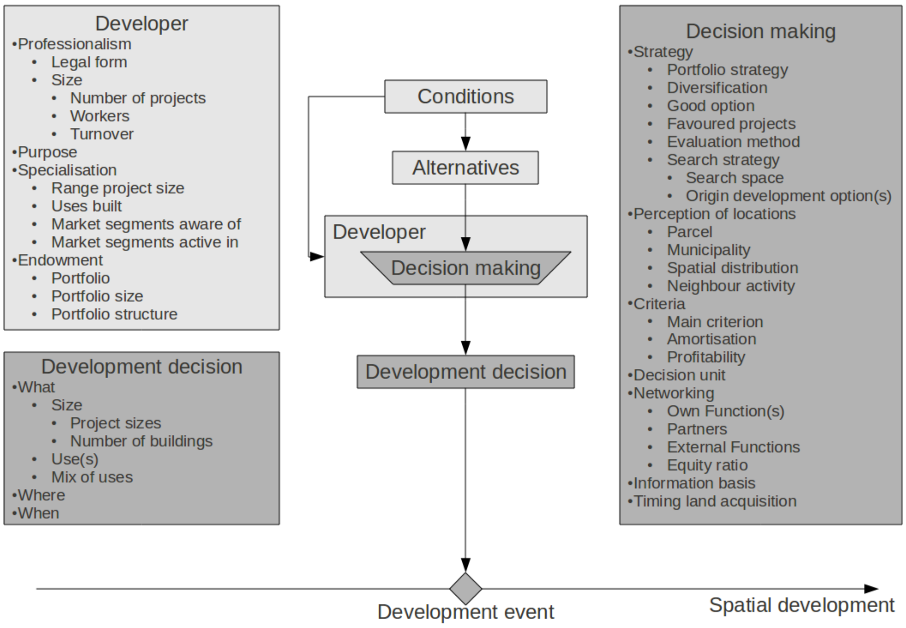

Modelling real estate development with heterogeneous agents

Christof Zöllig Renner

Prof. Dr. Kay W. Axhausen

Institute for transport planning and systems

April 2013

STRC | 13th Swiss Transport Research Conference

Monte Verità / Ascona, April 24 – 26, 2013

Contents

List of Tables

Institute for transport planning and systems

Modelling real estate development with heterogeneous agents

Christof Zöllig Renner
Institute for Transport Planning and Systems
ETH Zurich
Wolfgang-Pauli-Str. 15, CH-8093 Zurich
phone: +41-44-633 27 19
fax: +41-44-633 10 57
christof.zoellig@ivt.baug.ethz.ch

Prof. Dr. Kay W. Axhausen
Institute for Transport Planning and Systems
ETH Zurich
Wolfgang-Pauli-Str. 15, CH-8093 Zurich
phone: +41-44-633 39 43
fax:
axhausen@ivt.baug.ethz.ch

April 2013

Abstract

This paper deals with real estate supply modelling. In includes a quick summary of current approaches and makes an argument for considering heterogeneous real estate developer agents within a microsimulation framework. The hypothesis thus is that heterogeneity of real estate developers matters for the overall development process of an urban area. I.e. an area where all properties belong to the same machine manufacturer has very different development perspectives than an area with dozens of property owners.

On the basis of qualitative research a model is developed with heterogeneous agents supplying built space. A prototype is estimated in UrbanSim using a multinomial logit location choice model while discriminating three developer types, for each of which a submodel is presented. The discussion also includes data requirements and data preparation. The preliminary results show improvement of the models compared to models based on development type.

Keywords

Land use development, location choice, Multinomial logit (MNL) model, Zurich

1 Introduction

Spatial development of urban regions can be conceptualised as a series of decisions take by landowners. Thus we can try to model spatial development using discrete choice models. For a full description of a decision situation we should consider the decision maker, the alternatives and circumstances . We assume that the owner of a parcel decides on construction of built space within the regulations imposed. In the case of land development decisions this means to include developers into our analysis of development decisions which is rarely done in the context of Land Use Transport Interaction (LUTI) modelling. Therewith we define landowners which do development projects as developers.

In a first step we investigate in this paper if different developer types have different preferences in terms of location. We start examining location choice since we can argue that any developer must have made his location choice before he can start constructing since a piece of land is the principle resource for real estate construction. Also, location is very important because it can not be corrected at a later stage. However, the models presented here have to be seen as a first step towards a more general model which would ideally not only determine the location of a development project but also its size and characteristics.

We investigate different location choice behaviour by estimating discrete choice models. We group our observations of development projects according to developer type and analyse the differences in the effects of explanatory variables. The preparation of the necessary observations is described in section 3. Estimation results are shown and discussed in section 4 followed by conclusions and an outlook. The context is sketched in section 2.

2 Land development models

Land Use Models (LUMs) have been developed since von Thunen (1826) to have a better understanding of how land use evolves in a given area. In this context use means activities of economic actors, which use space for their purpose. In the beginning it was aggregate spatio-temporal modelling (Hotelling, 1929, Christaller, 1933, Lösch, 1962). Alonso (1964) developed bid-rent theory which sees prices and locations as the outcome of auction or bidding processes of urban actors. A convincing feature of this model is endogenous price determination within market clearing. Equilibria are assumed in these models. The next generation of models is based on microeconomics and thus are disaggregate. Further development of the bid-rent theory resulted in bid-choice theory (Martínez, 1992). This bid-choice theory is based on the assumption of consumer surplus maximising behaviour, which is shown to be the same as applying the utility maximisation approach of McFadden to location choice (McFadden, 1978). These models assume equilibria to determine the prices. Martínez and Donoso (2010) extended their model to include a supply model. The model is still base on static market equilibrium assumption. Another category of models operates as well on the basis of microeconomics and discrete choice theory but does not assume equilibria to solve for the prices. These models determine the prices not via equilibrium search. In the case of UrbanSim the prices are determined with hedonic regressions (Waddell, 2002b). Economic actors like firms and households react to price changes with their location choice. The location choices made then have again an influence on price in the next iteration. Recent work by Wang and Waddell (2013) suggest a price formation model reflecting competition with heterogeneous households. Before households and firms can choose their facility for activity performance the facilities have to be built. This process of built space provision is captured with Land Use Development Model (LUDM). Thus the purpose of this models with the modelling system is to provide facilities for the economic actors to locate. This process has been simplified in most previous models representing the supply side with a single representative agent which reacts to demand. Few previous studies attempted to capture heterogeneity within real estate developers (Dong and Gliebe, 2010, Haider and Miller, 2004).

3 A developer based land use development model

In this section we describe a first model of developer location choice discriminating three different developer types. In the following subsection the method is briefly described. Data and data preparation are discussed afterwords. The last subsection contains the description of the model.

3.1 Method

The method used in this study is discrete choice modelling (Train, 2009). The first step was a qualitative study (Zöllig and Axhausen, 2012b) to inform the modelling step presented here. Eleven developers were interviewed to get insights into the variety of developers in the Canton Zurich and their decision making. The qualitative part followed the methodology of Gläser and Laudel (2004) and suggested behavioural differences of developers according to purpose and professionalism. In this next step presented here we try to confirm the qualitative findings by estimating MNL choice models using UrbanSim (Waddell, 2002a, UrbanSimProject, 2007) made operational for the Canton Zurich. In a first approach we estimate separate models for three developer types and compare the effect and impact of considered variables.

3.2 Data and preparation

We have a dataset at hand which holds information on development projects and their developers. The dataset was bought from the firm DOCUMEDIA. We decide to use the DOCUMEDIA data to relate qualitative and quantitative analysis since it included contact details as well as an extensive number of projects from 2000 until 2010. However, there is no guarantee that the dataset contains all development projects carried out in that period.

In addition to the dataset which relates development projects and developers we used a lot of other data. The following list contains the most important data sources for study:

1. DOCUMEDIA
2. Federal Building and Housing Register (FBHR)
3. Cadastre data containing parcels
4. Land use plans
5. Population cencus 2000
6. Employment cencus 2001
7. Microcensus of travel behaviour
8. MATSim road network
9. Public transport stops of the cantonal office for spatial development with an impedance matrix

Most important for this study are the datasets of DOCUMEDIA and FBHR since they contain the information on developers and buildings. In the DOCUMEDIA data we have information on purpose and contact details regarding the developers. From the contact details we can infer the frequency of developers doing a project. We use this here as a proxy for professionalism.

The FBHR data contains more detail on the buildings constructed. Therefore we match the DOCUMEDIA and FBHR datasets via a spatio temporal join. We use an address matching procedure to match DOCUMEDIA projects to parcels and match then the buildings located on that parcel to the project. Here we only use projects in which new buildings are constructed. Due to the matching the number of observations is reduced from 21384 to 1301.

The rest of the data is used to enrich the development projects with location variables. These types of variables are supposed to be highly relevant for developers location choices. The qualitative study confirmed this assumption while showing as well other determinants of real estate development decisions like financial aspects, information used for decision making, evaluation methods, business strategy, search strategy and search space. Unfortunately, we find only few of these potentially explanatory variables in the data at hand. Thus, we focus on how

Table 1: Developer type definition

Developer type code	Developer type name	Attribute purpose	Number of projects
O1	Self-owning without portfolio strategy	Own use or letting	1
Om	Self-owning with portfolio strategy	Own use and or letting	Several
Smc	Commercial developer (Promoter)	Sale	Several or 1

location attributes are evaluated differently.

For each parcel we also calculate its accessibility using Multi-Agent Transport Simulation (MATSim). The usage of highly detailed transport simulation allows us to calculate individual accessibilities for each parcel for different modes. Here we use car and Public Transport (PT). Following Geurs and van Wee (2004) we use a location-based potential accessibility measure.

The result of the data preparation is a database comprising parcels with planning constraints, buildings with living units, households with persons and employment which are related with each other. Households are located in living units, living units are associated to buildings and buildings are standing on a parcel. This abstraction of spatial reality is the context of the development events. These events are the observation used for model estimation.

3.3 Model formulation

The prepared data is imported into UrbanSim where the modelling and model estimation is performed. We specify three sub-models within an agent location choice model according to three categories of developers. The categories and their definition according to the data used is shown in table 1. In this study we only use observations of projects residential . These can be of different type such as single family homes, multi family homes or residential buildings with mixed use. The overall choice set are all parcels in the Canton Zurich which are zoned for housing and have still capacity for new built space. Thirty alternatives are randomly sampled from the overall choice set. We estimate MNL models with a linear utility function.

Table 2 shows the definitions of the variables used. Most of these variables were used in the previous model which discriminates development types (Zöllig and Axhausen, 2012a). The variable of newcomers counts only persons younger than 30, a variable measuring the share of recreation areas, within the travel analysis zone is introduced.

The accessibilities are calculated with formula 1 in MATSim. The general travel costs c include only travel time here. MATSim is able to include other cost components such as toll costs .

$$Acc_i = \sum_j^J X_j e^{-\beta c_{ij}} \tag{1}$$

where

i= location of accessibility calculation
j= activity location index
X= number of persons and jobs
c= generalised travel costs
J= number of all activity locations considered
β= 0.2 (estimated parameter)

The expectations in terms of positive or negative effect on utility and differences in impact according to developer type are shown in table 3. The hypotheses are discussed in the following results section 4.

4 Estimation results and discussion

This section presents and discusses the model estimation. The results are very preliminary and have to be interpreted with care. Table 4 shows the signs and significance levels of the estimated parameters. We only show the signs due to unstandardised estimation. In the following we discuss these preliminary results on the basis of the hypothesis in table 3. We discuss first the signs of the variables and the occurring or not occurring differences.

Car accessibility has a negative sign which is unexpected. The negative sign might captures that new development often are at the fringe of the settlement where lots are cheaper. This is an endogeneigty problem Guevara and Ben-Akiva (2006) which demands for integrating an appropriate control variable.

In case of PT accessibility the signs are as expected. The low impact in case of *O1*- and *Om*-developers is surprising.

Table 2: Variable definitions

Variable Name	Description	Unit
Accessibility car	Accessibility of jobs and inhabitants by car according to formula (1). Travel times are calculated with MATSim.	[-]
Accessibility public transport	Accessibility of jobs and inhabitants by public transport according to formula (1). The calculations of travel times are based on the cantonal transport model.	[-]
Fit of development to parcel constraints	Step function differentiating between permitted and planned 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 neighbouring buildings	Number of buildings with year built later than 1995 within 150 m.	[-]
Young newcomers in neighbourhood	Number of residents younger than 30 which reported a different address five years ago within a radius of 300 m.	[-]
Price per permitted floor space	Price per permitted square meter floor-space.	[CHF/m2]
Slope	Slope of parcel in percent.	[%]
Share of recreation are in zone	Share of land area dedicated to recreational use within the traffic analysis zone of the development. This comprises parks, commons, camping and sport grounds.	[-]

Table 3: Expectations of variable effects

Variable	Hypothesis
Accessibility car	Positive sign. More impact for selling developers.
Accessibility public transport	Positive sign. More impact for selling developers
Fit of development to parcel constraints	Positive sign. We expect stronger impact for optimal fit for developers with purpose sale.
New neighbouring buildings	Positive sign.
Price per permitted floor space	Positive sign for type Smc, negative for type O1 and type Om.
Immigration of young households	Positive sign for type Smc and Om since they produce for the market. Type O1 might have negative preference due to noise concerns.
Slope	Negative sign since it is more complicated to build on. We expect stronger preference for types Smc and Om assuming more cost sensitivity.
Share of recreation are in zone	Positive sign, while less impact on developers of type Smc and Om since they are not personally affected.

For the variable describing the fit of a development project to the allowed volume on parcel has the expected sign and is of high significance. This suggests that further interaction variables could improve the model. However, it would be preferable to control as well the type of development to see if in cases of single family houses less dense development occurs.

All developer types have preferences for new buildings in the neighbourhood. The recent development decisions of others seems to give confidence rather than fear of too much competition.

The positive sign of land price is irritating for developers which develop for their own use. For developers which sell or let the estate later on higher priced land might be an indicator for good sales. The hypothesis regarding this variable is thus only confirmed for selling developers. Again it would be advisable to control for the type of use of the development. There could also be a correlation issue with the accessibility variables.

The positive sign for the endowment of a location with recreation areas cannot be found for *O1*-developers where we would have expected the strongest preference due to personal attachment and interest. The only developer type with a significant effect is *Smc*.

For the developer types *O1* and *Smc* we find counter intuitive signs for slope. Maybe the slope variable captures the positive effect of view.

Under the assumption that people move to attractive areas it is surprising that the sign for newcomers is negative. We expected that developers would see that as a signal for increasing demand. The variable is not discriminating according to income which might be necessary for a consistent result. Another interpretation is that the negative sign indicates disliking of unsettled neighbourhoods.

The adjusted likelihood ratio index of all three models is on the same level. However, the developer based models have less modal fit then those with segments according to development type (Zöllig and Axhausen, 2012a).

Table 4: Signs and significance of estimates parameters

	O1		Om		Smc	
	Sign	Sign.	Sign	Sign.	Sign	Sign.
Accessibility car	-	***	-	***	-	***
Accessibility PT	+	*	+		+	***
Fit of development to parcel constraints	+	***	+	***	+	***
New neighbouring buildings	+	***	+	*	+	***
Price per permitted floor space	+		+	***	+	***
Share of recreation are in zone	-		+		+	**
Slope	+	***	-		+	**
Young newcomers in neighbourhood	-	***	-	**	-	**
Adj. likelihood ratio index:		0.112		0.112		0.106
Number of observations:		340		320		443
Level of significance	1.0%	***	5.0%	**	10.0%	*

We also want to discuss the impact of the variables in respect of developer types. The estimated impact *I* is calculated with the following formula 2:

$$I = \frac{|\overline{(x_{ij})} * \beta_j|}{\sum_{j=1}^{J}(|\overline{(x_i)} * \beta_j|_j)} \tag{2}$$

with

i= alternative index
j= variable index
J= number of variables
β= estimated parameter
x= characteristic

The impacts of the developer models in table 5 show that car accessibility is the dominant variable for all three developer types. The hypothesis that car accessibility is more relevant for selling developers is not confirmed. This is only supported by the results in case of PT accessibilities.

We expect selling developers to exploit allowed density more rigorously than other developer types. This is only confirmed in respect of *O1* developers. The impact on *Om* is more than twice as large.

The hypothesis that more professional developer types *Om* and *Smc* are more cost sensitive and thus react stronger to steeper slope as a proxy of construction cost cannot be confirmed.

Table 5: Impact of variables

	O1	Om	Smc
Accessibility car	0.878	0.628	0.796
Accessibility PT	0.020	0.040	0.088
Fit of development to parcel constraints	0.025	0.180	0.072
New neighbouring buildings	0.010	0.008	0.006
Price per permitted floor space	0.002	0.088	0.012
Share of recreation are in zone	0.004	0.013	0.007
Slope	0.010	0.007	0.008
Young newcomers in neighbourhood	0.051	0.035	0.010

For completeness we put the estimated parameters and their t-values in table 7 and the descriptives of the estimation data in table 6 of the appendix.

5 Conclusion and outlook

In this paper we include some characteristics of developers to explain better their development decisions in terms of location choice. Therefore, we combined a dataset of DOCUMEDIA including development projects and developer information with the Federal Building and

Housing Register. The preliminary model estimates show unexpected signs and only little difference regarding the used explanatory variables. The fit of the models is not increased in comparison to development based models. Further work has to be done to find better explanatory variable according to developer types. Special attention should be paid to endogeneity issues. With the specification investigated here it is not recommended to use a developer based land development model. The reasons for the unsatisfying results can be seen in the following points:

- Data preparation includes some uncertainties and lacks a rigorous representation of the point in time a development took place.
- The development type is not controlled for even though one would assume that the same developer type has also different preferences for different project types.
- Reconstruction is not considered.
- The price variable is not adjusted to the time of the observation.

The next steps will be to estimate models which interact information about developer type and type of development. Also we have to explore the correlation among the explanatory variables with factor analysis. Further we want to investigate the role of the ownership structure. If the model estimation reveals the importance of the ownership structure it will be a next task to synthesise it for simulation.

6 References

Alonso, W. (1964) *Location and Land Use: Toward a General Theory of Land Rent*, Harvard University Press.

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

Dong, H. and J. Gliebe (2010) Exploring the taste heterogeneity in home developers' location choice, Denver.

Geurs, K. T. and B. van Wee (2004) Accessibility evaluation of land-use and transport strategies: review and research directions, *Journal of Transport Geography*, **12** (2) 127–140, ISSN 0966-6923.

Gläser, J. and G. Laudel (2004) *Experteninterviews und qualitative Inhaltsanalyse als In-*

strumente rekonsturierender Untersuchungen, 2 edn., VS Verlag für Sozialwissenschaften, Wiesbaden, ISBN 3825223485.

Guevara, C. A. and M. Ben-Akiva (2006) Endogeneity in residential location choice models, *Transportation Research Record: Journal of the Transportation Research Board*, **1977** (-1) 60–66.

Haider, M. and E. Miller (2004) Modeling location choices of housing builders in the greater toronto, canada, area, *Transportation Research Record: Journal of the Transportation Research Board*, **1898**, 148–156.

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

Lösch, A. (1962) *The economics of location*, 3 edn., Fischer, Stuttgart.

Martínez, F. J. (1992) The bid - choice land-use model: an integrated economic framework, *Environment and Planning A*, **24** (6) 871 – 885.

Martínez, F. J. and P. Donoso (2010) The MUSSA II land use auction equilibrium model, in F. Pagliara, J. Preston and D. Simmonds (eds.) *Residential Location Choice*, Advances in Spatial Science, 99–113, Springer Berlin Heidelberg, ISBN 978-3-642-12788-5.

McFadden, D. (1978) Modeling the choice of residential location, in A. Karlqvist (ed.) *Spatial Interaction Theory and Residential Location*, 75–96, North-Holland, Amsterdam.

Train, K. E. (2009) *Discrete Choice Methods with Simulation*, 2 edn., Cambridge University Press, Cambridge, ISBN 0521017157, 9780521017152.

UrbanSimProject (2007) WebHome, http://www.urbansim.org/Main/WebHome, `http://www.urbansim.org/Main/WebHome`.

von Thunen, J. H. (1826) *Der isolierte Staat in beziehung auf Landwirtschaft und Nationalokonomie, Gustav Fisher, Stuttgart. English edition: The isolated state (trans: Wartenburg CM (1966), edited by Hall P)*, Pergamon, Oxford.

Waddell, P. (2002a) UrbanSim: modeling urban development for land use; transportation and environmental planning, *Journal of the American Planning Association*, **68** (3) 297–314.

Waddell, P. A. (2002b) UrbanSim: Modeling urban development for land use, transportation, and environmental planning, *Journal of the American Planning Association*, **68** (3) 297–314.

Wang, L. and P. Waddell (2013) Disaggregated real estate demand model with price formation for integrated land use and transportation modeling.

Zöllig, C. and K. W. Axhausen (2012a) Assessment of infrastructural investment using agent based accessibility, in K. T. Geurs, K. J. Krizek and A. Reggiani (eds.) *Accessibility Analysis and Transport Planning*, 1 edn., NECTAR Series on Transportation and Communication Networks Research, 54–70, Edward Elgar Publishing, Cheltenham, UK, ISBN 9781781000106.

Zöllig, C. and K. W. Axhausen (2012b) Heterogeneity of real estate developers in canton zurich, Ascona.

Acronyms

FBHR Federal Building and Housing Register. 3

LUDM Land Use Development Model. 2
LUM Land Use Model. 1
LUTI Land Use Transport Interaction. 1

MATSim Multi-Agent Transport Simulation. 4, 5
MNL Multinomial logit. ii, 2, 4

PT Public Transport. 4, 5, 9

A Description of used data

A.1 Descriptives of estimation datasets

Table 6: Variable descriptives

Submodel O1	mean	sd	sum	min	max
Young newcomers in neighbourhood	152.4	162.13	1554480	0	1458
Share of recreation are in zone	1.11	0.54	11285.5	0.0915037	14.165
Accessibility PT	10.69	2.36	109022	-18.5865	12.897
Accessibility car	13.83	0.28	141114	13.0168	14.4499
Fit of development to parcel constraints	-925.54	2411.67	-9440510	-31898.4	11.6141
Price per permitted foor space	2173.73	1400.16	22172000	299.091	11015
New neighbouring buildings	4.63	8.67	47205	0	107
Slope	5.06	4.06	51590.1	0	24.3175
Submodel Om					
Young newcomers in neighbourhood	155.53	168.79	1493080	0	1450
Share of recreation are in zone	1.11	0.55	10699.4	0.63652	14.165
Accessibility PT	10.76	2.06	103253	-18.5865	12.897
Accessibility car	13.84	0.31	132819	0	14.4462
Fit of development to parcel constraints	-3460.67	10164	-33222400	-130840	10.6978
Price per permitted foor space	2244.53	1479.77	21547500	258.039	11015
New neighbouring buildings	4.52	8.67	43345	0	107
Slope	5.08	4.06	48721.5	0	24.8381
Submodel Smc					
Young newcomers in neighbourhood	154.14	165.8	2048560	0	1495
Share of recreation are in zone	1.1	0.49	14612.7	0.0915037	6.98158
Accessibility PT	10.7	2.31	142203	-18.5865	12.897
Accessibility car	13.84	0.27	183894	13.0327	14.4517
Fit of development to parcel constraints	-3484.44	6218.47	-46308200	-70676	10.9717
Price per permitted foor space	2176.37	1425.5	28923900	299.091	11015
New neighbouring buildings	4.67	8.7	62103	0	107
Slope	5.03	4.08	66834	0	30.8969

A.2 Estimates and t-values

Table 7: Parameter estmates

	O1		Om		Smc	
	estimate	t-values	estimate	t-values	estimate	t-values
Accessibility car	-1.5709	-6.0	-0.3513	-3.7	-1.2729	-5.4
Accessibility PT	0.0468	1.6	0.0287	0.9	0.1822	3.1
Fit of development to parcel constraints	0.0007	7.7	0.0004	16.9	0.0005	18.5
New neighbouring buildings	0.0513	11.1	0.0145	1.6	0.0306	5.4
Price per permitted floor space	0.0000	0.5	0.0003	9.4	0.0001	3.5
Share of recreation are in zone	-0.0952	-0.7	0.0936	1.1	0.1505	1.8
Slope	0.0469	3.7	-0.0103	-0.6	0.0331	2.5
Young newcomers in neighbourhood	-0.0083	-8.7	-0.0017	-2.1	-0.0014	-2.1
Adj. likelihood ratio index:		0.112		0.112		0.106
Number of observations:		340		320		443