Comparing estimation results of land use development models using different databases available in Switzerland

Working Paper

Author(s):
Zöllig, Christof; Axhausen, Kay W.

Publication date:
2012

Permanent link:
https://doi.org/10.3929/ethz-a-007340033

Originally published in:
Arbeitsberichte Verkehrs- und Raumplanung 796

Comparing Estimation Results of Land Use Development Models Using Different Data Bases Available in Switzerland

Date of submission: 01/08/12

Authors:
Christof Zöllig Renner (Corresponding author)
IVT, ETH Zurich, CH-8093 Zurich
phone: +41-44-633 27 19
fax: +41-44-633 10 57
christof.zoellig@ivt.baug.ethz.ch

Kay W. Axhausen
IVT, ETH Zurich, CH-8093 Zurich
phone: +41-44-633 39 43
fax: +41-44-633 10 57
axhausen@ivt.baug.ethz.ch

Total document: 5937

Number of words (max 7500 words, incl. abstract and references):
Number of figures, tables, or photographs: 4
Word equivalent: 7945

Abstract (max 250 words)

Land development models are often the weak part of land use transport interaction models. This is partly due to the in transparent real estate market, e.g. decisions on the realisation of real estate development projects are seldom recorded. We understand the term *project* in this context as prepared plan for building new built space which materialises in form of buildings. Since we are ultimately interested in explaining development of built space with discrete choice modelling, we would ideally observe the decision of realisation, which happens probably just before applying for construction permit. However, it is more likely that register information on existing buildings are available rather than data on project decisions. If the register contains information on the construction year we can take this as an observation of a development event. One thing we miss with this conception is the fact that sometimes multiple buildings are constructed in the same project.

This paper discusses this issues in the context of an implementation of a parcel based UrbanSim land development model implemented for a land use transport model of Canton Zurich in Switzerland. We estimate four discrete choice models of developments choosing their location using UrbanSim. With in the whole modelling system these models have the purpose of updating the stock of real estate which is needed to provide alternatives for further models household or firm location choice models. Each development model is estimated on two differing data basis. One contains construction projects, also with multiple buildings, and the other contains only single buildings as observations. The estimation results are discussed focusing on the impact of disregarding project information.

We find that using building registers is a viable option if no data on development projects is at hand, especially if there are few multi building projects in the considered area. However, register information has the disadvantage of not containing proper observations which leads to certain bias. Variables such as the fit to development constraints are especially critical when formulating development location choice models on the level of parcels.

INTRODUCTION

The main idea of land use transport interaction models is the interdependence of both systems via accessibility (1). Most recent and recommendable reviews of land use transport interaction models are (2–5). For a discussion of the accessibility concept and its measurement see the papers by Karst Geurs and colleagues (6, 7).

The availability of integrated models is useful if the planners want to assess politics of land use and transport comprehensively considering the feedback loop between both systems. On the land use side this requires to get a better understanding of how regulations affect the development of built space and how consequently transport infrastructure is affected, and vis-versa.

In this paper we describe an implementation of a land use development model considering different information basis. This is a practical issue since data availability is not always given as desired. In the case described here we have two sources of information on land use development events at hand. The first source is the federal buildings and dwellings register and the second source contains development projects. The first source is more comprehensive and richer in detail in terms of building description and also allows to extract development events via the *year built* attribute. However, the second source is theoretically preferred since the development decisions in reality are not necessarily done per single building but rather per project, which can be composed of multiple buildings. Thus the term *project* is understood as prepared plan for building new built space which materialises in form of buildings in this work. We are fully aware that a construction project is a process with many decisions and complex dynamics, but in this context we simplify it to one main decision. The second data source allows us to derive this joint construction information and hence to estimate the proposed land use development model with or without the information on actual projects. In the following we refer to the first case as building location choice model (BLCM) and the latter we name project location choice model (PLCM). So, when estimating BLCM, on the basis of single buildings as observations, we implicitly assume that the development decisions are made for each building separately. In the case of PLCM we take projects as observations which can be composed of multiple buildings.

The research question is if it is adequate to use BLCM in land use transport interaction models. We expect biased estimation parameters when estimating BLCM since the observation in case of BLCM is not the real decision situation.

We investigate this question in the context of constructing an integrated land use model for Canton Zurich in Switzerland. We are using the open platform for urban simulation OPUS. The model is working at a high degree of spatial resolution as we are implementing the model for the parcel level. For the land use development model this high resolution is appealing since builders actually buy properties to realise their projects.

Researching land use development modelling can also be justify by refering to literature. Hunt points out that often the supply of built space is the weakest point in land use transport models (4). Also Haider points out that built space supply is rarely investigated (8). Literature in real eastate research indicates the same (9).

The paper is organised in a following section on land use modelling where theory is referenced and the methodology is laid out. In the third section we describe data preparation and the models. Estimation results are presented and interpreted in section four before we draw our conclusions.

LAND USE DEVELOPMENT MODELLING

The production of built space can be conceptualised as two basic markets, one market of land and one of built space (10). On the land market properties are traded as the basic resource for builders to realise a project. On the market for built space ready to use spaces are traded. Thus the builders provide new built space by modifying properties using various factor inputs. The general aim of a builder is to modify the property such that value is added i.e. the resulting product can be sold on the market with profit.

There are two basic behavioural approaches to land use modelling. Martinez (11) and followers use an aggregate (discriminating groups of builders) approach using bid-aution theory pioneered by Alonso (12). Waddell (13, 14) and colleagues focus on disaggregate, microsimulation approach using hedonic price modelling (15), which is nowadays a standard tool in real estate economics, first used in a spatial context by Rosen (16). The first approach is based on the paradigm of market equilibrium while the latter is more flexible in this respect. As we see the urban system characterised by complexity and not necessarily in equilibrium, we prefer to go with microsimulation at this point like other colleagues (17, 18).

Regardless of the approach taken the land use development models have to answer the main questions of when, where and how land use changes. The focus here is on the question *where* new building projects locate. We apply discrete choice modelling with random utility maximisation (RUM) pioneered by McFadden (19, 20) to the context of location choice of builders. Similar work as been done by Haider and Miller (8) and Dong and Gliebe (21).

Paul Waddell and colleagues (13) have been developing an open platform for micorsimulation for land use (OPUS / UrbanSim) since the nineties. In terms of usability it is probably the most advanced and comprehensive open source software for urban modelling at this time. UrbanSim is a system of loosely coupled submodels giving the user high flexibility in use. The core models of UrbanSim are location choice models for households and firms which locate in buildings. The land development model provides the buildings as alternatives by updating the building stock. In the newest parcel base version of UrbanSim this update decision is based on real estate prices given by a hedonic real estate price model and the resulting return on investment calculation of a representative developer (22) who chooses the project option with maximum expected return on investment.

However, it is also possible to formulate the development model as a location choice model of projects to be placed on parcels. The projects to be placed are then sampled from previously occurring projects in a number determined by the difference of target and actual vacancy rate. This approach is less attractive from a theoretical point of view but has its appeal for practical application since it is less data intensive.

A main motivation for land use transport models is the joint assessment of land use and transport infrastructure policies to organise the built environment in a better way. The land use development model thus has to consider development regulations. In UrbanSim such development constraints are applied to the geographies which are the alternatives for the land use development model. The constraints can then be used to restrict the choice set or by introducing the information directly into the utility function of the choice model.

Methodology

We use UrbanSim to estimate two sets of land development location choice models. The two sets are estimated on different data bases. The first set of models is estimated using data of development projects, while the second set is estimated on the basis of development events extracted from building register data. The extracted developments are not actually observed projects. In that case we rather observe the entry of a building to the register. We essentially miss the information on which buildings have been built as part of the same project. This is not ideal for behavioural modelling since the development decision is taken on the basis of projects rather than on the basis of individual buildings.

The comparison of the models cannot be done on the basis of goodness of fit measures like adjusted rho2 because different samples are used. Hence, we compare the models qualitatively and on the basis of parameter ratios. This requires using the same variables across the models.

In the first place we are estimating MNL models because of the choice set sampling problem (20). We then can apply random sampling of alternatives which is necessary in this case since considering full choice sets would generate very high computational costs. Another reason is that considering full choice sets is behaviourally unrealistic. No developer is assessing all parcels at the same time. Also, MNL models are usually the reference to compare more elaborated models with at a latter stage.

ESTIMATING LAND USE DEVELOPMENT LOCATION CHOICE MODELS ON TWO DIFFERENT SAMPLES

The study area comprises 151 municipalities of Canton Zurich. On the temporal axis we consider the time span from 2000 – 2010. A further description of the study area and the real estate market can be found in (23).

To be able to estimate location choice models of land development we first have to prepare the necessary data. This is a challenging effort since the original data is very heterogeneous and sometimes inconsistent. Here we focus on the preparation of the development events data and refer the reader interested in the preparation of the other data to Schirmer et al. (24).

Data used

The information on development events is taken from two main sources. The first source is the federal buildings and dwellings register. This data comprises information on the location of buildings by coordinates, size, type, age, and internal structure. In the context of this paper the attribute *year built* is crucial since we derive the point in time of construction from it. The second data source contains development projects from a database to inform potentially interested construction firms about planned projects. The project description comprises the number of buildings, the date of construction and the location by address.

Data preparation

The preparation of the development event data with *PostgreSQL* includes the matching of the project data to the building data from the register and dismissing data points with missing or implausible content. Namely in cases where development took place outside construction zones according to the data. Further we decided to use new construction only since we expect this to be the relevant events for the land use transport system. The matching of development projects to constructed buildings is achieved with spatio-temporal matching. Buildings are assigned to a project when they are located on the same parcel and have a *year built* after the project has been finished. Unfortunately, we loose already 74% of all recorded new construction projects when tying to match them to parcels due to poor address matching. In the following we loose another 75% when trying to match buildings to projects. When we further remove the projects located on parcels not inside the construction zone, we end up with a dataset of 1044 projects comprising 1147 buildings. The projects are located in 114 out of 151 municipalities all over Canton Zurich.

Note that the development projects have to be linked to buildings because at the end we want the development model to provide buildings. The buildings and dwellings register provides much more detail on the constructed buildings. The important information we use from the projects data is the information about joint construction. We used *PostgreSQL* for data integration.

Database for estimation

The final datasets for estimation are created in UrbanSim, when running a model estimation. The estimation dataset is different for each estimation since we apply random sampling of alternative. The variables used in the utility function are partly calculated outside UrbanSim using GIS and partly defined and calculated within UrbanSim using the domain specific modelling language (25).

Model formulation

We use simple MNL models for four types of development. The distinguished types are single family housing (SFH), multi family housing (MFH), residential with side use (RWSU) and non-residential developments. The types are based on intended main uses for the developments. Within the residential segment we could consider three sub segments. This was not possible for the non-residential segment due to the small number of observations.

There are four separate models to be able to consider the planning restrictions in the sampling of alternatives. With separate models it is possible to exclude certain alternatives from the sampling, e.g. a residential building is not allowed to be placed on a parcel zoned industrial. Formulating an overall model with dummies is therefore not an option. The utility functions are linear in parameters and consist of the variables listed in table 4. The variables are described in table 1 and the following subsection. Descriptive statistics are shown in table .

Variable Name	Description	Unit
acc_car	Accessibility of jobs and inhabitants by car according to formula (1). The calculations of travel times are based on the cantonal transport model.	[-]
acc_pt	Accessibility of jobs and inhabitants by public transport according to formula (1). The calculations of travel times are based on the cantonal transport model.	[-]
dist_to_school	Euclidean distance to next school facility.	[m]
dev_fit_permitted_floor_m2_step_ln	Step function differentiating between permitted and restricted volumes as follows: Ln of the difference between permitted floor-space of considered parcel and floor-space of development project, if the difference is larger than zero. Four times the difference between permitted floor-space of considered parcel and floor-space of development project, if the difference is smaller than zero.	[m2]
new_bldgs_within_150_of_parcel	Number of buildings with year built later than 1995 within 150 m.	[-]
newcomers_within_300_of_parcel	Number of residents which reported a different address five years ago within a radius of 300 m.	[-]
price_permitted_floor_m2	Price per permitted square meter floor-space.	[CHF/m2]
slope	Slope of parcel in percent.	[%]

Table 1 Variable Description

Variables

The reasoning of builders can be interpreted as anticipation of the consumers preferences. Therefore, it is not surprising that in case of residential buildings we think of similar variables to include as in household location choice models. This rises the question which variables are only of interest to the builder. To get an idea of what variables are important for the location of development events we studied the literature and conducted in-depth interviews with builders (26). The two sources show that characteristics of the parcel itself plus all externalities from the neighbourhood are considered. Examples of attributes mentioned are construction costs (slope as proxy), complicated legal situations, variables describing the land market (provision of same built space type, absorption rate), condition of soil (contamination demanding for sanitation before development), population growth or influx of young adults (newcomers). The consideration of variables is limited by data availability. We find it useful to classify the explanatory variables in four classes: Topography, infrastructure (networks and built space), socio-economic actors and norms (legal and social).

All variables which are not characterising the parcel directly are location externalities for which we also have to define how far their spatial reach is. This very similar problem is discussed by Guo and Bhat (27) where they try to pin down the neighbourhood concept or more generally to what spatial extent the endowment of the vicinity is perceived. For the operationalisation with circular units they come up with “neighbourhood radii” of 0.4 km, 1.6 km and 3.2 km. They also find that socio-economic variables “have significantly smaller spatial extent of influence than the land-use variables” (p. 44). A quick analysis of the open-field system (names of land/fields) from the cadastre data in the Canton Zurich shows a median size of 4.17 hectares which suggests a radius of 115 meters. This is small compared to the US numbers. A sample of ad-hoc measurements on a city map yield neighbourhood areas of 3.5 to 35 hectares suggesting radii of 105 to 334 meters. On this basis we chose radii of 150 to 300 meters for our variables. In our case this concerns the accessibility variables and the variables containing *within* in their names.

The accessibility variables are calculated with the following formula

$$acc = \ln\left(\sum_{j}^{J} X_j e^{-\beta tt_{ij}}\right) \quad (1)$$

where X is the number of persons and jobs in a travel analysis zone which is multiplied by a negative exponential weighting based on travel time tt (approximation of generalised travel costs c). β is set to 0.2. In case of acc_car the travel times are calculated on the basis of the car network and in case of public transport (acc_pt) on the basis of public transport network. The cantonal travel model has been used for travel time calculations. Applying the logarithm can lead to negative accessibility values.

Accessibility is a potential which captures the presence of alternatives, while access is a proximity measure to the closest satisfying option. An example of an access variable is the distance to school. This access variable is only based on euclidean distance.

We assume that a development is more likely if it fits the zoning constraints on the parcel under consideration. We capture this fit by calculating the difference between permitted floor-space and floor-space of the development under consideration for a given parcel alternative. The difference is included in the utility function using a step function. On the negative intercept we formulate a linear function with slope of 4. On the positive intercept the natural logarithm is applied expressing the decreasing marginal utility of development reserves.

The newcomers variable is motivated by a builder's statement saying they would analyse demographic development of candidate areas in respect of growth. Thus we try to capture upcoming areas by measuring the influx of people.

The price per permitted floor-space is calculated by multiplying the average land price in the respective municipality by the land area of the respective parcel, divided by permitted floor-space according to the zoning constraints. A parcel with low price per permitted floor-space is expected to be attractive for development.

Land prices vary substantially by zone type (28) and allowed density. To some extent the zone types also reflect centrality (city zone) and cultural heritage (“core” zone). While prices in the previous zone are approximately double of purely residential zones, the prices are about a 5/6 in the latter which reflects more constraints an thus less appropriate built space for commercial uses. Prices in non-residential zones are about two third of prices in residential zones.

With the slope variable we try to capture development costs. Groundwork is often necessary on hillsides which is usually making construction more expensive.

	PLCM				BLCM			
Parameter Name	mean	sd	min	max	mean	sd	min	max
SFH								
newcomers_within_300_of_parcel	310.2	268.3	0.0	2774.0	312.2	269.5	0.0	2867.0
dist_to_school	527.2	361.9	4.0	2868.0	524.1	348.4	3.0	2868.0
acc_car	9.7	0.5	0.0	11.0	9.7	0.5	0.0	11.0
acc_pt	10.7	1.8	-19.0	13.0	10.7	1.9	-19.0	13.0
dev_fit_permitted_floor_m2_step_ln	-894.5	2438.3	-50099.6	10.5	-801.6	1185.6	-18563.6	10.5
price_permitted_floor_m2	2414.2	1494.1	359.3	11015.0	2429.9	1502.9	325.0	11015.0
new_bldgs_within_150_of_parcel	5.5	9.3	0.0	86.0	5.4	9.2	0.0	107.0
slope	5.7	4.4	0.0	31.1	5.7	4.4	0.0	31.1
MFH								
newcomers_within_300_of_parcel	312.5	269.9	0.0	2851.0	317.5	278.1	0.0	2867.0
acc_car	9.7	0.5	0.0	11.0	9.7	0.5	0.0	11.0
acc_pt	10.7	1.8	-19.0	13.0	10.7	1.8	-19.0	13.0
dev_fit_permitted_floor_m2_step_ln	-3979.4	5583.3	-90548.8	10.7	-3460.7	3307.1	-31857.0	11.1
price_permitted_floor_m2	2442.0	1536.6	299.1	11015.0	2412.4	1508.8	299.1	11015.0
new_bldgs_within_150_of_parcel	5.3	9.0	0.0	107.0	5.4	9.2	0.0	107.0
slope	5.7	4.5	0.0	31.1	5.7	4.4	0.0	31.1
With Side Use								
dev_fit_permitted_floor_m2_step_ln	-5858.3	8360.3	-52911.0	10.7	-5344.3	5313.0	-23478.6	11.1
new_bldgs_within_150_of_parcel	5.3	9.2	0.0	66.0	5.3	8.9	0.0	86.0
acc_car	9.7	0.5	8.0	11.0	9.8	0.5	8.0	11.0
acc_pt	10.7	1.6	-19.0	13.0	10.8	1.6	-19.0	13.0
price_permitted_floor_m2	2389.4	1563.7	299.1	11015.0	2364.1	1432.5	299.1	11015.0
Non-Residential								
newcomers_within_300_of_parcel	291.8	283.7	0.0	1967.0	306.3	292.7	0.0	1957.0
acc_car	9.6	0.6	0.0	11.0	9.6	0.5	8.0	11.0
acc_pt	9.9	3.2	-19.0	13.0	10.0	3.0	-19.0	13.0
dev_fit_permitted_floor_m2_step_ln	-8572.5	22477.2	-173320.0	12.8	-7096.6	16178.9	-92838.0	12.9
price_permitted_floor_m2	1165.4	668.5	196.9	5188.6	1154.2	654.6	196.9	5188.6
new_bldgs_within_150_of_parcel	3.3	5.3	0.0	61.0	3.4	5.4	0.0	65.0
slope	3.8	3.3	0.0	27.3	3.7	3.1	0.0	27.3

Table 2 Descriptives of Observation Variables

Model estimation

The models are estimated with the development version of UrbanSim employing the B-triple-H algorithm (29) for maximum likelihood estimation.

RESULTS AND INTERPRETATION

The estimation results all four sub-models are shown in table 4 according to the data bases used. On the left side we list the parameters and t-values for model estimated with project information, on the right side the estimates using single buildings from the register only. We first want to focus on the variables hypothesised to be sensitive to consideration of projects versus buildings, which are namely *dev_fit_permitted_floor_m2* and *new_bldgs_within_150_of_parcel*.

The variable *dev_fit_permitted_floor_m2* shows a positive sign in all models indicating that developments locate on parcels with more floor-space permitted than used, hence development reserves are valued confirming previous research (30). To compare the models we calculate the value of reserves[1] (VOR) as the exchange rate of price per permittted squaremeter of floor-space with reserves for additional floor-space. We generally expect a positive value since we expect capitalisation on the real option of development reserves in the price per permitted floor space.

$$VOR = \frac{\beta_{dev-fit}}{\beta_{permitted-floor-price}} \quad (2)$$

The exchange rates shown in table 3 confirming the expectation of positive values of reserves. Only in case of residential buildings with side use the value is negative which might be an artefact of the small sample size.

Further we expect that VOR is overestimated on the basis of single projects because the parcel is actually filled with additional buildings of the same project. So, we are measuring reserves when there actually are none. This hypothesis is confirmed for developments with residential use but not for non-residential developments.

All models show positive signs for the variable measuring new developments in the vicinity. This confirms spatial inertia of land development found in previous studies (8, 21). The reasons for spatial inertia are manifold. The phenomenon reflects to some extent that settlements grow at their boarders. Another reason can be seen in the intention of planning authorities to concentrate development. A third explanation might be that builders tend to develop in areas they are familiar with. This hypothesis can not be verified at this stage. To do so requires observing which projects have been built by which builder. Qualitative research in Switzerland (26) and elsewhere (31) suggests that there is a fraction of local developers which operate mainly in areas they are familiar with.

Analogous to the previous example we calculate an exchange ratio to compare the model estimates on different data bases. We calculate the exchange rate of price per permittted squaremeter of floor-space with the number of new buildings in the vicinity. This gives us a value of new houses[2] (VONH) as

$$VONH = \frac{\beta_{new_{bldgs}}}{\beta_{permitted-floor-price}} \quad (3)$$

1 The unit is CHF/m2 floor-space reserve.
2 The unit of this value is CHF/new building.

	PLCM		BLCM	
	VDOR	VONH	VDOR	VONH
SFH	0.03	333.68	0.03	343.50
MFH	2.31	158.71	2.52	139.16
RWSU	-1.19	-29.08	-1.40	-40.27
Non-residential	0.70	75.90	0.63	76.54

Table 3 Exchange Rates

The calculated exchange rates are shown in table 3. Thus VONH is overestimated on the basis of buildings in case of single family developments and non-residential developments but not in the two other cases, which is unexpected. The reason might be insufficient accuracy of prices crudely considered dynamics. VONH is way larger than VOR and has also a much bigger range. This might reflect different locations and the importance of those.

All parameters except for slope in case of non-residential developments show consistent signs which indicates that single buildings are a viable option. We interpret the negative signs of the accessibility variable with preferences for quite, remote locations of single family home builders anticipating the preferences of potential costumers. In case of multi family housing the signs of the public transport accessibility variable becomes positive. This expresses preferences for denser living environments we can expect for this housing type. The negative sign for *acc_car* is expressing that MFH developments tend to locate in dense areas where *acc_car* is worse relative to *acc_pt* then in less dense areas. We find the same signs for residential developments with side uses. For non-residential buildings the signs of all accessibility variables are positive showing preferences for dense environments of firms.

We measure the attractiveness for residents with the variable newcomers_within_300_of_parcel and expect positive signs for residential developments. We have significant[3] estimates in case of single family housing, multi family housing projects and non-residential developments. The negative values for single family developments might be an artefact of lacking opportunities to move to in the vicinity, so newcomers are less frequently observed in areas of single family developments. Thus latent demand cannot be captured with this variable for SFH. One might better use absorption rates. The negative sign for non-residential development is interpreted as an effect of competition. In areas with a lot of influx of residents housing developments are preferred.

The slope variable is only significant in case of single family developments. The positive sign shows preference for location on hillsides. The model could be improve in this aspect by considering aspect and solar exposition.

The distance to school is only included for single family housing because we can expect families to live there. The expected negative sign shows in both data bases and the estimates are significant. Thus builders try to build near schools to satisfy that need.

For the price variable we would generally expect a negative sign indicating that builders tend to buy and develop land which is cheep. However, the signs in our models show positive signs for single family housing projects, multi family housing projects and non-residential projects and buildings. Also the effect of the price variable is surprisingly small. One explanation is that developers expect higher revenue from letting or higher expected sales price at these locations. Another possibilty is that the permitted floor space prices variable is insufficiently observed and thus the models probably suffer from an endogeneity problem as described by Guevara and Ben-Akiva (32).

3 We consider an estimate as significant on a 95% level.

	PLCM		BLCM	
Parameter Name	estimate	t-values	estimate	t-values
SFH				
acc_car	-0.3707	-5.17	-0.3740	-5.52
acc_pt	-0.0142	-0.65	-0.0216	-1.00
dev_fit_permitted_floor_m2_step_ln	0.0010	11.22	0.0012	10.95
dist_to_school	-0.0005	-3.47	-0.0005	-3.94
new_bldgs_within_150_of_parcel	0.0493	13.24	0.0512	14.17
newcomers_within_300_of_parcel	-0.0025	-7.61	-0.0024	-7.77
price_permitted_floor_m2	0.0001	5.21	0.0001	5.38
slope	0.0366	3.34	0.0349	3.29
# Observations	501		523	
LL(0)	-1703.9999		-1778.8262	
LL(conv.)	-1455.2932		-1493.9691	
Adj. rho2	0.141		0.156	
MFH				
acc_car	-0.1261	-1.04	-0.0800	-0.67
acc_pt	0.1628	3.63	0.1728	3.92
dev_fit_permitted_floor_m2_step_ln	0.0005	17.27	0.0006	16.87
new_bldgs_within_150_of_parcel	0.0339	5.74	0.0338	6.33
newcomers_within_300_of_parcel	0.0009	3.50	0.0008	3.63
price_permitted_floor_m2	0.0002	6.78	0.0002	7.97
slope	0.0115	0.80	0.0107	0.78
# Observations	405		445	
LL(0)	-1377.4849		-1513.5328	
LL(conv.)	-1168.1905		-1230.9202	
Adj. rho2	0.147		0.182	
With Side Use				
acc_car	-0.7055	-1.80	-0.4893	-1.27
acc_pt	0.4154	2.30	0.3273	2.19
dev_fit_permitted_floor_m2_step_ln	0.0005	4.69	0.0005	4.87
new_bldgs_within_150_of_parcel	0.0111	0.31	0.0148	0.50
price_permitted_floor_m2	-0.0004	-2.59	-0.0004	-2.75
# Observations	54		65	
LL(0)	-183.6647		-221.0778	
LL(conv.)	-143.4228		-157.4724	
Adj. rho2	0.192		0.265	
Non-Residential				
acc_car	0.4569	1.42	0.6646	2.23
acc_pt	0.1762	2.04	0.3245	4.03
dev_fit_permitted_floor_m2_step_ln	0.0005	12.96	0.0005	13.56
new_bldgs_within_150_of_parcel	0.0534	2.50	0.0584	3.64
newcomers_within_300_of_parcel	-0.0001	-0.15	-0.0012	-2.97
price_permitted_floor_m2	0.0007	4.74	0.0008	5.52
slope	0.0152	0.39	-0.0230	-0.64
# Observations	84		114	
LL(0)	-285.7006		-387.7365	
LL(conv.)	-196.0938		-261.6918	
Adj. rho2	0.289		0.307	

Table 4 Parameter Estimates of the Models, PLCM = Project Location Choice Model, BLCM = Building Location Choice Model, SFH = Single Family Housing, MFH = Multi Family Housing, RWSU = Residential with Side Use

The model statistics show only 54 and 84 observations the case of residential developments with side use and non-residential development respectively. According to the rule of thumb saying we need 30 observation per estimated parameter suggest the high adjusted rho2 shows over fitting of the data. We included this many variables for the sake of comparison. The situation of few observations in case of non residential developments is common and a general problem for land development models. In addition the variety of uses is much higher in the non-residential segment which is unfortunate.

CONCLUSIONS

In this paper we investigated the viability of extracting pseudo development event from building register data. The comparison of model estimation on the basis of single buildings with model estimation using development project information shows relatively little variation in the examined case of Canton Zurich. This suggests that building register data may be used for creating land development models. However, one should mind biased results due to the theoretically inadequate observations when taking newly registered buildings as development events. Also, some variables such as the fit of a development to a parcel as location alternative are more sensitive to the observations used. Therefore, it is still advisable to use information on development projects enabling the estimation of more sophisticated models.

The drawbacks are less severe in regions with few multi building projects. Hence, consulting aggregate information on the size of projects, in terms of number of buildings per project, is recommended before implementing a land development model in form of a building location choice model.

Formulating building location choice models offers practical advantages. The most crucial and evident one is the sheer lack of better data in many cases. A practical advantage when implementing a BCLM in the parcel context is the absence of projects that are actually located on multiple parcels. If we formulate a project location choice model, which is theoretically more sound, we have to model that parcels can be unified to accommodate larger projects which we expect due to increasing economies of scale. We may coin this *parcelling problem* which we leave for further research.

We presented location choice models for residential and non-residential developments. Recent literature focused on the residential segment only, probably also due to scarce and very heterogeneous data. For the sake of completeness (which is required for a meaningful overall model, since we also need to provide built space for workplaces) we also include non-residential buildings.

General problems encountered are the scarcity of development events on the non-residential segment, which is unfortunate because the non-residential segment too heterogeneous in terms of uses. We still lack spatio-temporal price information with suitable level of detail. This is owed to high data protection especially in the realm of real estate property (9).

This work focuses on the effects of missing project information when extracting newly registered buildings as development events, which reflects work in progress developing a land development model. We will improve the model by revisiting data preparation to gain more development events, adding further information such as vicinity endowment with green spaces, exposition and sunshine index and more accurate price information. At this point we do not distinguish between new construction on green land and replacement of buildings. It will be interesting to see if we find spatial inertia as well in case of redevelopment. Disregarding redevelopment is especially problematic in our case study region since unbuilt developable land is scarce (28). Finally, we want to add information on builders which take the development decision. We assume that considering ownership structures can considerably improve our understanding of land development produced in real estate markets. Previous studies found that the property structure influences land development (33). This issue is getting more relevant since the built environment is entering a phase of reconstruction and transformation. In terms of model comparison we will compare the performance on the basis of simulation results using similar methods as proposed by Jäggi et al (34) and Dong and Gliebe (21).

Acknowledgements

The data preparation was joint work with Kirill Müller, Patrick Schirmer, Timo Hostschäfer and Andri Mani whose efforts are very much appreciated.

This research has been partly funded through the SustainCity project (FP7-244557), co-financed by the EU within the Seventh Framework Programme (FP7). The authors wish to acknowledge the Commission for their support of the project, the efforts of the partners and the contributions of all those involved in SustainCity.

REFERENCES

[1] Wegener, M., and F. Fürst. *Land-Use Transport Interaction: State of the Art*. Publication 46. Institut für Raumplanung, Universität Dortmund, Dortmund, 1999.

[2] Iacono, M., D. Levinson, and A. El-Geneidy. Models of Transportation and Land Use Change: A Guide to the Territory. *Journal of Planning Literature*, Vol. 22, No. 4, 2008, pp. 323.

[3] Wegener, M. Current and Future Land Use Models. Presented at the 1st Land Use Model Conference, Dallas, 1995.

[4] Hunt, J. D., D. S. Kriger, and E. J. Miller. Current operational urban land-use–transport modelling frameworks: A review. *Transport Reviews*, Vol. 25, No. 3, 2005, pp. 329–376.

[5] Chang, J. S. Models of the Relationship between Transport and Land-use: A Review. *Transport Reviews*, Vol. 26, No. 3, 2006, pp. 325–350.

[6] Geurs, K., B. Zondag, G. de Jong, and M. de Bok. Accessibility appraisal of land-use/transport policy strategies: More than just adding up travel-time savings. *Transportation Research Part D: Transport and Environment*, Vol. 15, No. 7, Oct. 2010, pp. 382–393.

[7] Geurs, K. T., and B. van Wee. Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, Vol. 12, No. 2, Jun. 2004, pp. 127–140.

[8] Haider, M., and E. Miller. Modeling Location Choices of Housing Builders in the Greater Toronto, Canada, Area. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1898, Jan. 2004, pp. 148–156.

[9] DiPasquale, D. Why don’t we know more about housing supply? *The Journal of Real Estate Finance and Economics*, Vol. 18, No. 1, 1999, pp. 9–23.

[10] Farooq, B. A Unified Framework of Urban Built-Space Evolution. Presented at the 11th Swiss Transportation Research Conference, Ascona, 2011.

[11] Martínez, F. MUSSA: Land Use Model for Santiago City. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1552, Jun. 1996, pp. 126–134.
[12] Alonso, W. *Location and Land Use: Toward a General Theory of Land Rent*. Harvard University Press, 1964.
[13] Waddell, P. UrbanSim: Modeling urban development for land use; transportation and environmental planning. *Journal of the American Planning Association*, Vol. 68, No. 3, 2002, pp. 297–314.
[14] Waddell, P., A. Borning, M. Noth, N. Freier, M. Becke, and G. Ulfarsson. Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. *Networks and Spatial Economics*, Vol. 3, No. 1, 2003, pp. 43–67.
[15] Small, K. A., and H. S. Rosen. Applied Welfare Economics with Discrete Choice Models. *Econometrica*, Vol. 49, No. 1, Jan. 1981, pp. 105–130.
[16] Rosen, S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, Vol. 82, No. 1, Jan. 1974, pp. 34–55.
[17] Farooq, B., and E. J. Miller. Towards integrated land use and transportation: A dynamic disequilibrium based microsimulation framework for built space markets. *Transportation Research Part A: Policy and Practice*, Vol. 46, No. 7, Aug. 2012, pp. 1030–1053.
[18] Hurtubia, R., M. Bierlaire, and F. Martínez. Dynamic microsimulation of location choices with a quasi-equilibrium auction approach. Presented at the 12th Swiss Transport Research Conference, Ascona, 2012.
[19] McFadden, D. *Econometric models of probabilistic choice*. MIT Press, Cambridge, 1981.
[20] McFadden, D. MODELING THE CHOICE OF RESIDENTIAL LOCATION. *Transportation Research Record*, No. 673, 1978, pp. 72–77.
[21] Dong, H., and J. Gliebe. Forecasting Location of New Housing in Integrated Models of Land Use. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2255, Dec. 2011, pp. 79–90.
[22] Waddell, P., L. Wang, B. Charlton, and A. Olsen. Microsimulating parcel-level land use and activity-based travel: Development of a prototype application in San Francisco. *Journal of Transport and Land Use*, Vol. 3, No. 2, Sep. 2010.
[23] Zöllig, C., and K. W. Axhausen. A conceptual, agent-based model of land development for UrbanSim. Presented at the 51th ERSA Conference, Barcelona, 2011.
[24] Schirmer, P., C. Zöllig, K. Müller, B. R. Bodenmann, and K. W. Axhausen. The Zurich case study of UrbanSim. Presented at the 51th ERSA Conference, Barcelona, 2011.
[25] Borning, A., H. Ševčíková, and P. Waddell. A domain-specific language for urban simulation variables. Presented at the international conference on digital government research, 2008.
[26] Zöllig, C., and K. W. Axhausen. *Uncovering the heterogeneity of real estate developers in the canton Zurich*. Publication 3.6. IVT, ETH Zürich, Zurich, Feb. 2012, p. 24.
[27] Guo, J. Y., and C. R. Bhat. Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography*, Vol. 15, No. 1, Jan.

2007, pp. 31–45.
[28] Rey, U. *Zentrale Lagen begehrt*. Publication 11/12. Statistisches Amt des Kantons Zürich, Zürich, 2011, p. 17.
[29] Berndt, E. K., B. H. Hall, and R. E. Hall. Style not yet supported for "*Estimation and inference in nonlinear structural models.*"Please consult TRB Formatting Guidelines for assistance.
[30] Thalmann, P. Hoch hinaus oder in die Breite? Presented at the ETH Forum Wohnungsbau 2009, Zürich, 2009.
[31] Ruming, K. Developer Typologies in Urban Renewal in Sydney: Recognising the Role of Informal Associations between Developers and Local Government. *Urban Policy and Research*, Vol. 28, No. 1, Mar. 2010, pp. 65–83.
[32] Guevara, C. A., and M. Ben-Akiva. Endogeneity in Residential Location Choice Models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1977, No. -1, 2006, pp. 60–66.
[33] Triantafyllopoulos, N. Does Land Property Structure Affect Local Development Patterns? Evidence from a Greek Tourist Area. *Urban Studies*, Vol. 45, No. 4, Apr. 2008, pp. 797–824.
[34] Jäggi, B., C. Weis, and K. W. Axhausen. Stated Response and Multiple Discrete-Continuous Choice Models: Analyses of residuals. *Journal of Choice Modelling*, forthcoming.