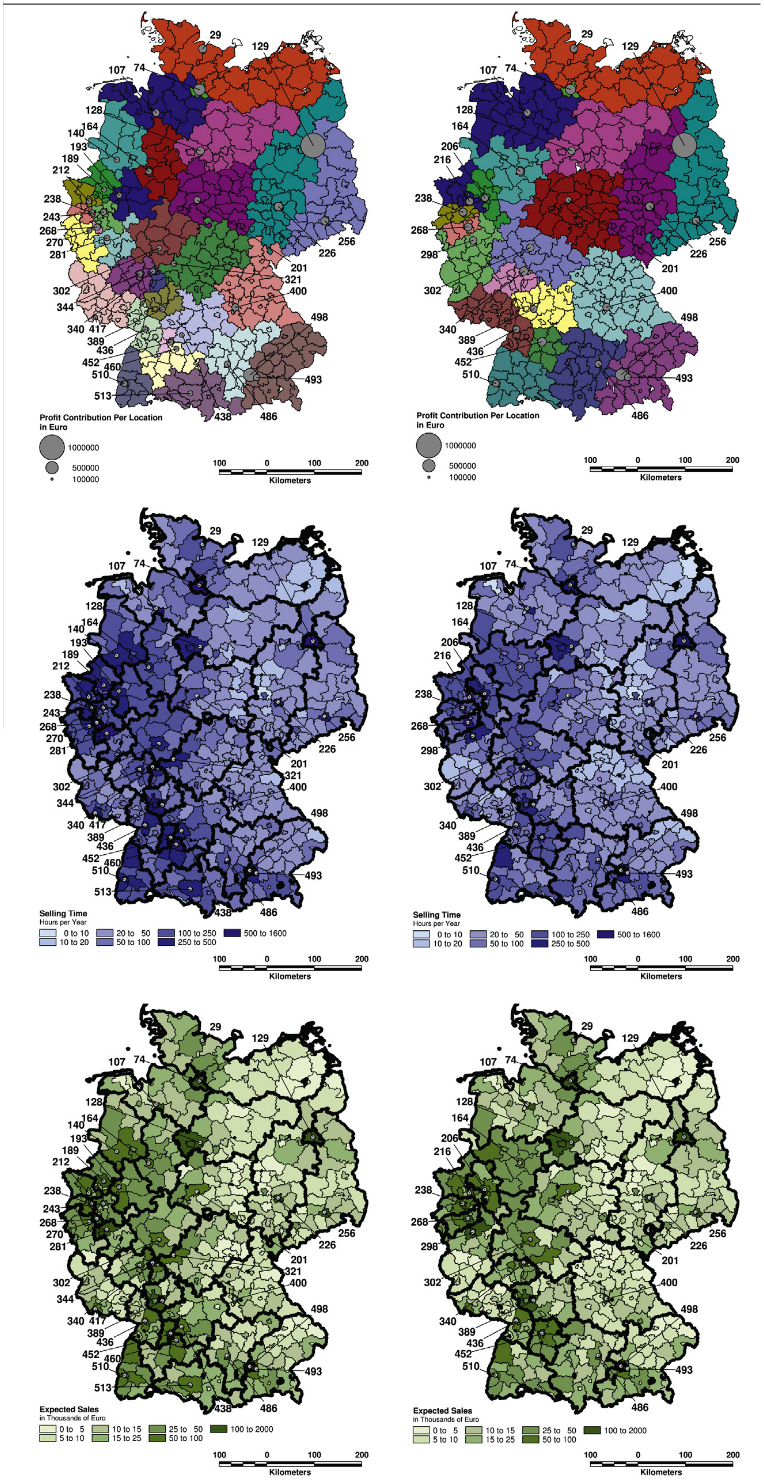


Fig. 4. Spatial patterns of selling time, expected sales and profit contribution. The maps are based on the results of Table 6.

Haase (1999)). The solution gap for our approach is 0.66% in the average (see column GAP in Table 4).

Randomized n_j and varying b. Since the parameters of the profit contribution function (4) are empirically obtained, a company might be uncertain about the reliability of the values. The most crucial parameters are the calling time profitability n_j and the calling time elasticity b. In order to reflect uncertainty about n_j, here we employ a randomized calling time profitability $\tilde{n}_j \in [n_j(1-\omega), n_j(1+\omega)]$. We analyze selected performance measures with varying b and ω. Because we assume that information about potential customers are relatively reliable, we consider $\omega = 0.1$ and $\omega = 0.2$. In terms of calling time elasticity b we consider values within the range of empirical observed values (cf. Skiera & Albers, 2008; Drexl & Haase, 1999 and see Fig. 1). For each value of b we compute 30 instances with random $\tilde{n}_j$. The parameter settings and the results are shown in Fig. 2. Most eye-catching is the dramatic increase in computational effort for $b = 0.4$. However, the solution gap remains stable – although the variance increases with b. As expected the lower bound increases. This is important from a managerial point of view: The variance of $\underline{F}$ is quite small for all b. We witness no remarkable difference in the results for $\omega = 0.1$ and $\omega = 0.2$.

Sales territory design and contiguity property. Now, we are interested in the results concerning the sales territory alignment, particularly the effect of the contiguity constraints (14) and (15). Therefore, we choose a relatively large instance ($R = 300$ and $|I| = 25$). To solve this problem we again employ the initial MIP-LB and branching strategy BS = 1. We solve the same problem two times: (i) with contiguity constraints (14) and (15) and (ii) without these constraints. All parameter settings and the results can be found in Fig. 3. Concerning the computational effort the differences are dramatic. Without contiguity constraints we are able to solve the problem in 50 seconds with complete implicit enumeration. In contrast, with contiguity constraints we do not achieve complete implicit enumeration within the maximum allowed computational time of 21,600 seconds. Of course, $\underline{F}$ is larger for the "non-contiguous model". However, the relative difference is only 0.11%. If we consider the maps in Fig. 3 the effect of the contiguity constraints become apparent: The sales territories of 74, 143, 206, 226, 232, 281, 298, 460 are not contiguous. Most evidently, this is the case for location 74. The price for contiguous sales territories in this example are at least 10,288.39 Euro per year. In contrast to our expectations, we see that some sales territories are far from being compact (340 and 460 for example).

Computational study with some relaxations and larger problem sets. In some applications the number of sales representatives might be fixed. Further, contiguous sales territories might not be compulsory. In order to account for such an application we consider our model (9)–(19) without (14) and (15). We further introduce

$$\sum_{i \in I} y_i = \mathfrak{S} \tag{32}$$

with $\mathfrak{S}$ as the sales force size, i.e. the number of sales representatives. Moreover, managers might be satisfied with an approximation of the selling time and the expected profit. Therefore, we employ the procedure to obtain an initial lower bound as outlined in Section 4.4. Finally, the number of potential accounts or SCUs may be larger than in our case example. Therefore, we artificially enlarge the set of SCUs of the case example and accordingly generate the distances and the demand based on the case example outlined in Section 5.1. The results are given in Table 5.

Sales territories and profit contribution. Finally, we are interested in detailed information about selling times per SCU, expected sales per SCU and expected profit contribution of each sales person. Therefore, we consider two scenarios: we assume fixed locational costs f_i of (i) 100,000 Euro and (ii) 150,000 Euro. Of course, these parameter values are arbitrary. However, these settings provide insights on how a managerial study using our approach could be designed. Parameter settings and summary results are given in Table 6. Maps displaying sales territories, selling times and expected profit contribution as well as expected sales are shown in Fig. 4. Of course profit declines if f_i increases. Here, $\underline{F}$ declines by more than 1.4 million Euro if f_i increases by 50,000 Euro. At the same time the number of established locations (i.e., sales force size) increases if f_i declines. The increase in sales force size results in a more cluttered sales territory design (see Fig. 4). Particularly in the western part of Germany the sales territories are cluttered. As expected the metropolitan areas as Berlin (129), Hamburg (74) and Munich (498) are the locations with the largest profit contribution. These locations are single SCU sales territories. Since the number of medical practices is positively related to the population, we witness highest selling times and expected sales in the most populated areas of Germany.

6. Summary and conclusions

We propose a novel and innovative approach for solving the sales force deployment problem. The four interrelated subproblems sales force sizing, sales representatives location, sales territory alignment, and sales resource allocation are solved simultaneously so that the total profit contribution is maximized. As common in literature, the profit contribution is measured by a separable objective function, i.e., by a sum of strictly concave sales response functions each depending on the continuous measured selling time a sales representative is allocating to a specific sales coverage unit. An important key element of the new approach is that each sales response function is exactly "approximated" by a piecewise linear function where each line segment has the length of zero. This results in a mathematical formulation with an infinite number of binary variables (basis points). As a second new insight to the sales force deployment problem, our approach provides tight upper and lower bounds. These are computed by solving the linear relaxation using column generation. The subproblems are solved analytically. In order to obtain an integral solution we use a Branch-and-Price approach. The third important contribution is that we incorporate explicit contiguity constraints in our model. This results in contiguous sales territories as demanded by sales companies. We evaluate our proposed approach in a series of computational studies. Therefore, we employ a large application-oriented example within the health care and pharmaceutical industry in Germany. The studies show that our approach generates a high-quality solution within reasonable time. The results are stable against parameter variation and we can show the effect of the contiguity constraints. A comparison with Drexl and Haase (1999) shows that we are able to reduce the solution gap from 3% reported in Drexl and Haase (1999) to 1.6% in the average if we use our upper bounds and the approximation method proposed by Drexl and Haase (1999) to compute a lower bound. If we employ our approach we achieve a solution gap of 0.66% in the average using the same artificial data.

Acknowledgments

We thank 5 anonymous reviewers for their helpful comments to improve the paper.

References

Albers, S. (1996). Optimization models for salesforce compensation. *European Journal of Operational Research, 89*, 1–17.

Beckmann, M., & Golob, T. (1972). A critique of entropy and gravity in travel forecasting. In Gordon, F. (Ed.), *Traffic flow and transportation* (pp. 109–117).

Capron, L., & Hulland, J. (1999). Redeployment of brands, sales forces and general marketing management expertise following horizontal acquisitions: A resource-based view. *Journal of Marketing, 63*, 41–54.

Charnsethikul, P. (2011). A column generation approach for the products mix based semi-infinite linear programming model. *International Journal of Management Science and Engineering Management, 6*(2), 106–109.

Churchill, G., Ford, N., & Walker, O. (1993). *Sales force management: Planning, implementation and control* (4th ed.). Irwin, Homewood.

Church, R., & Murray, A. (2009). *Business site selection, location analysis, and GIS*. Wiley and sons.

Cravens, D., & LaForge, R. (1983). Salesforce deployment analysis. *Industrial Marketing Management, 12*, 179–192.

Darmon, R. Y. (2002). Salespeople's management of customer information: Impact on optimal territory and sales force sizes. *European Journal of Operational Research, 137*(1), 162–176.

Desaulniers, G., Desrosiers, J., & Solomon, M. M. (Eds.). (2005). *Column generation. GERAD 25th anniversary series*. Springer.

Drexl, A., & Haase, K. (1999). Fast approximation methods for sales force deployment. *Management Science, 45*(10), 1307–1323.

du Merle, O., Villeneuve, D., Desrosiers, J., & Hansen, P. (1999). Stabilized column generation. *Discrete Mathematics, 194*, 229–237.

Golalikhani, M., & Karwan, M. H. (2013). A hierarchical procedure for multi-skilled sales force spatial planning. *Computers & Operations Research, 40*(5), 1467–1480.

Haase, K., Lange, K., & Missong, M. (1999). *Designing the dealer network of a motor-car producer*. Tech. Rep. 123/1999, University of Kiel, Economics Working Papers.

Howick, R., & Pidd, M. (1990). Sales force deployment models. *European Journal of Operational Research, 48*, 295–310.

Hruschka, H. (2006). Relevance of functional flexibility for heterogeneous sales response models: A comparison of parametric and semi-nonparametric models. *European Journal of Operational Research, 174*(2), 1009–1020.

King, D. M., Jacobson, S. H., Sewell, E. C., & Cho, W. K. T. (2012). Geo-graphs: An efficient model for enforcing contiguity and hole constraints in planar graph partitioning. *Operations Research, 60*(5), 1213–1228.

Lasdon, L. (2002). *Optimization theory for large systems*. Dover Publication.

Lee, C.-Y., & Yang, R. (2013). Compensation plan for competing salespersons under asymmetric information. *European Journal of Operational Research, 227*(3), 570–580.

Lodish, L. (1975). Sales territory alignment to maximize profit. *Journal of Marketing Research, 12*(1), 30–36.

Lodish, L. (1976). Assigning salesmen to accounts to maximize profit. *Journal of Marketing Research, 13*(4), 440–444.

Lodish, L., Curtis, E., Ness, M., & Simpson, M. K. (1988). Sales force sizing and deployment using a decision calculus model at syntex laboratories. *INTERFACES, 18*, 5–20.

Lübbecke, M., & Desrosiers, J. (2005). Selected topics in column generation. *Operations Research, 53*(6), 1007–1023.

Mantrala, M., Albers, S., Caldieraro, F., Jensen, O., Joseph, K., Krafft, M., et al. (2010). Sales force modeling: State of the field and research agenda. *Marketing Letters, 21*(3), 255–272.

Mesak, H. I., & Ellis, T. S. (2009). On the superiority of pulsing under a concave advertising market potential function. *European Journal of Operational Research, 194*(2), 608–627.

Müller, S., Wilhelm, P., & Haase, K. (2013). Spatial dependencies and spatial drift in public transport seasonal ticket revenue data. *Journal of Retailing and Consumer Services, 20*(3), 334–348.

Pinals, D. (2001). *Effective sales force deplyoment: Use or misuse of your most critical resource*. <http://www.alignstar.com/wp2_use_or_misuse.html>.

Rangaswamy, A., Sinah, P., & Zoltners, A. (1990). An integrated model-based approach for sales force structuring. *Marketing Science, 9*(4), 279–298.

Segura-Ramiro, J., Ríos-Mercado, R., Álvarez Socarrás, & de Alba Romenus, K. (2007). A location-allocation heuristic for a territory design problem in a beverage distribution firm. In *Proceedings of the 12th annual international conference on industrial engineering*. Cancun, Mexico.

Shirabe, T. (2005). A model of contiguity for spatial unit allocation. *Geographical Analysis, 37*, 2–16.

Shirabe, T. (2009). Districting modeling with exact contiguity constraints. *Environment and Planning B, 36*, 1053–1066.

Skiera, B., & Albers, S. (1998). COSTA: Contribution optimizing sales territory alignment. *Marketing Science, 17*(3), 196–213.

Skiera, B., & Albers, S. (2008). Prioritizing sales force decision areas for productivity improvements using a core sales response function. *Journal of Personal Selling & Sales Management, 28*(2), 145–154.

Zolterns, A., & Sinah, P. (2005). Sales territory design: Thirty years of modeling and implementation. *Marketing Science, 24*(3), 313–331.

Zoltners, A. (1976). Integer programming models for sales territory alignment to maximize profit. *Journal of Marketing Research, 13*(4), 426–430.

Zoltners, A., Sinah, P., & Lorimer, S. (2009). *Building a winning sales force: Powerful strategies for driving high performance*. AMACOM, New York (No. ISBN 0814410405).

Zoltners, A., & Lorimer, S. (2000). Sales territory alignment: An overlooked productivity tool. *Journal of Personal Selling & Sales Management, 20*(3), 139–150.

Zoltners, A., & Sinah, P. (1983). Sales territory alignment: A review and model. *Management Science, 29*(11), 1237–1256.

2011 ERDKUNDE Vol. 65 · No. 4 · 401–414

ASSESSMENT OF SCHOOL CLOSURES IN URBAN AREAS BY SIMPLE ACCESSIBILITY MEASURES

Sven Müller

With 8 figures
Received 02. March 2010 · Accepted 19. May 2011

Summary: The demographic processes in the eastern regions of Germany have yielded a dramatic decline in student numbers for the time period 1992–2002. This in turn implicates a remarkable school consolidation. In application scenarios, simple measures of the assessment of school closures are needed. In this paper we discuss simple measures of school-accessibility based on public transport travel-times. Moreover, an efficient network flow model to determine the travel-times is presented. Furthermore, a guidance of how a network graph might be constructed is given. As a result of the accessibility analysis, we find that proximate areas are affected by increased travel-times. However, outskirt districts are affected as well. This finding is not obvious, though. The easy-to-understand measures of accessibility presented in this paper might be implemented in the educational planning process. The case study of Dresden is exemplary for other (western) regions in Germany with comparable demographic processes.

Zusammenfassung: Ein wesentliches Merkmal des derzeit diskutierten demographischen Wandels, sind die in einigen Städten Ostdeutschlands zum Teil stark rückläufigen Schülerzahlen. In Folge dieser Entwicklungen werden häufig Schulstandorte geschlossen. Bei der Entscheidung, welcher Standort aufzugeben ist, werden Aspekte der räumlichen Erreichbarkeit zumeist vernachlässigt. Dieser Beitrag diskutiert quantitative Indikatoren, die eine Diskussion über zu schließende Standorte, hinsichtlich zu erwartender Veränderung der Erreichbarkeit bereichern können. Hierbei werden sowohl einfache Erreichbarkeitsindikatoren, wie zum Beispiel die Anzahl der ÖPNV-Linien, die eine schulstandortnahe Haltestelle bedienen, als auch komplexere, aber weiterhin leicht verständliche Erreichbarkeitsmaße, die zusätzlich die Aktivitätenmenge berücksichtigen, verwendet. Darüber hinaus wird ein effizientes Kürzeste-Wege-Problem zur Ermittlung der Reisezeiten sowie die Vorgehensweise zur Graphkonstruktion des Liniennetzes vorgestellt. Im Rahmen der Analysen zeigt sich, dass trotz einer Vielzahl an Schulschließungen, für die Gymnasien der Stadt Dresden eine gute Erreichbarkeit erhalten bleibt. Gleichwohl werden Wohnorte, die bereits vor Schulschließung verhältnismäßig schlechte Erreichbarkeitswerte aufwiesen überproportional von den Verschlechterungen der Erreichbarkeit in Folge der Schulschließungen getroffen.

Keywords: Accessibility, travel-times, public transport, shortest-path problem, graph construction, geographic information systems, school closure, urban areas

1 Introduction

During the time period 1995–2005, the city of Dresden, Germany, faced the problem of a dramatic decline (30%) in student numbers. Recently however, the student numbers have increased slightly. This phenomenon is typical for many urban areas in eastern regions of Germany (except for Berlin). Consequently, a lot of schools have been closed in recent years. It is a very difficult task to decide which schools should be closed. Often the political debate on this topic is focused on a few selected criteria like locational costs. Notably, in real world applications, a complex view of the problem is needed. Unfortunately, this is unusual due to the complexity of the related measures and models (Müller 2010; Müller et al. 2009). In this paper we'd like to propose simple measures that can help to decide school closures (and openings as well). We focus on measures that describe the accessibility of Gymnasium-schools to students in terms of travel-time. It is well known that long travel-times have a wide range of negative impacts on students (Talen 2001). Therefore, we'd like to investigate spatial equity of accessibility variation due to school consolidation.

We consider Gymnasium-schools only, because the public debate concentrates on these (Peter 2004; Peter 2005; Klameth 2005; Richter 2005). Roughly speaking, a Gymnasium-school is equivalent to an American high-school that qualifies for university study. For more details, see Müller (2009) and Müller (2008). Since nearly 70% of all students who qualify for a Gymnasium-school use public transport on their commute to school, we will con-

DOI: 10.3112/erdkunde.2011.04.06 ISSN 0014-0015 http://www.erdkunde.uni-bonn.de

sider public transport in our analysis (Müller 2006). Our study area is the city of Dresden, Germany. Commute to school flows from the city of Dresden to surrounding regions and vice versa are negligible (Müller 2009). All measures presented here can be used as single indicators in order to assess the effects of school closures. However, one might apply these measures to location-allocation models as well (Leonardi 1978). The remainder of the paper is structured as follows: In section 2, we define accessibility. In section 3, an efficient shortest-path problem is presented in order to determine travel-times on the commute to school. Note that efficient solutions to shortest-path problems are one major issue in recent accessibility research (Kwan et al. 2003). Section 4 comprises the construction of the underlying network graph. The discussion of the measures to assess school closures and the application to the city of Dresden can be found in section 5. Finally, a conclusion is given in section 6.

2 What is accessibility and how can we measure it?

A very general definition of accessibility can be found in Rodrigue et al. (2009). They define accessibility as a key element to transport geography, and to geography in general, since it is a direct expression of mobility either in terms of people, freight or information. Well-developed and efficient transportation systems offer high levels of accessibility (if the impacts of congestion are excluded), while less-developed ones have lower levels of accessibility. Thus, accessibility is linked to an array of economic and social opportunities. Accessibility is defined as the measure of the capacity of a location to be reached by, or to reach different locations. Therefore, the capacity and the arrangement of transport infrastructure are key elements in the determination of accessibility. All locations are not equal because some are more accessible than others; this implies inequalities. The notion of accessibility consequently relies on two core concepts: The first is location, where the relativity of space is estimated in relation to transport infrastructures, since they offer the mean to support movements. The second is distance, which is derived from the connectivity between locations. Connectivity can only exist when there is a possibility to link two locations through transportation. It expresses the friction of distance. The location that has the least friction relative to others is likely to be the most accessible. Commonly, distance is expressed in units such as in kilometers or in time, but variables such as cost or energy spent can also be used.

Now the question arises: How should accessibility be measured? As Kwan (1998) states, conventional accessibility measures are based on three fundamental elements. First, a reference location serves as the point from which access to one or more other locations is evaluated. The reference location most often used is the home location of an individual, or the zone where an individual's home is located when zone-based data are used. Second, a set of destinations in the urban environment is specified as the relevant opportunities (here schools) for the measure to be enumerated. Further, each opportunity may be weighted to reflect its importance or attractiveness. Third, the effect of the physical separation between the reference location and the set of urban opportunities upon such access is modeled by an impedance function, which represents the effect of distance decay on the attractiveness of the relevant opportunities. Based on these three elements, various types of accessibility measures can be specified (Brunisma and Rietveld 1998): In general, we consider relative and integral measures of accessibility. Relative accessibility measures describe the degree of connection between two locations. They are expressed in terms of the presence or absence of a transport link, or the physical distance or travel time between two locations. Integral measures, on the other hand, represent the degree of interconnection between a particular reference location and all, or a set of, other locations in the study area. When impedance between the reference location and the other locations is expressed in the form of a distance decay function similar to those found in gravity models, the access measure is a gravity-based measure. In the case where an indicator function is used as the impedance function to exclude opportunities beyond a given distance limit, the measure is a cumulative-opportunity measure. This measure indicates how many opportunities are accessible within a given travel time or distance from the reference location. A further distinction can be made depending on whether an access index is enumerated and used as an indicator of physical or place accessibility (how easily a place can be reached or accessed by other places), or personal or individual accessibility (how easily a person can reach activity locations). For more details, see Kwan (1999).

One important area in applied accessibility research is the provision of social services such as hospitals, clinics, senior centers, parks and schools. Studies within this research area evaluate whether access to a particular social service is socially equitable or discriminatory, and seek to identify areas of service deprivation that need special attention (Kwan et al. 2003). However, as Talen (2001) asserts, there is scant research and experience devoted to school accessibility. In order to rectify this lack of literature, we will consider simple relative and integral measures of school accessibility here. Since most of these measures are based on travel-time, we elucidate how travel-times can be computed efficiently.

3 An efficient network flow model for the shortest-path problem

We assume, that students who commute to school by public transport choose the shortest-path (in terms of travel-time) from home (reference location) to school (destination). In order to determine the shortest-path, we can use either a shortest-path algorithm or a network flow model (Domschke and Drexl 2005; Longley et al. 2001; Ahuja et al. 1993). In both cases, the length of the shortest-path in minutes of travel-time is the weighted sum of arcs or edges of the shortest-path. Note that travel-time includes access- and egress-time, in-vehicle- and waiting-time as well as transfer-time. Here we will regard a network flow model, by considering a graph that comprises the public transport network, the schools, and the students' homes. How we actually construct such a graph is described in detail in section 4. From a theoretical viewpoint, the graph consists of nodes $i \in V$ and arcs or edges $(i, j) \in \vec{E}$ that connect the nodes i and j. Moreover, $(i, j) \in \vec{E}$ are weighted by δ_{ij}, which is the travel-time in our case. Our mathematical program (i.e., model) determines the paths from a given reference location $q \in Q$ to **all** destinations $s \in S$ with minimum weights δ_{ij} of the corresponding arcs or edges $(i, j) \in \vec{E}$. $Q \subseteq V$ and $S \subseteq V$. We introduce the positive variable X_{ij}, which is the flow from node $i \in V$ to node $j \in V$. "Flow" has to be interpreted from a theoretical viewpoint (i.e., we do not mean real entities like students). Now we define the objective as

$$\min \sum_{(i,j) \in \vec{E}} \delta_{ij} X_{ij} \qquad (1)$$

such that

$$\sum_{i|(i,k) \in \vec{E}} X_{ik} - \sum_{j|(k,j) \in \vec{E}} X_{kj} = \begin{cases} -|S| & k=q \\ 1 & \text{else} \end{cases} \qquad k \in v \qquad (2)$$

and

$$X_{ij} \geq 0 \qquad (i,j) \in \vec{E} \qquad (3)$$

The objective (1) minimizes the travel-time between a given q and all s. The flow constraints (2) guarantee a contiguous path from q to s. Therefore we assume that one entity per destination s departs from source q. So if $k = q$, we need exactly the amount of entities that equals the demand of all destinations s. That is $\sum_{j|(k,j) \in \vec{E}} X_{kj} = |S|$ for $k = q$. For all other nodes, there is one more in-flow entity than out-flow entity. Hence $\sum_{i|(i,k) \in \vec{E}} X_{ik} - \sum_{j|(k,j) \in \vec{E}} X_{kj} = 1$ for $k \neq q$.

The shortest-path problem outlined here is efficient in various ways: (i) the domain (3) of our variable X_{ij} is $\mathbb{Q}^{\geq 0}$. Hence, our model is a linear program. However, due to the special structure of the model (i.e., the flow-constraints (2)) X_{ij} takes either the value 1 or the value 0 in the solution. This enables the use of a powerful network simplex method in order to solve our problem optimally and efficiently. (ii) Our model is called a single-source-shortest-path-problem (SSSP), because we compute each shortest-path to all destinations s of one source q. Thus, in order to compute a travel-time matrix between all pairs of q and s we need to solve the problem $|Q|$-times. A single-pair-shortest-path-problem (SPSP) determines the shortest-path between one q and one s. Hence, SPSP has to be solved $|Q| x (|S| - 1)$ times. The expected computation time for a $q \times s$ travel-time matrix is remarkably lower for the SSSP compared to the SPSP (Cormen et al. 2001). (iii) in our specific situation, we know that $|S| < |Q|$, i.e., we have less schools s than student locations q. So if we switch q and s, the number of repetitions of our problem reduces from $|Q|$-times to $|S|$-times.

4 How is a network graph set up?

Generally, we have to consider three steps in order to set up a comprehensive graph for the determination of students' travel-times on the commute to school. First, we have to set up a graph of the public transport network of the city of Dresden. We considered lines and routes of the time-period of interest 2002–2008.[1] We only considered routes valid for weekdays between 6:00 AM and 3:00 PM, which is the peak time-period in school commuting. However, we accounted for special routes due to commute to school flows. At this point the nodes of our graph are stops (tram, bus and rail) and the arcs are the connections between these nodes. All arcs are weighted by the drive time between two adja-

[1] Most school closures have taken place in this period.

cent nodes. So far, we have been able to compute the in-vehicle travel-time only. In order to account for transfer-times we considered the schedule of all bus-, tram- and rail-lines. We considered multiple nodes at stops where interchange between different lines is possible. Moreover, we accounted for the opposite direction of a given line (inbound and outbound) as well. Therefore, a given stop might consist of multiple virtual nodes (see Fig. 1). The arcs between the nodes of the same stop are weighted by the difference of arrival-times of the two lines considered. Note that if these time differences changed within the time-period (6:00 AM to 3:00 PM), we computed the average of these values.

In a second step we have to connect the school locations and the locations of the students to public transport stops. Therefore, we buffered all stops with an arc radius of 800 meters using a standard geographic information system. All schools within the buffer of a given stop are connected to the respective stop. As a result, schools might be connected to more than one bus stop. This is necessary, since students

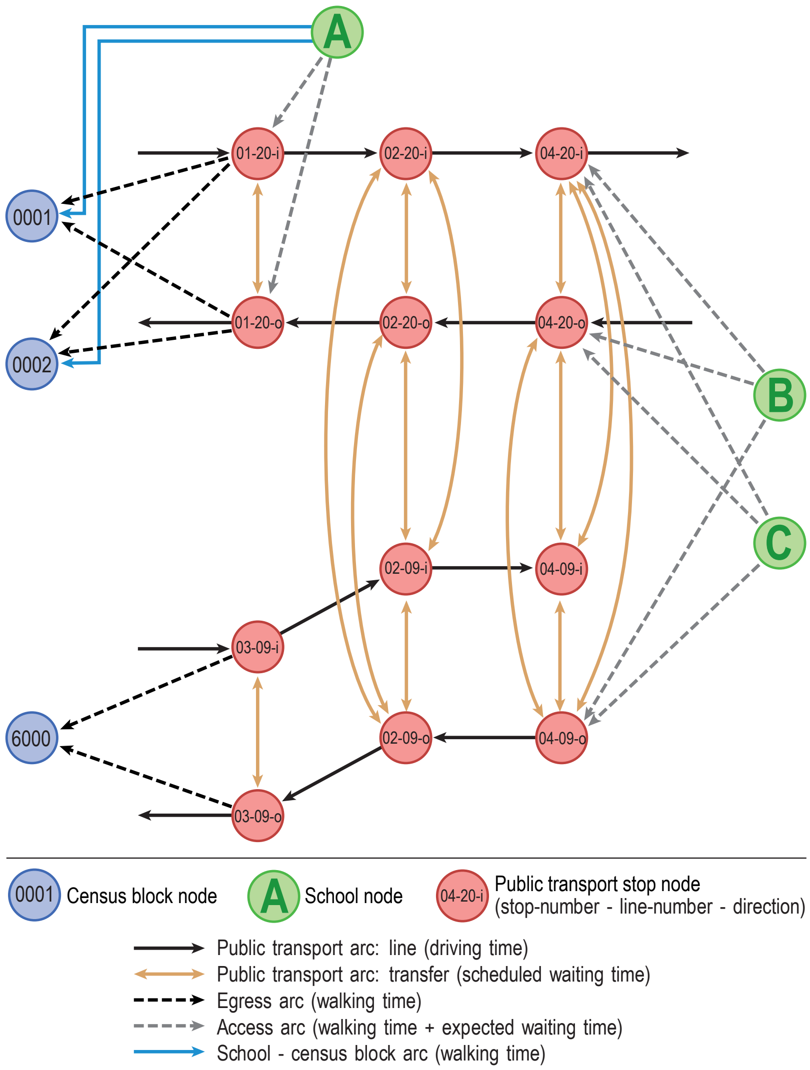


Fig. 1: Graphical representation of the arcs and nodes used for the network graph. We use multiple nodes for stops which are served by two or more lines (stops 2 and 4 for example). For a more general construction we might add the opposite direction for the access and egress arcs

commute to school from different directions using different lines that might terminate at different stops close to the respective school. If a school is not located within any buffer, we increased the arc radius of proximate stops stepwise as long as the respective school is connected to at least one stop. Note that we accounted for physical barriers like embankments and rivers in order to make sure the schools are accessible from the stop (see Fig. 2). The results of this procedure are additional nodes (schools) and arcs (connection between school and stops) of our graph. These arcs are weighted by

$$\delta_{ij} = \sigma(|a_i - a_j| + |b_i - b_j|) \qquad (4)$$

where either $i \in S$ or $j \in S$. σ is the assumed travel-speed by foot (here: 1.49 meters per second) and a_i, a_j, b_i and b_j are geographic coordinates of the nodes i and j. The Manhattan metric of (4) accounts for detours in an urban environment.

The last step consists of the assignment of the students' homes to departure stops. Since we do not have the exact addresses of the students, we use aggregated student numbers at the geographical scale of census blocks. The city of Dresden is subdivided into more than 6 400 census blocks. The area of an average census block is nearly 0.05 square kilometers. We assigned the centroid of each census block to at least one stop. Therefore, we used the same procedure as for the assignment of the schools to stops. However, the weighting of the resulting arcs is different. Namely

$$\delta_{ij} = \sigma(|a_i - a_j| + |b_i - b_j|) + a(1 - \beta \exp(\gamma Z)) \qquad (5)$$

where either $i \in Q$ or $j \in Q$. The term $a(1 - \beta \exp(\gamma Z))$ is the expected waiting time at the departure stop with Z as the headway. Parameters a, β and γ have to be determined empirically. Here we assume that students have precise information about the public transport schedule and, hence, we set the maximum waiting time to 8 minutes (i.e., $a = 8$). Values for β and γ (1.1045 and -0.0852 respectively) are taken from Grosse (2003). The expected waiting time dependent on the headway is shown in figure 3. Finally, if a census block is located within the 800 meters buffer of a given school, this census block is directly assigned to the respective school. That is because given a distance of 800 meters or less, the probability of commuting to school by foot or bike is higher than for any other transport means (Müller et al. 2008).

We used this graph with Q as the schools (24) and S as the census blocks (more than 6 400) and employed our model from section 3 in order to determine a $Q \times S$ travel-time matrix. To do so, we implemented the model in GAMS Version 22.2 (www.gams.com). The CPU-time with a Pentium 4 and 3 GHz and 2 GB DDR Ram under OS Windows XP is nearly 40 minutes.

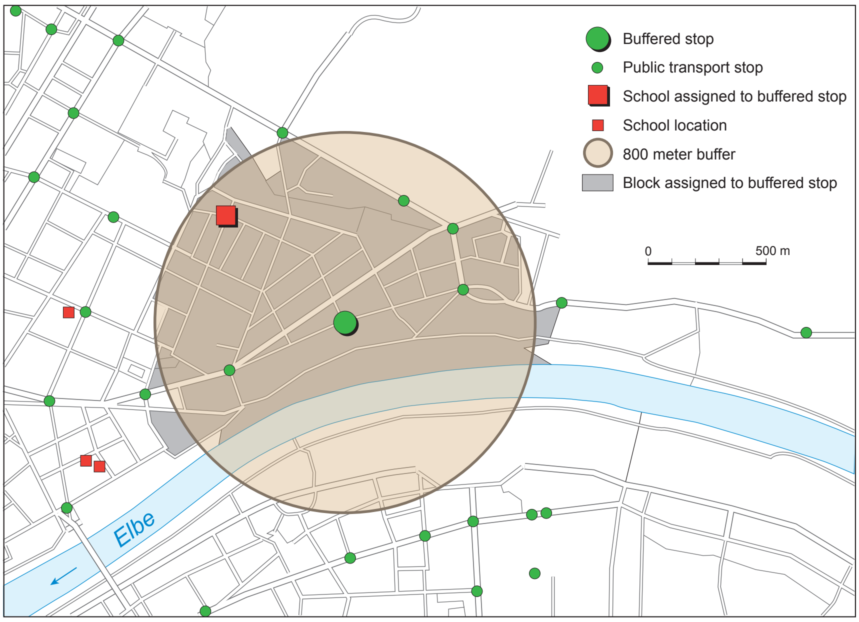


Fig. 2: Assignment of census blocks and schools to public transport stops

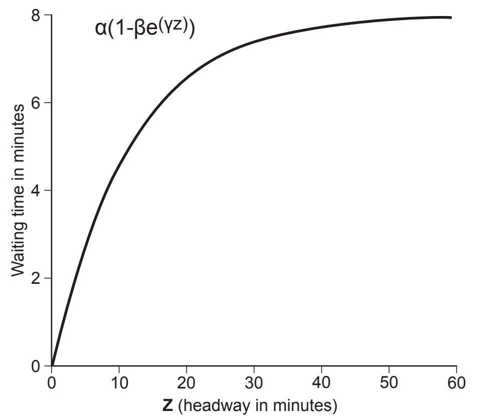


Fig. 3: Expected waiting time dependent on line headway Z: α(1 - β exp (γZ))

5 Impact of school closures on accessibility

Following the definitions given in section 2, here we discuss two relative measures and three cumulative-opportunity measures. All measures have in common that they are quite simple to understand and to compute (we can employ standard GIS-techniques). As a reference, figure 4 shows the spatial population pattern of the city of Dresden for the year 2002 – the base year of our analysis.

5.1 Relative measures of accessibility

A very general measure of accessibility of Gymnasium-schools, which is related to service quality, is the minimum travel-time. This measure is based on the assumption that the probability of enrolling in the nearest school is highest compared

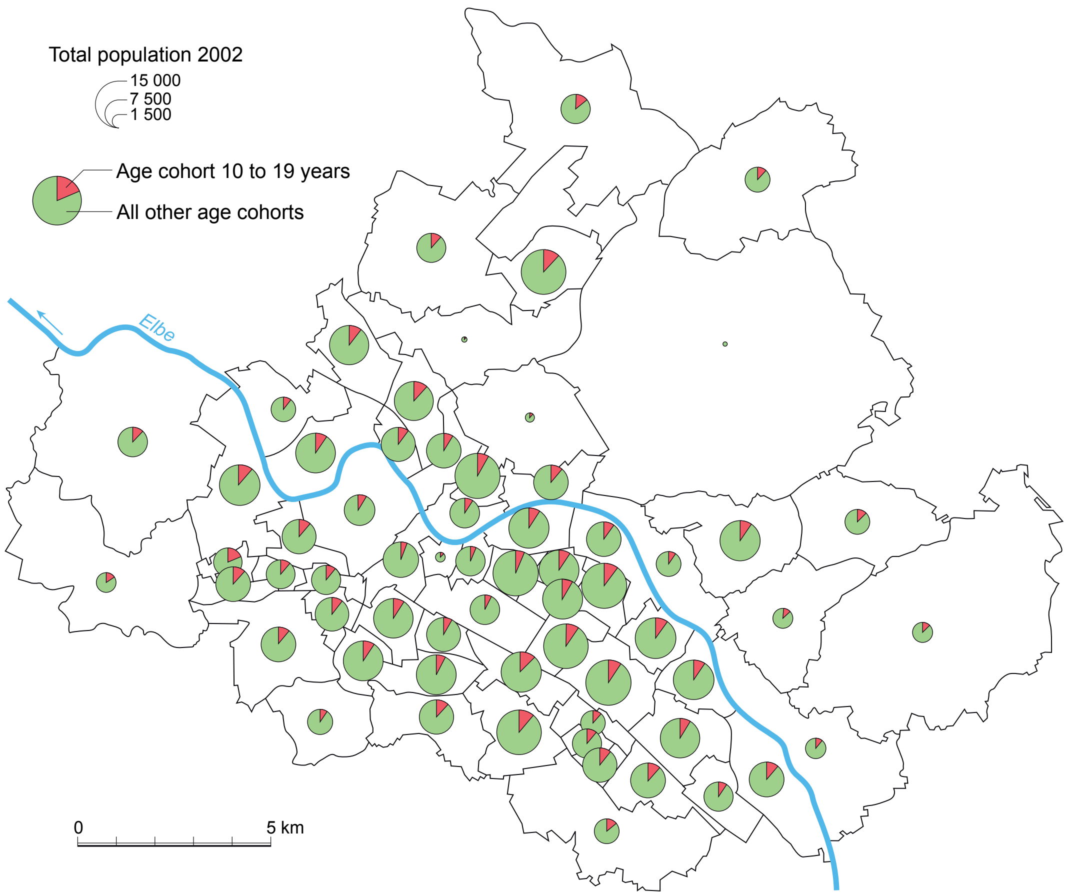


Fig. 4: Population of the city of Dresden on city district level. Numbers are given for the year 2002

to all other schools (Müller et al. 2011). We define the measure as

$$A_i = t_{ij} | (t_{ij} = \min\{t_{i1},...,t_{iQ}\}) \quad (6)$$

t_{ij} denotes the travel-time on the shortest-path between a given census block $i \in S$ and a school $j \in Q$. A_i is the travel-time of a given census block to the most proximate school. Hence, the lower the value of A_i, the better the accessibility for the given census block. Here, the absolute measure is of less relevance. We are more interested in the change of A_i over the time-period of school consolidation (2002–2008). 25% of the Gymnasium-schools were closed in this period. Figure 5 depicts the deterioration of this measure in proximate areas where schools were closed. However, the increase in minimum travel-time in the north of our study area was not anticipated. This phenomenon is related to the closure of very central schools, which in turn are located close to the central railway station. The considered census blocks in the north are closely located to a railway station. In terms of travel-time, the closest schools for the census blocks located in the north are schools located in the city center. Without this measure, the interdependency between school closure and decline in accessibility of census blocks located far away from the closed school site would not have been detected. The closure of schools located at the outskirts reveals an increase in minimum travel-time from 10 minutes to 30 minutes to some extent. However, most of the census blocks do not show an increase in minimum travel-time.

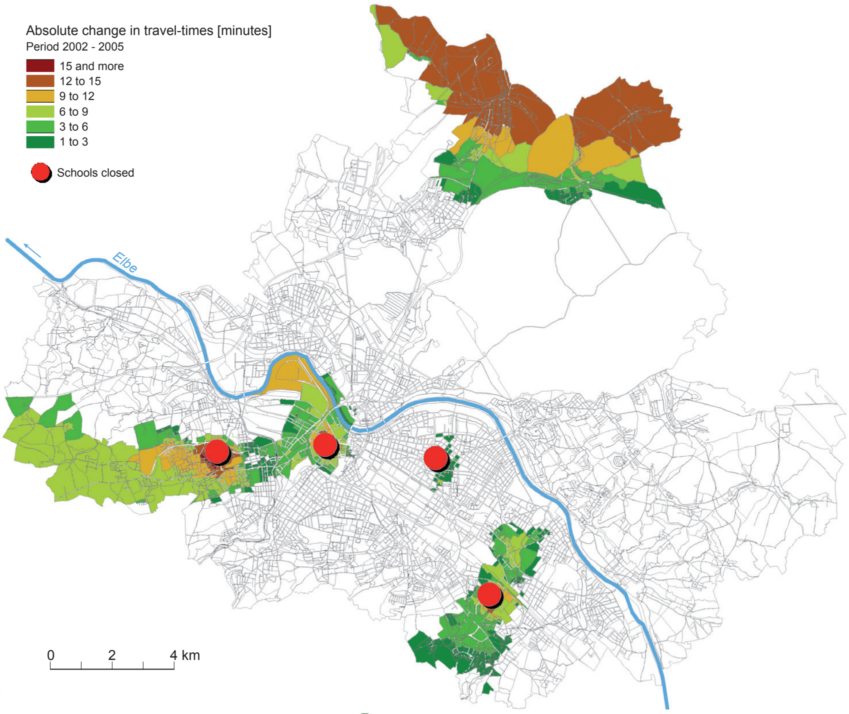
So far, we do not know how many students are faced with an increase in travel-time. Therefore we introduced a second relative measure of accessibility as

$$\tilde{A}_i = \frac{L_i^{\varrho+1}}{\max\{L_j^{\varrho+1} | j=1,...,S\}} A_i^{\varrho+1} - \frac{L_i^{\varrho}}{\max\{L_j^{\varrho} | j=1,...,S\}} A_i^{\varrho} \quad (7)$$

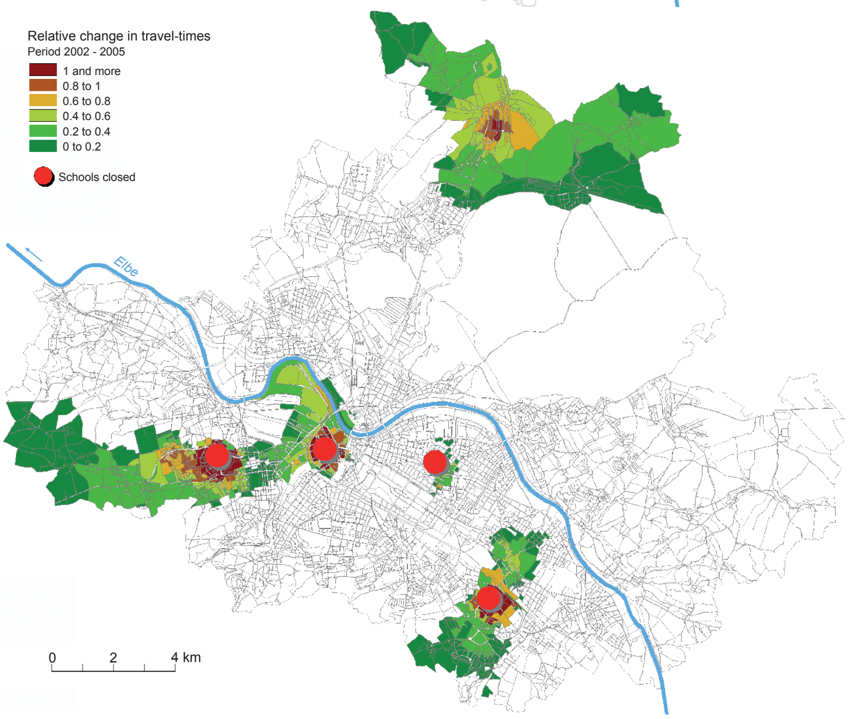
ϱ stands for the time period, A_i^{ϱ} the accessibility measure of (6) in period ϱ and L_i^{ϱ} is the absolute number of students in census block $i \in S$ who qualify for a Gymnasium-school in period ϱ. $i,j \in S$. This measure translates the change in absolute minimum travel-time between two periods into a relative measure weighted by the number of students of a given census block i relative to the maximum number of students over all census blocks. Hence, $\tilde{A}_i$ takes large positive values if many students are faced with a remarkable increase in travel-time. A high positive number has a negative meaning since travel-time increases from ϱ to $\varrho+1$ for a relatively large number of students. Therefore, $\tilde{A}_i$ is of particular interest for applications because we are able to evaluate school consolidation very quickly with one measure (change in relative student-minutes). As expected, the census blocks located proximate to closed schools show the highest values of this measure (see Fig. 6). Moreover, this measure gives us some more information about the impact of school consolidation. If we compare the change in A_i (see Fig. 5) and the outcome of $\tilde{A}_i$ in figure 6, we observe an interesting pattern: There are areas of census blocks - particularly in the center, the north, and the south-east – where only small increases (absolute and relative) in travel-time occur. However, we see that these areas exhibit large positive values of $\tilde{A}_i$. In contrast, some of the regions that show a large increase in travel-time (particularly the most northern areas) do not exhibit large positive values of $\tilde{A}_i$ as expected. Thus, if we only focus on the simple travel-time measure, we miss the effect of the travel-time increase on the respective students. $\tilde{A}_i$ and A_i tell us that students who were particularly located at the outskirts were most affected by an increase in travel-time due to school consolidation in the period 2002–2008.

5.2 Cumulative-opportunity measures of accessibility

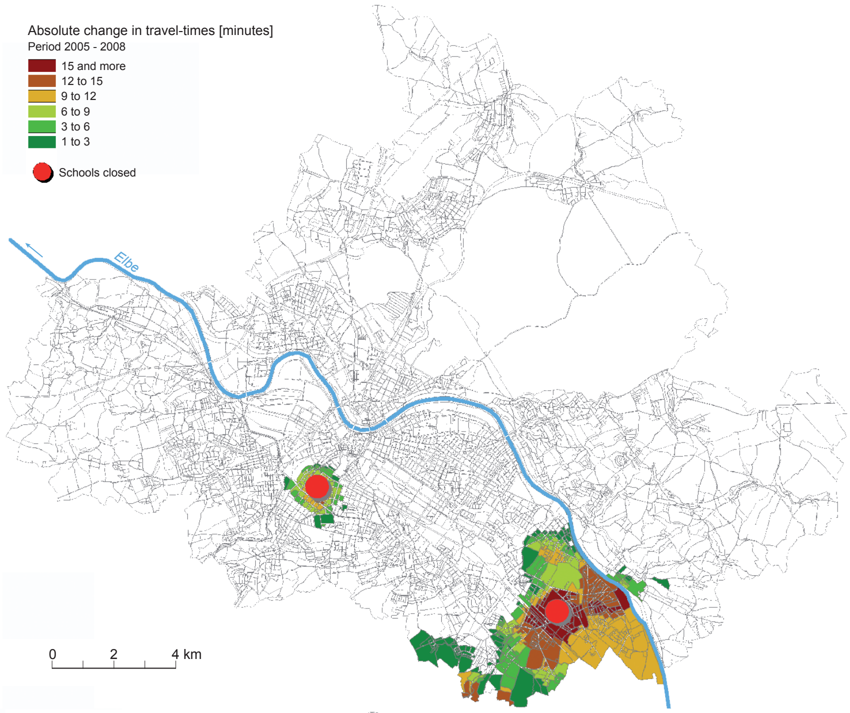
Here we consider two measures. The first measure is related to the accessibility of public transport infrastructure to schools. Therefore, we assume the higher the number of stops within an 800 meter buffer of a given school, the higher the accessibility of this school. The same should be true if we replace stops by lines. Now we have to define an evaluation scale. The range of this scale over all 24 Gymnasium-schools that opened in the year 2002 is 0–18 stops and 0–16 lines. Further, we partition these ranges in three equally large sub-ranges, i.e., 0–6 stops, 7–12 stops, and 13–18 stops; 0–5 lines, 6–10 lines and 11–16 lines. As figure 7 depicts, there are 6 schools with 7–12 stops and 18 schools with 13–18 stops. If we consider the number of lines, we find 7 schools with less than 6 lines, 17 schools with 6–10 lines, and only 2 schools with more than 9 lines. All together as expected, we see that the most central school locations have the best accessibility. However, in the outskirts, we expected schools to be less accessible. Apparently, this is not the case in the far south-east. The spatial structure of the public transport infrastructure results in a good accessibility for most of the schools (particularly in the south-east).


Absolute change in travel-times [minutes]
Period 2002 - 2005
15 and more
12 to 15
9 to 12
6 to 9
3 to 6
1 to 3
Schools closed
Elbe
0 2 4 km

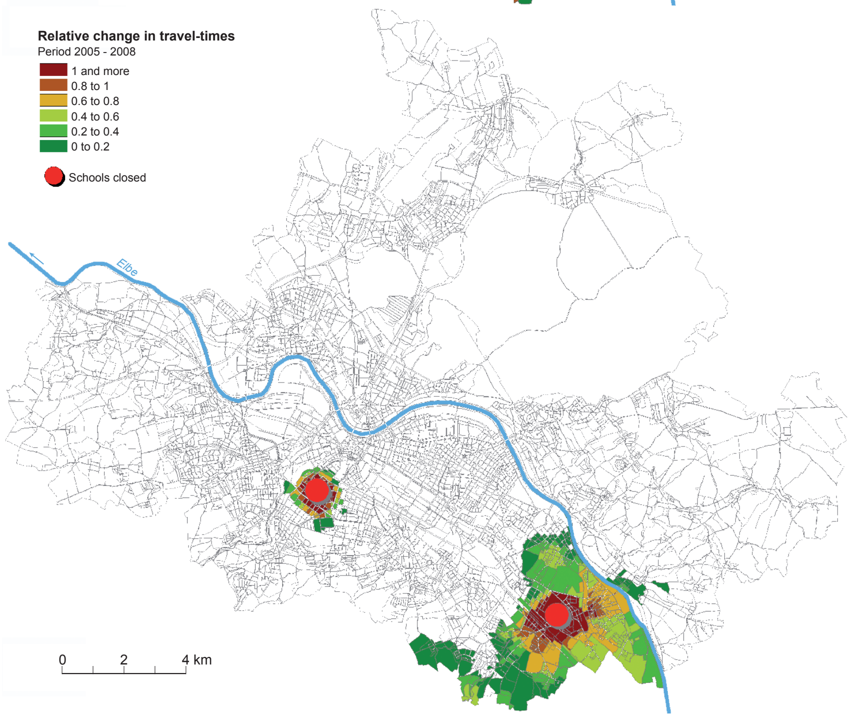
(a)


Relative change in travel-times
Period 2002 - 2005
1 and more
0.8 to 1
0.6 to 0.8
0.4 to 0.6
0.2 to 0.4
0 to 0.2
Schools closed
Elbe
0 2 4 km

(b)



(c)



(d)

Fig. 5: Accessibility measure A_i: Absolute (a, c) and relative (b, d) change in public transport travel-times on commute to school for the periods 2002–2005 (a, b) and 2005–2008 (c, d). The numbers are the absolute and relative increase in travel-time to the closest school. The maps are displayed on census block level

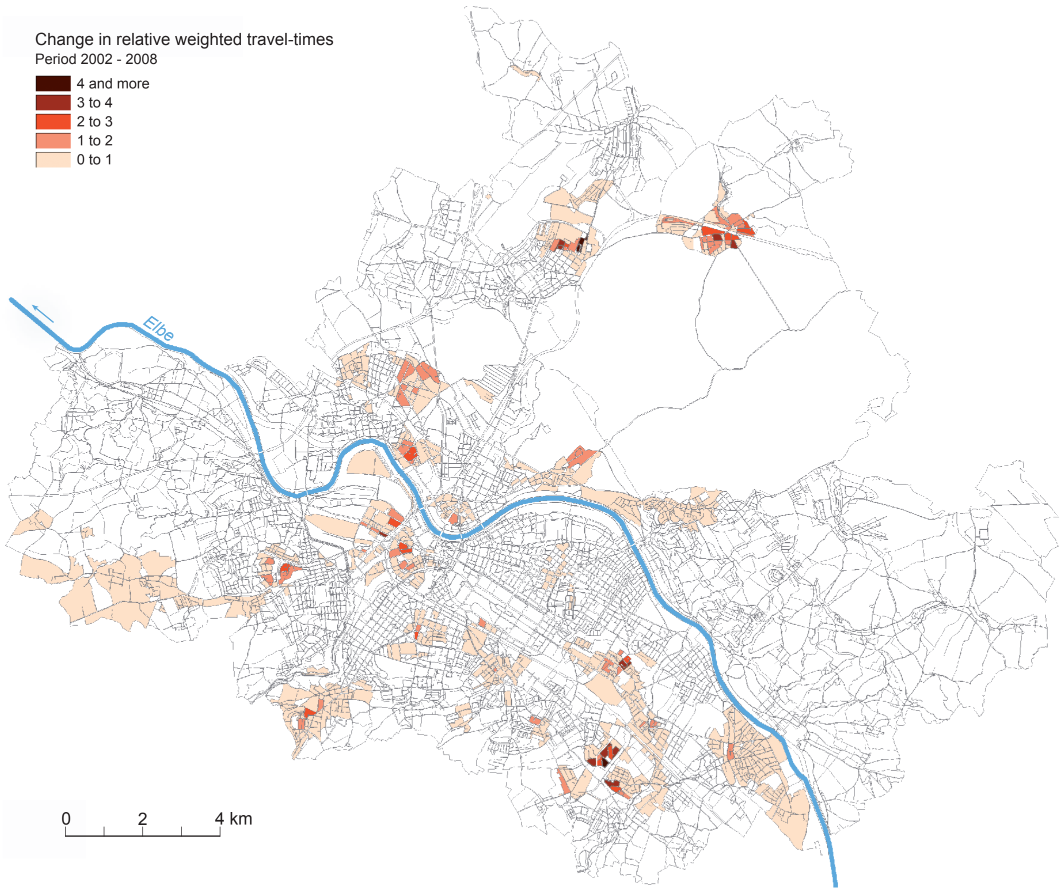


Fig. 6: Relative accessibility measure $\tilde{A}_i$: Change in travel-times to the closest school for the period 2002–2008. Travel-times are weighted by relative student numbers. The map is displayed at the census block level

Finally, we discuss a measure that employs an indicator function to exclude schools beyond a given travel-time T. Moreover, we consider the number of students per census block $i \in S$ if the travel-time from this census block to all schools open is less than T. The accessibility is measured as.

$$\Omega_i = L_i \,|\, (t_{ij} \leqslant T \quad \forall\, j \in Q) \tag{8}$$

In our study we set T = 45 minutes. The city council of the city of Dresden constitutes a maximum reasonable travel-time to Gymnasium-schools of 45 minutes (STADTRAT DER LANDESHAUPTSTADT DRESDEN 1997). The larger the student numbers of a census block $i \in S$ given that all schools $i \in Q$ are accessible within 45 minutes of public transport travel-time the larger is our measure Ω_i. This measure is an interesting supplement to the measures of section 5.1. Ω_i tells us where we find a given number of students who are privileged in terms of their school choice set. That is, these students may choose a school from the full choice set of schools. Figure 8 shows the values of Ω_i for all blocks in the years 2002 (a) and 2008 (b). As a general pattern we see that accessibility in terms of availability of choice alternatives is spatially discriminatory. That is, the outskirts do not have students who are able to access all schools within 45 minutes travel-time. Moreover, if we consider the absolute change between year 2002 and year 2008 this discrimination becomes even more obvious (Fig. 8 (c)).

6 Conclusion

Demographic processes have yielded a dramatic decline in student numbers in most regions of eastern Germany during the period of 1992–

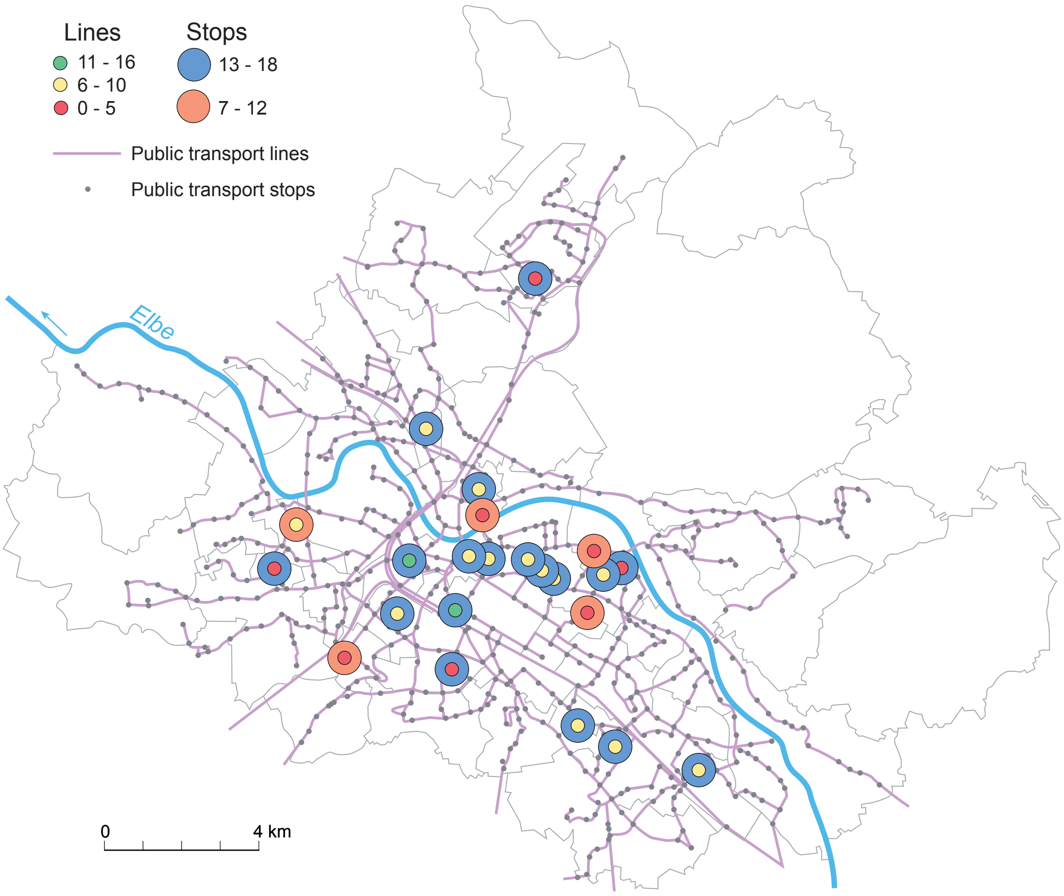
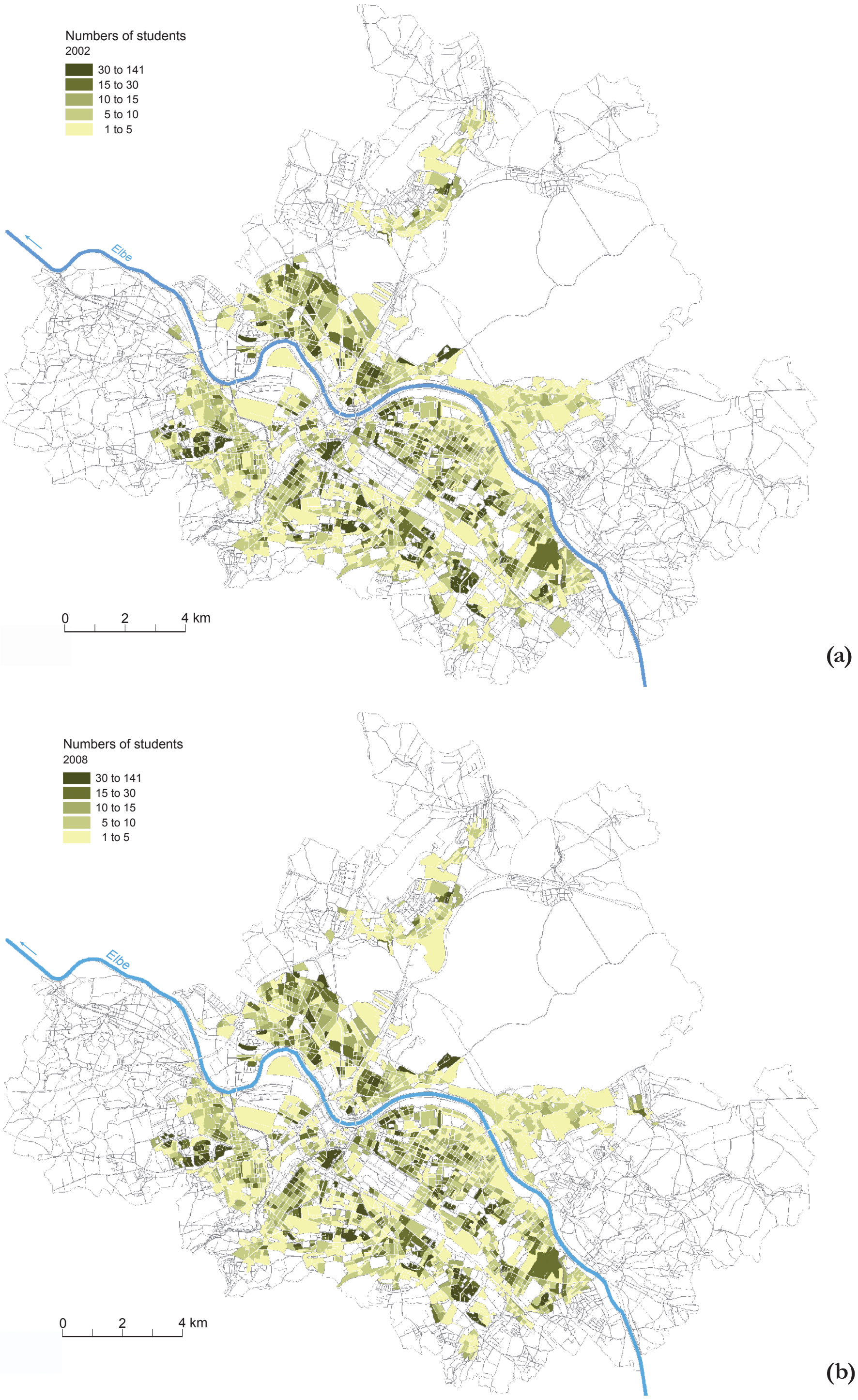


Fig. 7: Accessibility of Gymnasium-schools dependent on the access to public transport infrastructure in the period of 2002–2008. The number of stops within 800 meters arc radius is given. The number of lines given in the map depends on the number of lines serving the stops assigned to each school. The map is displayed on the city district level

2002. This in turn has led to a remarkable school consolidation of Gymnasium-schools in the time period 2002–2008. The question as to which school should be closed at a given point in time is a very difficult one. However, the political discussion about school consolidation in Germany lacks a complex and differentiated view. Müller (2010) pointed out that this might be due to models and measures that are too complex. Therefore, we discuss rather simple measures of school accessibility based on public transport travel-times. We show how these travel-times can be computed efficiently. This is of particular interest for prospective applications in planning and for local authorities. We use a simple mathematical program and standard GIS procedures in order to measure whether access to schools is spatially equitable or discriminatory. Due to our analysis, we were able to identify areas of service deprivation that need special attention. The measures presented here are straightforward and thus might be implemented in the planning process more easily than complex models and procedures.

The case study of school consolidation in the city of Dresden, Germany shows that areas that are located proximate to closed school sites are mostly affected by an increase in travel-time. Moreover, taking into account the number of students who are affected by an increase in travel-time to the closest school, we see that particularly outskirt areas are discriminated.

The demographic processes in eastern Germany are to be expected to take place in western Germany as well. Hence, it would be interesting to see, whether these (or other measures) will be implemented by educational authorities in the planning process.


Numbers of students
2002
30 to 141
15 to 30
10 to 15
5 to 10
1 to 5
Elbe
0
2
4 km
(a)
Numbers of students
2008
30 to 141
15 to 30
10 to 15
5 to 10
1 to 5
Elbe
0
2
4 km
(b)

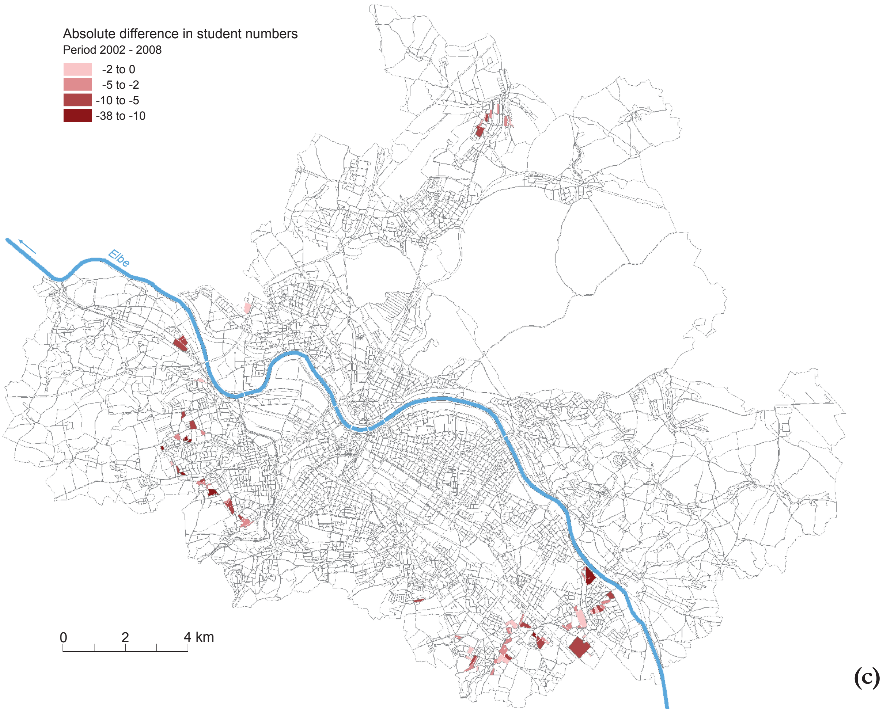


Fig. 8: Accessibility measure Ω_i: Number of students if travel-time to all schools open is less than 45 minutes. The absolute difference of Ω_i (c) between the year 2002 (a) and the year 2008 (b) is computed as min $\{0, \Omega_i^{2008} - \Omega_i^{2002}\}$. All maps are displayed on census block level

Acknowledgements

The author is grateful to two anonymous referees for their comments on an earlier version of this paper. Of course, the author retains responsibility for all remaining errors, omissions and opinions.

References

Ahuja, R.; Magnanti, T. and Orlin, J. (1993): Network flows. Theory, algorithms, and applications. Englewood Cliffs.

Brunisma, F. and Rietveld, P. (1998): The accessibility of European cities: theoretical framework and some comparison of approaches. In: Environment and Planning A 30, 499–521. DOI: 10.1068/a300499

Cormen, H.; Leiserson, C.; Rivest, R. and Stein, C. (eds.) (2001): Single-source shortest paths. Cambridge.

Domschke, W. and Drexl, A. (2005): Einführung in Operations Research. Heidelberg.

Grosse, C. (2003): Verkehrsabhängige Betriebsführung bei Stadtschnellbahnen. Ph.D. thesis, Technische Universität. Berlin.

Klameth, S. (2005): Drei Stunden Schulweg. In: Sächsische Zeitung (10. Juni).

Kwan, M.-P. (1998): Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework. In: Geographical Analysis 30, 191–215. DOI: 10.1111/j.1538-4632.1998.tb00396.x

– (1999): Gender and individual access to urban opportunities. A study using space-time measures. In: The Professional Geographer 51, 211–227. DOI: 10.1111/0033-0124.00158

Kwan, M.-P.; Murray, A. T.; O'Kelly, M. E. and Tiefelsdorf, M. (2003): Recent advances in accessibility research: representation, methodology and applications. In: Journal of Geographical Systems 5, 129–138. DOI: 10.1007/s101090300107

Leonardi, G. (1978): Optimum facility location by accessibility maximizing. In: Environment and Planning A 10, 1278–1305. DOI: 10.1068/a101287

Longley, P.; Goodchild, M.; Maguire, D. and Rhind, D. (2001): Geographic information systems and science. Chichester.

Müller, S. (2006): Analyse des Verkehrsmittel- und Schulwahlverhaltens Dresdner Gymnasiasten. In: Aurada, K. and Rödel, R. (eds.): Kooperation und Integration. In: Greifswalder Geographische Arbeiten 36, 145–155.

– (2008): Dynamic school network planning in urban areas. Urban and Regional Planning 5. Berlin.
– (2009): A spatial choice model based on random utility. In: Journal of the Institute of Transport and Economics 11 (2), 2–25.
– (2010): Bildungsstandorte und demographischer Wandel. Langfristig optimale Schulstandorte unter Berücksichtigung schwankender Schülerzahlen. In: Standort – Zeitschrift für angewandte Geographie 34 (1), 6–10.
Müller, S.; Haase, K. and Kless, S. (2009): A multi-period school location planning approach with free school choice. In: Environment and Planning A 41 (12), 2929–2945. DOI: 10.1068/a40285
Müller, S.; Tscharaktschiew, S. and Haase, K. (2008): Travel-to-school mode choice modelling and patterns of school choice in urban areas. In: Journal of Transport Geography 16 (5), 342–357. DOI: 10.1016/j.jtrangeo.2007.12.004
Müller, S.; Haase, K. and Seidel, F. (2011): Exposing unobserved spatial similarity: evidence from German school choice data. In: Geographical Analysis (Accepted for publication).
Peter, J. (2004): Das große Schulsterben. In: Die Welt 26. August, 3.
– (2005): Koalitionsstreit in Sachsen. In: Die Welt 20. Mai, 4.
Richter, K. (2005): Schwärzeste Prognosen übertroffen. In: Dresdner Neueste Nachrichten 31. Mai, 11.
Rodrigue, J.-P.; Comtois, C. and Slack, B. (2009): The Geography of transport systems. New York.
STADTRAT DER LANDESHAUPTSTADT DRESDEN (1997): Satzung der Landeshauptstadt Dresden über die Erstattung der notwendigen Schülerbeförderungskosten. Satzung Schülerbeförderungskosten-Erstattung. In: Dresdner Amtsblatt 37. Dresden.
Talen, E. (2001): School, community, and spatial equity: an empirical investigation of access to elementary schools in West Virgina. In: Annals of the Association of American Geographers 91 (3), 465–486. DOI: 10.1111/0004-5608.00254

Author

Dr. Sven Müller
University of Hamburg
Institute for Transport Economics
Von-Melle-Park 5
20146 Hamburg, Germany
sven.mueller@wiso.uni-hamburg.de

3 Impact of Articles

Chp.	Reference	5yIF	h5	HB	JQ	GS	Scop
2.1.1	Müller, S., Haase, K.: **Local revenue response to service quality: spatial effects in seasonal ticket revenue data**. *European Journal of Marketing*, accepted for publication.	1.575	35	0.4	6.07	-	-
2.2.1	Haase, K., Müller, S. (2014): **Insights in clients' choice in preventive health care facility location planning**. *OR Spectrum*, forthcoming.	1.984	28	0.5	8.10	-	-
2.2.1	Müller, S., Haase, K. (2014): **Customer segmentation in retail facility location planning**. *Business Research*, in print.	-	-	0.2	7.21	1	-
2.2.1	Haase, K., Müller, S. (2014): **A comparison of linear reformulations for multinomial logit choice probabilities in facility location models**. *European Journal of Operational Research* **232**(3), 689-691.	2.524	82	0.7	8.09	3	-
2.2.2	Haase, K., Müller, S. (2014): **Upper and lower bounds for the sales force deployment problem with explicit contiguity constraints**. *European Journal of Operational Research* **237**(2), 677-689.	2.524	82	0.7	8.09	-	-
2.2.1	Haase, K., Müller, S. (2013): **Management of school locations allowing for free school choice**. *Omega - The International Journal of Management Science* **41**(5), 847-855.	3.474	50	0.7	7.31	3	1
2.1.2	Müller, S., Rode, J. (2013): **The adoption of photovoltaic systems in Wiesbaden, Germany**. *Economics of Innovation and New Technology* **22**(5), 519-535.	-	22	0.3	6.58	4	-
2.1.1	Müller, S. Wilhelm, P., Haase, K. (2013): **Spatial dependencies and spatial drift in public transport seasonal ticket revenue data**. *Journal of Retailing and Consumer Services* **20**(3), 334-348.	-	29	0.3	6.47	1	-
2.1.2	Müller, S., Haase, K., Seidel, F. (2012): **Exposing unobserved spatial similarity: Evidence from German school choice data**. *Geographical Analysis* **44**(1), 65-86.	1.814	16	-	-	7	3
2.1.1	+Müller, S. (2012): **Identifying spatial nonstationarity in German regional firm start-up data**. *Review of Regional Research* **32**(2), 113-132.	-	-	-	-	1	-
2.2.1	+Müller, S. (2011): **Assessment of school closures in urban areas by simple accessibility measures**. *Erdkunde* **65**(4), 401-414.	1.073	11	-	-	2	-
2.1.2	*Müller, S., Tscharaktschiew, S., Haase, K. (2008): **Travel-to-school mode choice modelling and patterns of school choice in urban areas**. *Journal of Transport Geography* **16**(5), 342-357.	2.520	34	0.3	-	55	28

5yIF: The 5-year journal Impact Factor (Thomson Reuters, 2012). http://admin-apps.webofknowledge.com/JCR/help/h_impfact.htm
h5: Google h5-index 2013. http://scholar.google.de/intl/de/scholar/metrics.html#metrics
HB: Handelsblatt Ranking 2012. http://www.handelsblatt.com/politik/oekonomie/bwl-ranking/
JQ: VHB JOURQUAL 2.1. http://vhbonline.org/service/jourqual/vhb-jourqual-21-2011/
GS: Google Scholar citations. http://scholar.google.de/citations?user=CV1-hdt1y1UC&hl=de
Scop: Scopus citations. http://www.researcherid.com/rid/A-4104-2010
*: Publication immediately related to PhD thesis.; +: single author paper.

Table 3.1: Chronological ranking of contributions. Rankings of journals is given in the appendix

4 Declaration of Authenticity

I certify that:

(a) the thesis being submitted for examination is my own account of my own research - that is, to all papers included in this thesis I have contributed to at least 50% of the total workload

(b) my research has been conducted ethically

(c) the data and results presented are the genuine data and results actually obtained by me during the conduct of the research

(d) where I have drawn on the work, ideas and results of others this has been appropriately acknowledged

(e) where part of the work described in the thesis has previously been incorporated in another thesis submitted by me for a higher degree (PhD), this has been identified and acknowledged accordingly

(f) this thesis has not been submitted elsewhere

Sven Müller, Hamburg, 1. July 2014

Bibliography

Benoit, D. and G. Clarke (1997, October). Assessing GIS for retail location planning. *Journal of Retailing and Consumer Services 4*(4), 239–258.

Crossland, M., B. Wynne, and W. Perkins (1995, July). Spatial decision support systems: An overview of technology and a test of efficacy. *Decision Support Systems 14*(3), 219–235.

Day, D. L., W. S. DeSarbo, and T. A. Oliva (1987, December). Strategy maps: A spatial representation of intra-industry competitive strategy. *Management Science 33*(12), 1534–1551.

Graf, M. and S. M. Mudambi (2005, June). The outsourcing of it-enabled business processes: A conceptual model of the location decision. *Journal of International Management 11*(2), 253–268.

Grimshaw, D. (2000). *Bringing Geographical Information Systems into Business*. Jon Wiley and Sons.

Miller, F., W. G. Mangold, and T. Holmes (2006, November). Integrating geographic information systems (gis) applications into business courses using online business geographics modules. *Journal of Education for Business 82*(2), 74–79.

Porter, M. and S. Stern (2001). Innovation: Location matters. *MIT Sloan Management Review 42*(4), 28–36.

Vlachopoulou, M., G. Silleos, and V. Manthou (2001, May). Geographic information systems in warehouse site selection decisions. *International Journal of Production Economics 71*, 205–212.

5 Appendix

ISI Web of Knowledge℠

Journal Citation Reports®

WELCOME HELP MARKED LIST

2012 JCR Science Edition

Journal Summary List

Journal Title Changes

Journals from: **subject categories OPERATIONS RESEARCH & MANAGEMENT SCIENCE** VIEW CATEGORY SUMMARY LIST

Sorted by: 5-Year Impact Factor SORT AGAIN

Journals 1 - 20 (of 79) [1 | 2 | 3 | 4] **Page 1 of 4**

MARK ALL UPDATE MARKED LIST

Ranking is based on your journal and sort selections.

Mark	Rank	Abbreviated Journal Title *(linked to journal information)*	ISSN	JCR Data						*Eigenfactor*® Metrics	
				Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Articles	Cited Half-life	*Eigenfactor*® Score	*Article Influence*® Score
☑	1	J OPER MANAG	0272-6963	4825	4.400	7.130	0.351	37	8.0	0.00984	2.281
☑	2	TRANSPORT RES B-METH	0191-2615	4064	2.944	3.520	0.347	101	9.4	0.00953	1.487
☑	3	OMEGA-INT J MANAGE S	0305-0483	3316	3.024	3.474	1.135	89	7.3	0.00680	1.049
☑	4	TECHNOVATION	0166-4972	2536	3.177	3.449	0.596	52	6.5	0.00400	0.726
☑	5	MANAGE SCI	0025-1909	18351	1.859	3.057	0.374	139	>10.0	0.02881	2.360
☑	6	DECIS SUPPORT SYST	0167-9236	3895	2.201	3.037	0.196	209	5.9	0.00989	0.863
☑	7	TRANSPORT RES E-LOG	1366-5545	1751	2.272	2.764	0.419	86	5.1	0.00692	1.077
☑	8	TRANSPORT SCI	0041-1655	2642	1.814	2.623	0.182	33	>10.0	0.00476	1.559
☑	9	INT J PROD ECON	0925-5273	8186	2.081	2.594	0.585	347	6.2	0.01745	0.698
☑	10	M&SOM-MANUF SERV OP	1523-4614	1064	1.712	2.561	0.244	45	6.5	0.00666	2.019
☑	11	EUR J OPER RES	0377-2217	23560	2.038	2.524	0.311	589	8.1	0.05778	0.972
☑	12	OPER RES	0030-364X	8311	1.786	2.484	0.208	106	>10.0	0.01870	1.961
☑	13	RELIAB ENG SYST SAFE	0951-8320	4783	1.901	2.441	0.411	151	6.6	0.01267	0.875
☑	14	COMPUT OPER RES	0305-0548	5810	1.909	2.374	0.350	314	6.2	0.02030	1.022
☑	15	MATH PROGRAM	0025-5610	5751	2.090	2.351	0.255	106	>10.0	0.01396	2.065
☑	16	EXPERT SYST APPL	0957-4174	13933	1.854	2.339	0.401	1363	3.3	0.04216	0.498
☑	17	PROD OPER MANAG	1059-1478	1794	1.315	2.316	0.129	70	7.6	0.00542	1.212
☑	18	SYST CONTROL LETT	0167-6911	4239	1.667	2.054	0.218	170	9.7	0.01083	1.048
☑	19	OR SPECTRUM	0171-6468	850	1.410	1.984	0.250	40	6.7	0.00308	0.962
☑	20	INFORMS J COMPUT	1091-9856	1060	1.370	1.819	0.152	46	7.5	0.00540	1.246

MARK ALL UPDATE MARKED LIST

Journals 1 - 20 (of 79) [1 | 2 | 3 | 4] **Page 1 of 4**

Figure 5.1: JCR Impact Factor 2012 *Operations Research*

ISI Web of Knowledge℠

Journal Citation Reports®

WELCOME ? HELP

2012 JCR Social Science Edition

Journal Summary List

Journal Title Changes

Journals from: subject categories **TRANSPORTATION** VIEW CATEGORY SUMMARY LIST

Sorted by: 5-Year Impact Factor SORT AGAIN

Journals 1 - 20 (of 26) [1 | 2] **Page 1 of 2**

MARK ALL UPDATE MARKED LIST

Ranking is based on your journal and sort selections.

Mark	Rank	Abbreviated Journal Title (linked to journal information)	ISSN	JCR Data						*Eigenfactor®* Metrics	
				Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Articles	Cited Half-life	*Eigenfactor®* Score	*Article Influence®* Score
☐	1	TRANSPORT RES B-METH	0191-2615	4064	2.944	3.520	0.347	101	9.4	0.00953	1.487
☐	2	TRANSPORT RES A-POL	0965-8564	3062	2.725	3.000	0.346	127	8.0	0.00699	1.084
☐	3	TRANSPORT RES E-LOG	1366-5545	1751	2.272	2.764	0.419	86	5.1	0.00692	1.077
☐	4	TRANSPORT SCI	0041-1655	2642	1.814	2.623	0.182	33	>10.0	0.00476	1.559
☐	5	J TRANSP GEOGR	0966-6923	1575	1.942	2.520	0.568	118	4.9	0.00448	0.742
☐	6	ACCIDENT ANAL PREV	0001-4575	6386	1.964	2.391	0.367	286	7.0	0.01490	0.778
☐	7	TRANSPORT RES F-TRAF	1369-8478	1181	1.577	2.354	0.149	67	6.6	0.00325	0.835
☐	8	TRANSPORT REV	0144-1647	918	1.887	2.261	0.314	35	6.7	0.00281	0.915
☐	9	TRANSPORT POLICY	0967-070X	1366	1.541	2.161	0.232	112	6.5	0.00318	0.681
☐	10	TRANSPORTATION	0049-4488	1358	1.657	2.131	0.344	64	8.6	0.00328	0.850
☐	11	J SAFETY RES	0022-4375	1665	1.379	1.970	0.192	52	7.3	0.00459	0.781
☐	12	TRANSPORT RES D-TR E	1361-9209	1303	1.291	1.801	0.190	84	6.6	0.00370	0.657
☐	13	TRANSPORTMETRICA	1812-8602	182	1.357	1.575	0.591	22	3.8	0.00088	0.631
☐	14	INT J SHIP TRANS LOG	1756-6517	107	1.319	1.493	0.174	23	2.5	0.00034	0.279
☐	15	J TRANSP ECON POLICY	0022-5258	756	0.860	1.462	0.192	26	>10.0	0.00133	0.719
☐	16	INT J SUSTAIN TRANSP	1556-8318	130	0.944	1.407	0.381	21	3.5	0.00082	0.546
☐	17	J AIR TRANSP MANAG	0969-6997	829	0.800	1.204	0.077	104	7.0	0.00240	0.483
☐	18	TRANSPORT J	0041-1612	331	0.250	0.961	0.045	22	9.6	0.00036	0.265
☐	19	TRANSP LETT	1942-7867	43	0.500	0.717	0.000	19		0.00034	0.326
☐	20	INT J TRANSP ECON	0391-8440	143	0.393	0.333	0.000	15	>10.0	0.00019	0.129

MARK ALL UPDATE MARKED LIST

Journals 1 - 20 (of 26) [1 | 2] **Page 1 of 2**

Figure 5.2: JCR Impact Factor 2012 *Transportation*

Operations Research - Google Scholar-Kennzahlen http://scholar.google.com/citations?view_op=top_venues&hl=de&vq=...

Web **Bilder** **Mehr...**

svennemulatti@gmail.com

In Google Scholar suchen

Englisch
- Betriebs- und Volkswirtschaft, Management
- Bio- und Geowissenschaften
- Chemie und Werkstoffwissenschaften
- Geisteswissenschaften, Literatur und Kunst
- Gesundheit und Medizin
- Ingenieurwissenschaft und Informatik
  - Operations Research
- Physik und Mathematik
- Sozialwissenschaften

Chinesisch

Portugiesisch

Deutsch

Spanisch

Französisch

Italienisch

Japanisch

Niederländisch

Top-Publikationen - Operations Research Weitere Informationen

	Publikation	h5-Index	h5-Median
1.	European Journal of Operational Research	82	113
2.	International Journal of Production Economics	63	84
3.	Computers & Operations Research	59	71
4.	Journal of Operations Management	55	87
5.	Computers & Industrial Engineering	49	63
6.	Operations Research	49	63
7.	Mathematical Programming	45	59
8.	Transportation Research Part E: Logistics and Transportation Review	38	55
9.	International Journal of Production Research	37	47
10.	Manufacturing & Service Operations Management	34	60
11.	Journal of the Operational Research Society	34	44
12.	Supply Chain Management: An International Journal	33	42
13.	International Journal of Operations & Production Management	32	50
14.	International Journal of Physical Distribution & Logistics Management	30	47
15.	Computational Optimization and Applications	30	40
16.	Transportation Science	30	39
17.	Production and Operations Management	29	46
18.	Operations Research Spectrum	28	40
19.	Annals of Operations Research	28	37
20.	Mathematics of Operations Research	28	34

Datumsangaben und Zitierhäufigkeiten werden automatisch von einem Computerprogramm ermittelt und stellen Schätzwerte dar.

Über Google Scholar Über Google Datenschutzerklärung & Nutzungsbedingungen Feedback geben

Figure 5.3: Google Scholar 2013 *Operations Research*

Transportation - Google Scholar-Kennzahlen http://scholar.google.com/citations?view_op=top_venues&hl=de&vq=...

Web Bilder Mehr... svennemulatti@gmail.com

In Google Scholar suchen

Englisch
- Betriebs- und Volkswirtschaft, Management
- Bio- und Geowissenschaften
- Chemie und Werkstoffwissenschaften
- Geisteswissenschaften, Literatur und Kunst
- Gesundheit und Medizin
- Ingenieurwissenschaft und Informatik
  - Transportation
- Physik und Mathematik
- Sozialwissenschaften

Chinesisch

Portugiesisch

Deutsch

Spanisch

Französisch

Italienisch

Japanisch

Niederländisch

Top-Publikationen - Transportation   Weitere Informationen

	Publikation	h5-Index	h5-Median
1.	Accident Analysis & Prevention	44	62
2.	Transportation Research Part B: Methodological	40	55
3.	Transportation Research Part E: Logistics and Transportation Review	38	55
4.	IEEE Transactions on Intelligent Transportation Systems	37	51
5.	Transportation Research Part A: Policy and Practice	36	48
6.	Transportation Research Part C: Emerging Technologies	35	44
7.	Journal of Transport Geography	34	43
8.	Meeting Transportation Research Board	32	41
9.	Transportation Research Record: Journal of the Transportation Research Board	32	37
10.	Transportation Science	30	39
11.	Transport Policy	29	37
12.	Journal of Safety Research	27	37
13.	Computer-Aided Civil and Infrastructure Engineering	27	36
14.	Transportation Research Part F: Traffic Psychology and Behaviour	27	36
15.	Transportation	26	36
16.	Transportation Research Part D: Transport and Environment	26	34
17.	Institute of Transportation Studies, Working Paper Series	25	42
18.	Tunnelling and Underground Space Technology	25	31
19.	Transport Reviews	24	37
20.	IEEE Intelligent Vehicles Symposium	24	31

Datumsangaben und Zitierhäufigkeiten werden automatisch von einem Computerprogramm ermittelt und stellen Schätzwerte dar.

Über Google Scholar   Über Google   Datenschutzerklärung & Nutzungsbedingungen   Feedback geben

Figure 5.4: Google Scholar 2013 *Transportation*

Strategic Management - Google Scholar-Kennzahlen http://scholar.google.com/citations?view_op=top_venues&hl=de&vq=...

Web Bilder Mehr... svennemulatti@gmail.com

In Google Scholar suchen

Englisch
- Betriebs- und Volkswirtschaft, Management
  - Strategic Management
- Bio- und Geowissenschaften
- Chemie und Werkstoffwissenschaften
- Geisteswissenschaften, Literatur und Kunst
- Gesundheit und Medizin
- Ingenieurwissenschaft und Informatik
- Physik und Mathematik
- Sozialwissenschaften

Chinesisch
Portugiesisch
Deutsch
Spanisch
Französisch
Italienisch
Japanisch
Niederländisch

Top-Publikationen - Strategic Management Weitere Informationen

	Publikation	h5-Index	h5-Median
1.	Strategic Management Journal	70	96
2.	Academy of Management Journal	69	109
3.	Management Science	65	98
4.	Journal of Marketing	63	101
5.	Organization Science	63	97
6.	Academy of Management Review	61	99
7.	Journal of International Business Studies	59	82
8.	Harvard Business Review	58	97
9.	Journal of Management	57	103
10.	Journal of Business Research	57	81
11.	Journal of Management Studies	56	98
12.	Journal of Operations Management	55	87
13.	Entrepreneurship Theory and Practice	52	80
14.	Journal of Business Venturing	51	79
15.	Omega	50	82
16.	Industrial Marketing Management	43	64
17.	Technological Forecasting and Social Change	41	54
18.	Journal of the Academy of Marketing Sciences	40	73
19.	Marketing Science	40	59
20.	Journal of Corporate Finance	39	62

Datumsangaben und Zitierhäufigkeiten werden automatisch von einem Computerprogramm ermittelt und stellen Schätzwerte dar.

Über Google Scholar Über Google Datenschutzerklärung & Nutzungsbedingungen Feedback geben

Figure 5.5: Google Scholar 2013 *Management*

Marketing - Google Scholar-Kennzahlen http://scholar.google.com/citations?view_op=top_venues&hl=de&vq=...

Web **Bilder** **Mehr...** **svennemulatti@gmail.com**

In Google Scholar suchen

Englisch
- Betriebs- und Volkswirtschaft, Management
  - Marketing
- Bio- und Geowissenschaften
- Chemie und Werkstoffwissenschaften
- Geisteswissenschaften, Literatur und Kunst
- Gesundheit und Medizin
- Ingenieurwissenschaft und Informatik
- Physik und Mathematik
- Sozialwissenschaften

Chinesisch
Portugiesisch
Deutsch
Spanisch
Französisch
Italienisch
Japanisch
Niederländisch

Top-Publikationen - Marketing Weitere Informationen

	Publikation	h5-Index	h5-Median
1.	Journal of Marketing	63	101
2.	Journal of Consumer Research	48	66
3.	Journal of Marketing Research	47	69
4.	Industrial Marketing Management	43	64
5.	Journal of the Academy of Marketing Sciences	40	73
6.	Marketing Science	40	59
7.	Journal of Product Innovation Management	39	60
8.	European Journal of Marketing	35	51
9.	Journal of Retailing	35	49
10.	Journal of Service Research	33	50
11.	Psychology & Marketing	31	41
12.	Journal of Retailing and Consumer Services	29	41
13.	Journal of Interactive Marketing	28	41
14.	The Services Industries Journal	28	38
15.	Journal of Supply Chain Management	27	59
16.	Journal of International Marketing	27	40
17.	International Marketing Review	26	35
18.	Journal of Services Marketing	26	34
19.	International Journal of Research in Marketing	25	37
20.	Journal of Consumer Psychology	24	43

Datumsangaben und Zitierhäufigkeiten werden automatisch von einem Computerprogramm ermittelt und stellen Schätzwerte dar.

Über Google Scholar Über Google Datenschutzerklärung & Nutzungsbedingungen Feedback geben

Figure 5.6: Google Scholar 2013 *Marketing*

PUBS/JAMS/ranking article

Share |

A Ranking of Marketing Journals

G. Tomas M. Hult, Michigan State University
William T. Neese, University of North Alabama
R. Edward Bashaw, University of Arkansas at Little Rock

MICHIGAN STATE UNIVERSITY

The following ranking is a component of a comprehensive three-sample study published in the Journal of Marketing Education (Spring 1997). The objective of the study was to rank marketing journals based on their importance in disseminating scholarly marketing knowledge. Two indices were used to rank the journals: (1) importance/prestige index and (2) popularity/familiarity index. To be included on the ranking list, a journal had to be ranked by at least 5% of the number of respondents that ranked the top journal in the field. A stratified sample of 1,000 marketing academicians at the assistant, associate, and full professor levels was used to compile the rankings. In addition, two samples of 500 academicians each were used to validate the initial results using different ranking methods. The complete study includes rankings pertaining to the overall sample and also segmented samples (i.e., doctoral, non-doctoral, AACSB accredited, and non-AACSB accredited institutions). Please look for the complete study in the Spring 1997 issue of the Journal of Marketing Education. The complete reference of the article is:

> Hult, G. Tomas M., William T. Neese, and R. Edward Bashaw (1997), "Faculty Perceptions of Marketing Journals," Journal of Marketing Education, 19 (1), 37-52.

Overall Ranking

1. Journal of Marketing
2. Journal of Marketing Research
3. Journal of Consumer Research
4. Journal of Retailing
5. Journal of the Academy of Marketing Science
6. Marketing Science
7. Harvard Business Review
8. Journal of Business Research
9. Journal of Advertising
10. Journal of Advertising Research
11. Management Science
12. Journal of Personal Selling and Sales Management
13. Advances in Consumer Research Proceedings
14. Journal of Public Policy and Marketing
15. Journal of Marketing Education
16. Psychology and Marketing
17. Sloan Management Review
18. Journal of Business
19. Journal of International Business Studies
20. Industrial Marketing Management
21. Journal of Consumer Marketing
22. California Management Review
23. Business Horizons
24. Journal of International Marketing
25. Journal of Services Marketing
26. International Journal of Research in Marketing
27. Journal of Consumer Psychology
28. Journal of Marketing Theory and Practice
29. AMA Educators' Conference Proceedings
30. European Journal of Marketing
31. Journal of Marketing Management (USA)
32. Journal of Health Care Marketing
33. Decision Sciences
34. Marketing Letters
35. Journal of Product Innovation Management
36. Journal of Business and Industrial Marketing
37. Journal of Global Marketing
38. Marketing Management
39. Journal of Business Ethics
40. Journal of Business Logistics
41. Marketing Education Review

Doctoral Institutions

1. Journal of Marketing
2. Journal of Marketing Research
3. Journal of Consumer Research
4. Marketing Science
5. Journal of Retailing
6. Journal of the Academy of Marketing Science
7. Harvard Business Review
8. Management Science
9. Journal of Business Research
10. Journal of Advertising
11. Journal of Advertising Research
12. Journal of Public Policy and Marketing
13. Sloan Management Review
14. Journal of Business
15. International Journal of Research in Marketing
16. Psychology and Marketing
17. Advances in Consumer Research Proceedings
18. Journal of Personal Selling and Sales Management
19. Industrial Marketing Management
20. Decision Sciences
21. Journal of Consumer Psychology
22. Journal of International Business Studies
23. Marketing Letters
24. California Management Review
25. Journal of Product innovation Management
26. Journal of International Marketing
27. European Journal of Marketing
28. Journal of Health Care Marketing
29. Journal of Consumer Marketing
30. Journal of Business Logistics
31. AMA Educators' Conference Proceedings
32. Journal of Services Marketing
33. Journal of Marketing Education
34. Business Horizons

Non-Doctoral Institutions

1. Journal of Marketing
2. Journal of Marketing Research
3. Journal of Consumer Research
4. Journal of Retailing
5. Journal of the Academy of Marketing Science
6. Harvard Business Review
7. Journal of Advertising
8. Journal of Business Research
9. Journal of Advertising Research
10. Marketing Science
11. Journal of Personal Selling and Sales Management
12. Journal of Marketing Education
13. Advances in Consumer Research Proceedings
14. Journal of Consumer Marketing
15. Business Horizons
16. Journal of Public Policy and Marketing
17. Journal of International Business Studies
18. Journal of Services Marketing
19. Management Science
20. Psychology and Marketing
21. Industrial Marketing Management
22. Journal of International Marketing
23. California Management Review
24. Sloan Management review
25. Journal of Marketing Theory and Practice
26. Journal of Business
27. AMA Educators' Conference Proceedings
28. Journal of Marketing Management (USA)
29. European Journal of Marketing
30. Journal of Business and Industrial Marketing
31. Journal of Health Care Marketing
32. Journal of Business Ethics
33. Journal of Global Marketing
34. Journal of Consumer Psychology

Figure 5.7: Ranking of Marketing Journals (Hult et al. (1997): Faculty perceptions of Marketing journals. Journal of Marketing Education, 19(1), 37-52.)

Table 2

Journals Ranked by Mean Quality and then Median Quality

Journal	mean quality	median quality	mode quality	standard deviation	Range	# rating quality	# rating quality as % of total responses	audience rating[1]	# giving this audience type as % of total # rating audience
Operations Research	1.02	1	1	0.16	1	82	96.5	ag	58.7
Management Science	1.09	1	1	0.5	4	85	100	ag	80.8
Mathematics of Operations Research	1.45	1	1	0.76	3	73	85.9	as	89.9
Mathematical Programming	1.68	1	1	1.01	4	63	74.1	as	96.7
Journal of the American Statististical Association	1.68	1	1	1.16	4	53	62.4	as	70.9
Manufacturing and Service Operations Management (INFORMS journal)	1.85	2	1	1.06	4	75	88.2	as	62.9
Naval Research Logistics	2.44	2	2	0.86	4	78	91.8	as/ag	50.7
SIAM Review	2.47	2	1	1.33	5	47	55.3	as	68.1
IIE Transactions	2.5	2	2	0.98	4	68	80	as/ag	52.4
Transportation Science	2.52	2	2	1.05	4	60	70.6	as	88.1
Interfaces	2.58	2	2	1.22	6	79	92.9	p	78.9
INFORMS Journal on Computing	2.71	2	2	1.35	6	48	56.5	as	89.8
Operations Research Letters	2.72	3	3	1.19	6	71	83.5	ag	55.4
Networks	2.83	3	2	1.26	5	41	48.2	as	92.5
Annals of Operations Research	2.89	3	3	1.13	5	64	75.3	as	65
European Journal of Operational Research	3	3	3	0.92	4	81	95.3	ag	61.8
Production and Operations Management	3.2	3	3	1.05	6	71	83.5	as	61.5
Journal of Operations Management	3.22	3	3	1.22	5	65	76.5	as	58.1
Journal of the Operational Research Society	3.25	3	3,4	1.18	5	64	75.3	ag/as	52.6
Decision Sciences	3.54	4	4	1.31	6	72	84.7	ag	66.7
Computers and Operations Research	4.07	4	4	1.14	5	45	52.9	as	75.6
Mathematical and Computer Modelling	4.09	4	4	1.38	6	34	40	as	87.2
International Journal of Production Research	4.18	4	4	1	4	57	67.1	as	63.6
International Journal of Production Economics	4.31	4	4	1.16	5	51	60	as	62.5
Decision Support Systems	4.37	4	4	1.13	5	43	50.6	as	88.1
Computers and Industrial Engineering	4.49	4	4	1.07	4	39	45.9	as	72.5
Omega	4.5	4	4	1.22	5	54	63.5	ag/as	52
American Journal of Mathematical and Management Sciences	4.75	5	4,5	1.5	5	32	37.6	as	71

[1]The audience of the journal is classified into ag, as or p if more than 50% of those rating the audience gave it that classification.

Figure 5.8: Top Journals in OM and OR. Mean/median quality (Olson, J. (2000): Top Journals in Operations Management and Operations Research. Working Paper.

Table 3

Journals Ranked by Number Who Rated Quality

Journal	mean quality	median quality	mode quality	standard deviation	Range	# rating quality	# rating quality as % of total responses	audience rating[1]	# giving this audience type as % of total # rating audience
Management Science	1.09	1	1	0.5	4	85	100	ag	80.8
Operations Research	1.02	1	1	0.16	1	82	96.5	ag	58.7
European Journal of Operational Research	3	3	3	0.92	4	81	95.3	ag	61.8
Interfaces	2.58	2	2	1.22	6	79	92.9	p	78.9
Naval Research Logistics	2.44	2	2	0.86	4	78	91.8	as/ag	50.7
Manufacturing and Service Operations Management (INFORMS journal)	1.85	2	1	1.06	4	75	88.2	as	62.9
Mathematics of Operations Research	1.45	1	1	0.76	3	73	85.9	as	89.9
Decision Sciences	3.54	4	4	1.31	6	72	84.7	ag	66.7
Operations Research Letters	2.72	3	3	1.19	6	71	83.5	ag	55.4
Production and Operations Management	3.2	3	3	1.05	6	71	83.5	as	61.5
IIE Transactions	2.5	2	2	0.98	4	68	80	as/ag	52.4
Journal of Operations Management	3.22	3	3	1.22	5	65	76.5	as	58.1
Annals of Operations Research	2.89	3	3	1.13	5	64	75.3	as	65
Journal of the Operational Research Society	3.25	3	3,4	1.18	5	64	75.3	ag/as	52.6
Mathematical Programming	1.68	1	1	1.01	4	63	74.1	as	96.7
Transportation Science	2.52	2	2	1.05	4	60	70.6	as	88.1
International Journal of Production Research	4.18	4	4	1	4	57	67.1	as	63.6
Omega	4.5	4	4	1.22	5	54	63.5	ag/as	52
Journal of the American Statististical Association	1.68	1	1	1.16	4	53	62.4	as	70.9
International Journal of Production Economics	4.31	4	4	1.16	5	51	60	as	62.5
INFORMS Journal on Computing	2.71	2	2	1.35	6	48	56.5	as	89.8
SIAM Review	2.47	2	1	1.33	5	47	55.3	as	68.1
Computers and Operations Research	4.07	4	4	1.14	5	45	52.9	as	75.6
Decision Support Systems	4.37	4	4	1.13	5	43	50.6	as	88.1
Networks	2.83	3	2	1.26	5	41	48.2	as	92.5
Computers and Industrial Engineering	4.49	4	4	1.07	4	39	45.9	as	72.5
Mathematical and Computer Modelling	4.09	4	4	1.38	6	34	40	as	87.2
American Journal of Mathematical and Management Sciences	4.75	5	4,5	1.5	5	32	37.6	as	71

[1]The audience of the journal is classified into ag, as or p if more than 50% of those rating the audience gave it that classification.

Figure 5.9: Top Journals in OM and OR. Number of quality ratings (Olson, J. (2000): Top Journals in Operations Management and Operations Research. Working Paper.