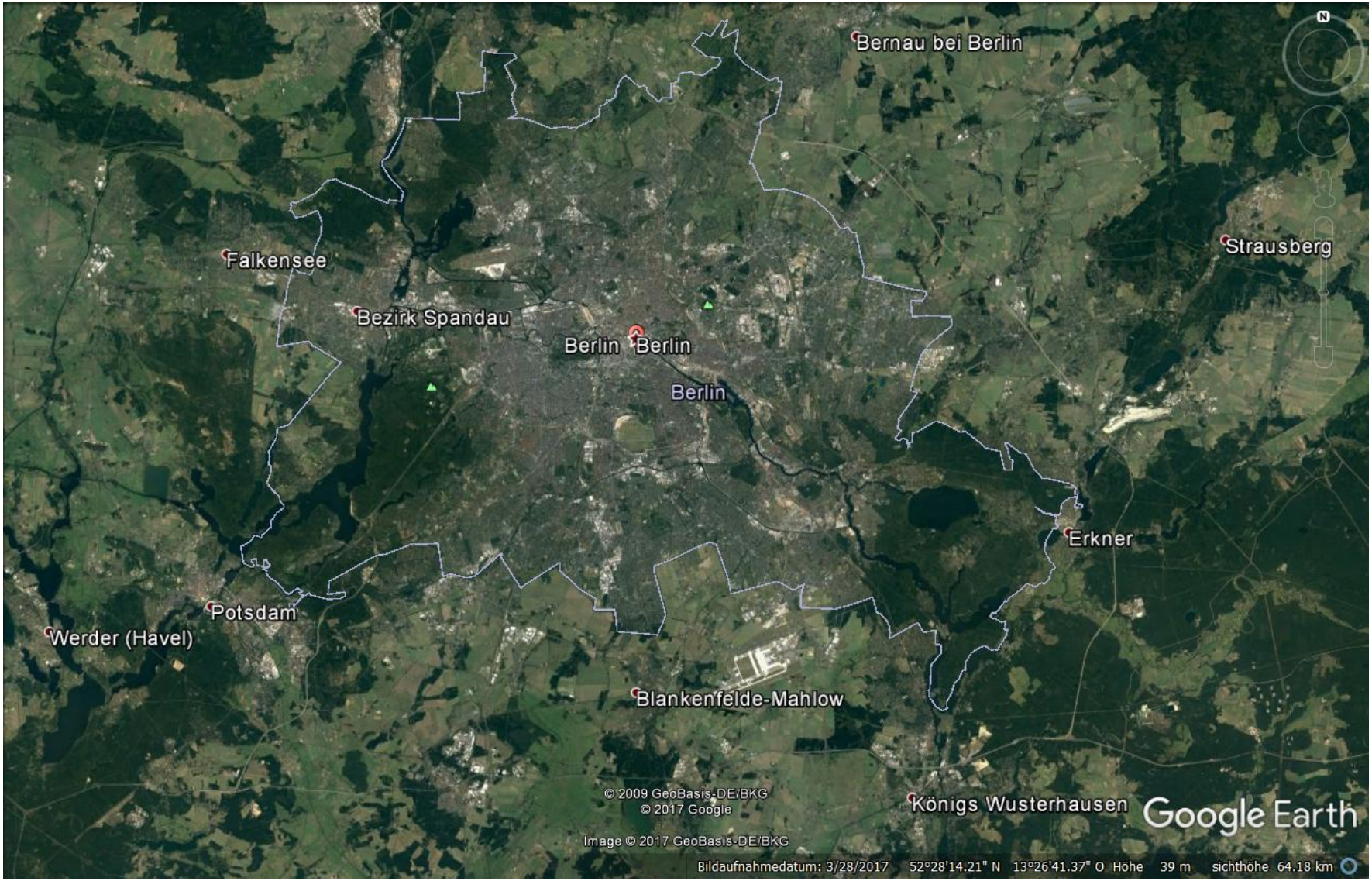
Estimating bid-auction models of residential location using census data with imputed household income

Benjamin Heldt
Francisco Bahamonde-Birke
Pedro Donoso
Dirk Heinrichs

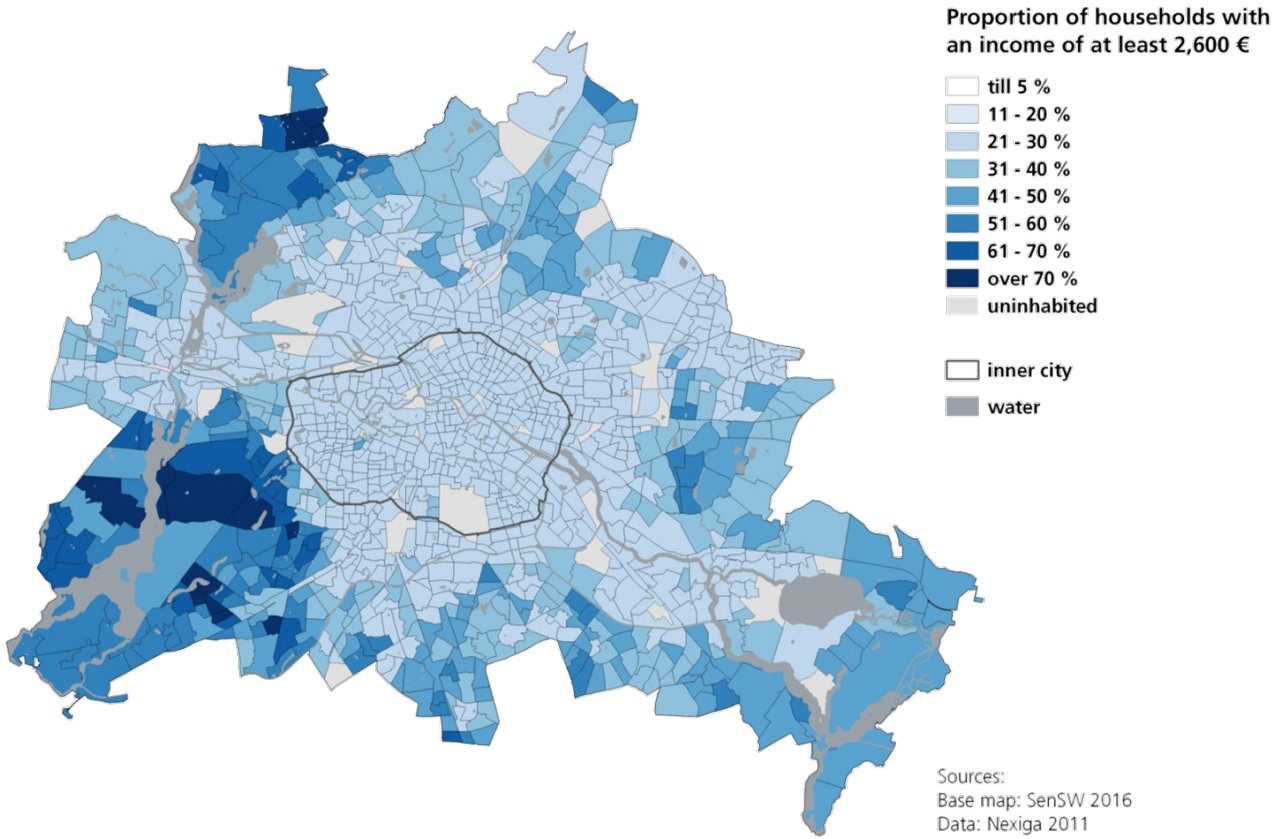
DLR


Bernau bei Berlin
Falkensee
Strausberg
Bezirk Spandau
Berlin
Berlin
Berlin
Erkner
Potsdam
Werder (Havel)
Blankenfelde-Mahlow
Königs Wusterhausen
© 2009 GeoBasis-DE/BKG
© 2017 Google
Image © 2017 GeoBasis-DE/BKG
Google Earth
Bildaufnahmedatum: 3/28/2017 52°28'14.21" N 13°26'41.37" O Höhe 39 m sichthöhe 64.18 km

DLR

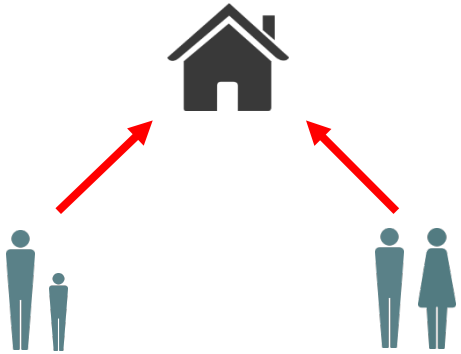
Income determines residential location choice

Which locations can a household afford?
Income influences preferences and life style
Which mobility resources do they have?



Different types of location choice models

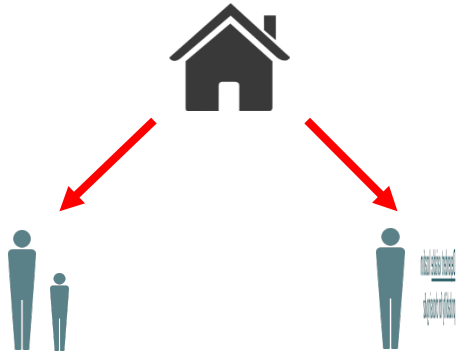
Utility-based models:
Households choose locations that maximise their utility (McFadden 1978)



$$P_{(dz|h)} = \frac{\exp(V_{hdz})}{\sum_{dz} \exp(V_{hdz})}$$

Dependent variable: location probability for choosing ***dz***

Bid-auction - based models:
Owners of locations choose households with highest bid (Alonso 1964, Martínez 1992)



$$P_{(h|dz)} = \frac{H_h \exp(\mu B_{hdz})}{\sum_g H_g \exp(\mu B_{gdz})}$$

Dependent variable: location probability for choosing (tenant) ***h***

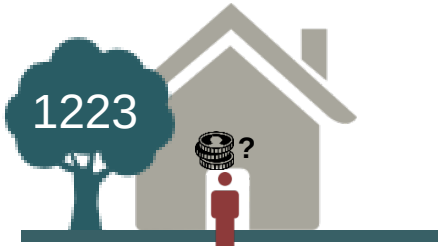
h, g	Household type
d	Dwelling type
z	Zone
B	Bid function
H	Number of househ.
V	Utility

The need for imputing data
(when modeling residential location)

Bid function (linear in parameters)

$$B_{hdz} = f(x_{household}, x_{dwelling}, x_{zones}, x_{accessibility})$$

Zensus 2011



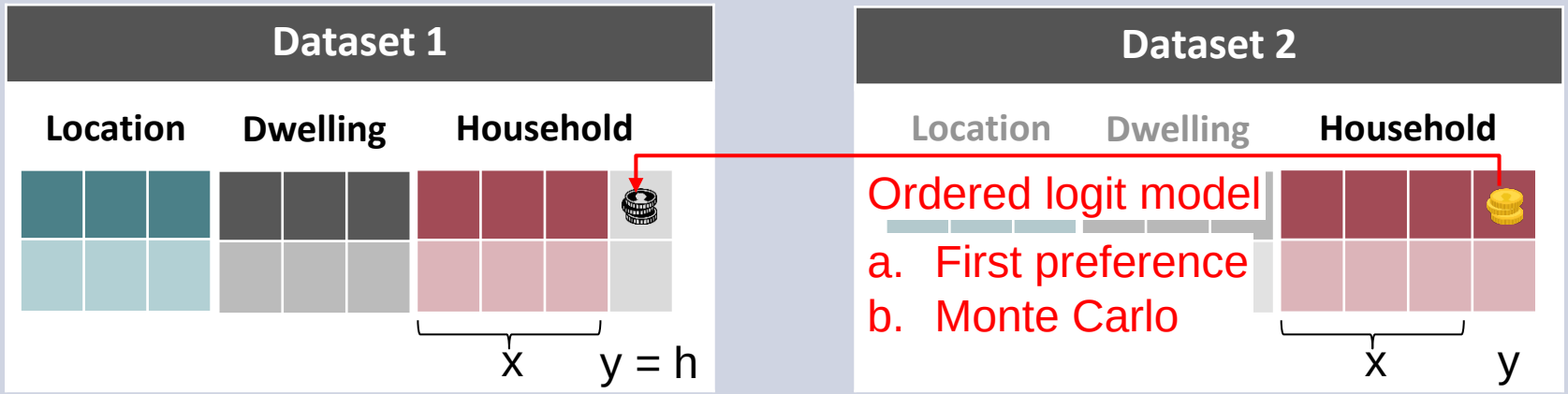
1.800.000 households
Location attributes
No household income

Mikrozensus 2010

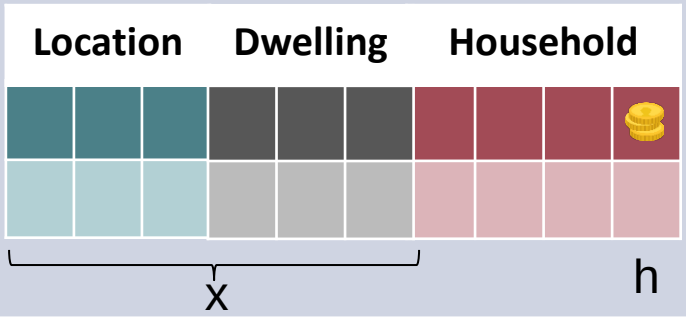
15.000 households
Household income class

Combined imputation and estimation
General methodology

Step 1: Imputation model



Step 2: Estimation of bid function parameters



Multinomial logit model

$$\longrightarrow \quad B_{hdz} = f(x_{household}, x_{dwelling}, x_{zones}, x_{access})$$

a. First preference
b. Monte Carlo

Case study Berlin Step1: Income imputation model

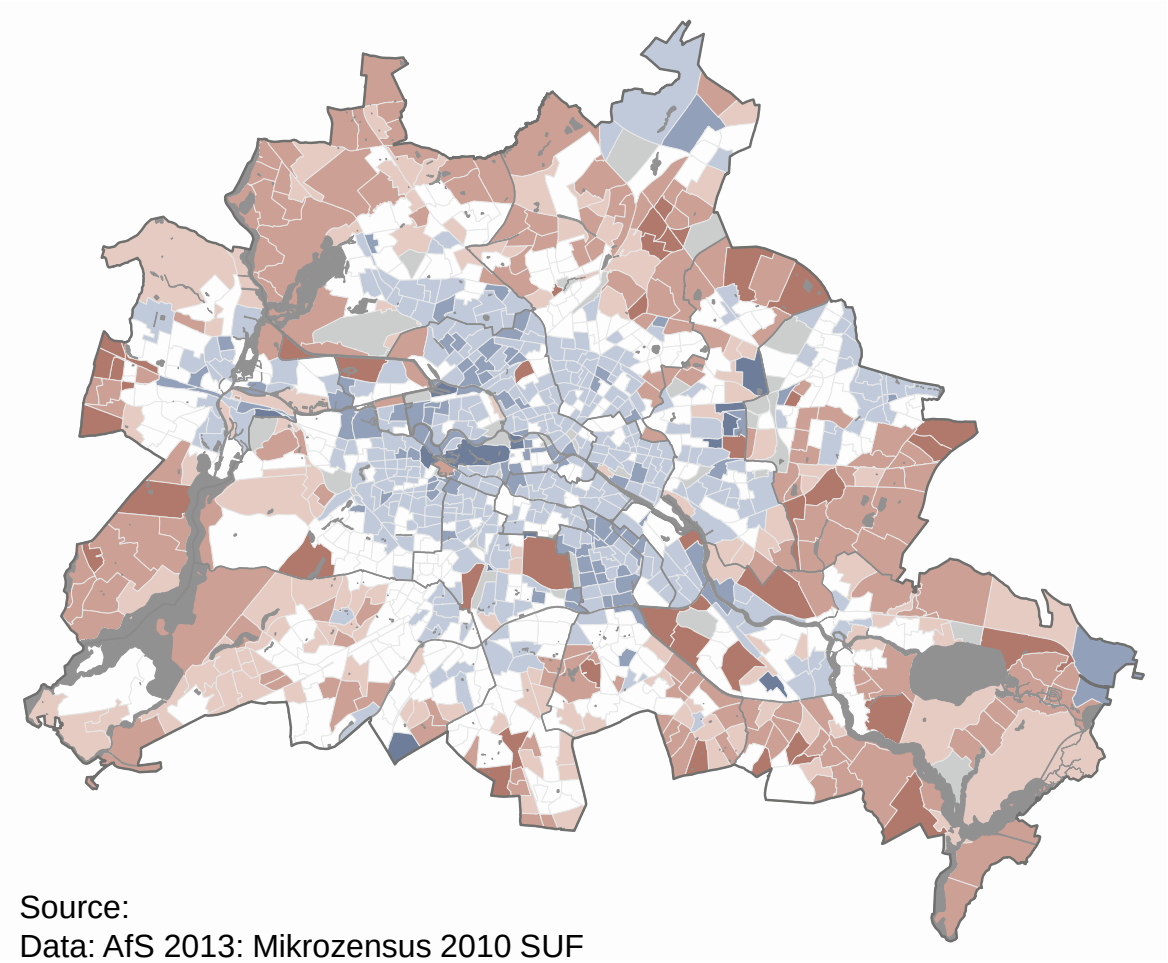
Variable		β
Thresholds for four income classes below 900 € 900 to below 1,500 € 1,500 to below 2,600 € 2,600 € and above	1	-0.631*
	2	-2.704*
	3	-5.089*
Household representative is employed in industry group …	1 (e.g. administration, services)	0.651*
	2 (e.g. retail, construction, health)	1.190*
	3 (e.g. manufacturing, education, arts)	1.960*
	4 (e.g. ITC, science, public services)	2.366*
	5 (e.g. finance and insurance, electricity)	2.998*
Household is single-person household		-0.723*
Household representative has migration background		-0.837*
Number of household members:	below the age of 18	0.438*
	18 until 30	0.867*
	31 until 50	1.820*
	51 until 64	1.835*
	65 and above	2.363*

Log-likelihood at convergence *-15,473*

Log-likelihood at zero *-20,465*

* p < 0.001

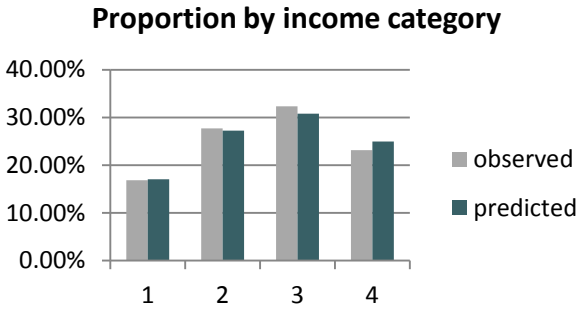
Case study Berlin Step1: Income imputation model



Proportion of households with at least 2,600 net household income € by zone

Standardised values

- less than −1.96
- −1.96 to less than −0.98
- −0.98 to less than −0.49
- −0.49 to less than 0.49
- 0.49 to less than 0.98
- 0.98 to less than 1.96
- 1.96 and higher
- Missing



Source:
Data: AfS 2013: Mikrozensus 2010 SUF
Geodata: SenSW Berlin 2016

Case study Berlin Step 2: Bid functions

Attribute	First preference imputation				Monte Carlo imputation			
Income category	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Constant	0	-0.883*	-3.21*	-5.7*	0	0.0797*	-0.74*	-2.57*
Dwelling size * is multi-family house	0	0.0164*	0.044*	0.0602*	0	0.00565*	0.0151*	0.0257*
Is single-family house	0	1.89*	4.35*	6.25*	0	0.611*	1.59*	2.9*
Proportion of high income hh.	0	3.7*	4.26*	6.03*	0	1*	1.97*	3.1*
Travel time to center	0	0.245*	0.317*	0.388*	0	0.0579*	0.115*	0.143*

Log-Likelihood at 0	-24,833	-24,833	* p < 0.001
Log-Likelihood	-20,671	-22,966	

Conclusions

- Low cost way to deal with missing data (as compared to new survey)
- Ability to determine significant location characteristics
- Probabilistic model considers error term

- Statistical explanation power limited due to increasing error with each step

- Improvement of the approach by:
  - Simultaneous estimation
  - More computatively effective methods (than Monte Carlo)
  - Increase sample size → determine more significant attributes

- Need for research data lobbying

References

Bibliography

ALONSO, W. 1964. *Location and land use. Toward a general theory of land rent,* Cambridge, United Kingdom, Harvard University Press.

MARTÍNEZ, F. 1992. The bid-choice land-use model: An integrated economic framework. *Environment and Planning A,* 24**,** 871-885.

MCCULLAGH, P. 1980. Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, 109-142.

MCFADDEN, D. 1978. Modelling the choice of residential location. *In:* KARLQVIST, A., LUNDQVIST, L., SNICKARS, F. & WEIBULL, J. (eds.) *Spatial interaction theory and planning models.* Amsterdam: North Holland.

Data

FDZ 2010. Mikrozensus. 2010. Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, dataset received on 30th of April 2016.

FDZ 2011. Zensusgesamtdatenbestand Berlin. 2011. Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, dataset received on 29th of July 2016.

Thank you

Dipl.-Geogr. Benjamin Heldt
Research Associate

German Aerospace Center (DLR)
Institute of Transport Research
Department for Mobility and Urban Development

M: benjamin.heldt@dlr.de
P: +4930670557971