Joep Steegmans

Online housing search: An economics cloud use case

June 1, 2023 6th Barcelona TechnoWeek





Universiteit Leider Tax Law and Economics

The Funda project (and online housing search)

- Funda, largest online housing platform in the Netherlands
- Web analytics (website interactions)
- User-generated data (10 TB)
- Micro-behaviour of house buyers (the ultimate goal)
- Municipal hit flows (origin-destination search flows)

tform in the Netherlands ons)

s (the ultimate goal) ation search flows)

Online housing search (current sub-projects)

- Research discipline: Economics
- P1: Online housing search and residential mobility (with K. Bakker)
 - Relating online behaviour to real outcomes (cross-section, 2018m1-2018m6)
- P2: Matching on the housing market (with J. Rouwendal)
 - Estimating (aggr.) matching functions (panel data, 2016m9-2018m6) —
- P3: Micro-behaviour of house buyers (with J. de Bruin)
 - Data prep. on individual search histories (panel data+, 2015m12-2018m6)

The Google Cloud















Research question

- To what extent does online search behaviour of housing platform users forebode real residential mobility flows?
- Can online search flows be used to improve the modelling of real residential mobility flows?
- And if so, do the effects of search behaviour increase with the seriousness of the platform users?

Theory: Housing search and residential mobility

Housing platforms, online listing portals:

- Spatial dimensions of search; curiosity search (Rae 2015) -
- Online listings (Loberto, Luciani & Pangallo 2020)
- Gravity model of housing search flows; hit flows (Steegmans & De Bruin 2021a)
- Residential mobility:
 - Job mobility, residential mobility, commuting (van Ommeren et al. 2000)
 - Residential location search; household average work distance (Rashidi et al. 2012)
 - Build environment, socioeconomic envir., points of interest, accessibility (Schirmer et al. 2014)

Data sets

- Funda search flows (hit flows)
 - Mendeley Data repository; Steegmans and De Bruin 2021c) -
- Relocations
 - CBS StatLine 2011-2021; Persons moving between municipalities
 - CBS relocation dashboard 2017-2020; Residential relocations
- Commuting

- CBS StatLine 2014-2020; Jobs of employees by employment and residence regions



Search flow data (2018H1)

1.1 billion hits

- 30 million users
- 124 million sessions
- 280 thousand properties
- 388 municipalities
- 148,216 flows (382*388)

(January-June 2018)





Hit flows from the Hague

Hit flows to the Hague





Search flow data (2018H1)

■ 1.1 billion hits

- 30 million users
- 124 million sessions
- 280 thousand properties
- 388 municipalities
- 148,216 flows (382*388)

(January-June 2018)

Table 1

Hits $(\times 10^6)$ Events $(\times 10^6)$ Time on site $(\times 10^6 \text{ h})$ Non-zero flows Total flows

Hits (percent) Events (percent) Time on site (percent) Non-zero flows (percent

Notes: Account, users that have registered and created a user account; Mortgage, users that have done the online mortgage calculation; Telephone, users that have clicked the button for the real estate agent's telephone number; Email service, users that have signed up for an email service of new listings within their preferences; Message, users that have contacted the real estate agent through an online form; Viewing, users that scheduled a viewing with the online tool; Buyer, users that registered themselves as the buyer of a property. See also: Steegmans and de Bruin [3, p. 11].

	Full sample	Account	Mortgage	Telephone	Email service	Message	Viewing	Buyer
	1096.6	244.3	213.8	230.4	118.3	88.2	92.2	4.4
	632.8	150.0	125.7	143.1	70.4	54.3	56.7	2.8
	12.6	2.8	2.6	2.9	1.3	1.0	1.1	0.1
	147,903	143,997	142,317	142,072	135,050	126,757	122,913	43,704
	148,216	148,216	148,216	148,216	148,216	148,216	148,216	148,216
t)	100	22.3	19.5	21.0	10.8	8.0	8.4	0.4
	100	23.7	19.9	22.6	11.1	8.6	9.0	0.4
	100	22.1	20.2	22.8	10.6	8.0	8.6	0.4
	99.8	97.2	96.0	95.9	91.1	85.5	82.9	29.5

Flow descriptives of the full sample and split by type of platform user.







Relocations (2019, 2018)

CBS StatLine

2019: 706,890 relocations between municipalities; 117,306 flows. 2018: 738,620 relocations between municipalities; 141,752 flows

CBS Dashboard

2019: 85 municipalities; 498,760 relocations from adults; 27,120 flows; 210,645 relocations of adults to owner-occupied houses).



Relocations (2019)

Relocations (2018)





Commuting (2017-12)

CBS StatLine

December 2017: 7,508,804 commuters; 150,544 flows. December 2018: 7,624,406 commuters; 150,544 flows.

Commuting flows (2017/12)



Empirical model (and methodology)

 $\ln(moves_{ii}) = \alpha_0 + \beta_1 \ln(hitflow_{ii}) + \beta_2 \ln(hitflow_{ii})$ $+\beta_5 contiguity_{ii} + \beta_6 province_{ii} +$

*destination*_{*i*} are the origin and destination fixed effects, ε_{it} is an idiosyncratic error term.

$$n(commute_{ij}) + \beta_{3} \ln(distance_{ij}) + \beta_{4} internal_{ij}$$

$$\sum_{i=1}^{382} \gamma_{i} origin_{i} + \sum_{j=1}^{388} \delta_{j} destination_{j} + \varepsilon_{ij} \quad (1)$$

where moves_{ii} are residential moves from municipality i to j, hitflow_{ii} is the flow of hits from i to j, commute_{ii} is the number of people commuting between i and j, *distance_{ii}* is the Euclidean distance between the centroids of i and j, internal_{ii} indicates whether search flows are within the municipality itself, contiguity_{ii} indicates whether i and j are neighbours, province_{ii} indicates whether i and j share the same province, origin_i and





Estimation results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Benchmark	Âİ	Account	Mortgage	Telephone	Email	Message	Viewing	Buyer
ln(hitflow)		0.267^{***}	0.153^{***}	0.119^{***}	0.118^{***}	0.0959^{***}	0.0820***	0.0752^{***}	0.109^{***}
		(0.00396)	(0.00292)	(0.00265)	(0.00262)	(0.00230)	(0.00205)	(0.00192)	(0.00157)
ln(commute)	1.102^{***}	1.099^{***}	1.099^{***}	1.108^{***}	1.107^{***}	1.107^{***}	1.109^{***}	1.115^{***}	1.027^{***}
	(0.0107)	(0.0105)	(0.0106)	(0.0106)	(0.0106)	(0.0106)	(0.0106)	(0.0106)	(0.0105)
$\ln(distance)$	-0.926***	-0.696***	-0.782^{***}	-0.812^{***}	-0.809***	-0.830***	-0.840***	-0.839***	-0.840***
	(0.00488)	(0.00588)	(0.00555)	(0.00547)	(0.00549)	(0.00537)	(0.00531)	(0.00534)	(0.00494)
within mun	0	0	0	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
neighbor mun	0.612^{***}	0.432^{***}	0.527^{***}	0.531^{***}	0.549^{***}	0.559^{***}	0.571^{***}	0.574^{***}	0.563^{***}
	(0.0161)	(0.0160)	(0.0160)	(0.0161)	(0.0160)	(0.0160)	(0.0160)	(0.0160)	(0.0158)
within prov	0.371^{***}	0.297***	0.323***	0.335***	0.334^{***}	0.340***	0.341^{***}	0.345^{***}	0.339***
	(0.00777)	(0.00769)	(0.00773)	(0.00774)	(0.00774)	(0.00774)	(0.00775)	(0.00774)	(0.00762)
constant	7.253^{***}	4.371***	5.785***	6.108***	6.126***	6.306***	6.430***	6.465^{***}	6.493***
	(0.214)	(0.214)	(0.213)	(0.213)	(0.213)	(0.213)	(0.213)	(0.213)	(0.210)
FE origin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE destination	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.872	0.877	0.875	0.874	0.874	0.874	0.874	0.873	0.877
Adj. R-squared	0.871	0.876	0.874	0.873	0.873	0.873	0.873	0.873	0.876
Observations	115,938	$115,\!938$	$115,\!938$	115,938	115,938	115,938	115,938	115,938	115,938

Standard errors in parentheses

* p < 0.05,** p < 0.01,*** p < 0.001

Table 1: Log regressions weighted by user mass (CBS StatLine moves)

Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Benchmark	All	Account	Mortgage	Telephone	Email	Message	Viewing	Buyer
ln(hitflow)		0.462^{***}	0.253^{***}	0.224^{***}	0.191^{***}	0.157^{***}	0.132^{***}	0.122^{***}	0.111***
		(0.0106)	(0.00796)	(0.00739)	(0.00724)	(0.00629)	(0.00562)	(0.00521)	(0.00357)
ln(commute)	0.896^{***}	0.858^{***}	0.878^{***}	0.882^{***}	0.890^{***}	0.898^{***}	0.896^{***}	0.904^{***}	0.829^{***}
	(0.0268)	(0.0259)	(0.0263)	(0.0264)	(0.0265)	(0.0265)	(0.0266)	(0.0266)	(0.0265)
$\ln(distance)$	-0.994***	-0.623***	-0.773***	-0.798***	-0.822***	-0.848***	-0.867***	-0.864***	-0.895***
	(0.0116)	(0.0141)	(0.0134)	(0.0132)	(0.0132)	(0.0129)	(0.0127)	(0.0128)	(0.0119)
within mun	0.187	-0.612^{***}	-0.164	-0.184	-0.0762	-0.0721	0.0161	0.0195	0.135
	(0.116)	(0.114)	(0.115)	(0.115)	(0.115)	(0.115)	(0.115)	(0.115)	(0.114)
neighbor mun	0.702^{***}	0.341^{***}	0.530^{***}	0.527^{***}	0.574^{***}	0.590^{***}	0.618^{***}	0.625^{***}	0.639^{***}
	(0.0392)	(0.0388)	(0.0388)	(0.0390)	(0.0390)	(0.0390)	(0.0390)	(0.0389)	(0.0386)
within prov	0.273^{***}	0.114^{***}	0.179^{***}	0.189***	0.204^{***}	0.214^{***}	0.223***	0.228***	0.243***
	(0.0196)	(0.0193)	(0.0195)	(0.0195)	(0.0196)	(0.0196)	(0.0196)	(0.0195)	(0.0193)
constant	6.740^{***}	1.427^{***}	4.323^{***}	4.791^{***}	5.005^{***}	5.556^{***}	5.782^{***}	5.832^{***}	6.284^{***}
	(0.184)	(0.221)	(0.201)	(0.197)	(0.199)	(0.192)	(0.191)	(0.190)	(0.183)
FE origin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE destination	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.805	0.818	0.812	0.812	0.810	0.810	0.809	0.809	0.812
Adj. R-squared	0.802	0.815	0.809	0.809	0.807	0.807	0.806	0.806	0.809
Observations	27,120	27,120	$27,\!120$	27,120	$27,\!120$	27,120	$27,\!120$	$27,\!120$	27,120

Standard errors in parentheses

* p < 0.05,** p < 0.01,*** p < 0.001

Table 2: Log regressions weighted by user mass (CBS Dashboard moves)

Robustness (ii)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Benchmark	All	Account	Mortgage	Telephone	Email	Message	Viewing	Buyer
ln(hitflow)		0.428^{***}	0.247^{***}	0.208^{***}	0.186^{***}	0.148^{***}	0.120^{***}	0.113^{***}	0.0927***
		(0.0101)	(0.00756)	(0.00703)	(0.00688)	(0.00598)	(0.00535)	(0.00495)	(0.00341)
$\ln(\text{commute})$	1.210^{***}	1.175^{***}	1.192^{***}	1.197^{***}	1.204^{***}	1.212***	1.210^{***}	1.217^{***}	1.154^{***}
	(0.0255)	(0.0247)	(0.0250)	(0.0251)	(0.0252)	(0.0252)	(0.0253)	(0.0253)	(0.0252)
ln(distance)	-0.688***	-0.344^{***}	-0.472^{***}	-0.505^{***}	-0.520^{***}	-0.550^{***}	-0.572^{***}	-0.567^{***}	-0.604^{***}
	(0.0111)	(0.0134)	(0.0127)	(0.0125)	(0.0126)	(0.0123)	(0.0121)	(0.0122)	(0.0113)
within mun	0.0331	-0.708***	-0.310**	-0.312^{**}	-0.223^{*}	-0.211	-0.122	-0.123	-0.0105
	(0.110)	(0.108)	(0.109)	(0.109)	(0.109)	(0.110)	(0.110)	(0.110)	(0.109)
neighbor mun	0.991^{***}	0.656^{***}	0.822^{***}	0.828***	0.866^{***}	0.885^{***}	0.914^{***}	0.919^{***}	0.938^{***}
	(0.0372)	(0.0369)	(0.0369)	(0.0371)	(0.0370)	(0.0371)	(0.0371)	(0.0370)	(0.0368)
within prov	0.273^{***}	0.126^{***}	0.181^{***}	0.195^{***}	0.205^{***}	0.217^{***}	0.228^{***}	0.231^{***}	0.248^{***}
	(0.0187)	(0.0184)	(0.0185)	(0.0186)	(0.0186)	(0.0186)	(0.0186)	(0.0186)	(0.0184)
constant	4.702^{***}	-0.212	2.331^{***}	2.900^{***}	3.007^{***}	3.591^{***}	3.852^{***}	3.868^{***}	4.363^{***}
	(0.175)	(0.210)	(0.191)	(0.187)	(0.189)	(0.183)	(0.182)	(0.181)	(0.175)
FE origin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE dest	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.746	0.762	0.756	0.754	0.753	0.752	0.751	0.751	0.753
Adj. R-squared	0.742	0.758	0.752	0.750	0.749	0.748	0.747	0.747	0.749
Observations	27,120	27,120	27,120	27,120	27,120	27,120	27,120	27,120	$27,\!120$

Standard errors in parentheses

* p < 0.05,** p < 0.01,*** p < 0.001

Table 3: Log regressions weighted by user mass (CBS Dashboard moves to owner-occupied houses)

Findings

- The model with online search performs better than the benchmark (adj. R-sq 0.876 vs. 0.871, hit flow highly significant).
- If the online search flow increases with 1%, relocations increase with 0.3-0.1 percent, ceteris paribus.
- Findings are robust to alternative variable definitions, samples.
- Findings are robust to estimation methodology (LSDV; weights; ln[x], ln[x+c]; PPML).

Conclusion

- Online platform search behaviour affects (predicts) real residential mobility flows.
- The effect does not increase with user seriousness (and smaller underlying observation numbers).
 - Measurement error and attenuation bias?
- The findings are independent of methodological choices.

Micro-behaviour of individual house buyers



Micro-behaviour of individual house buyers



2018-07

Micro-behaviour of individual house buyers



Thank you for your attention!

Joep Steegmans j.w.a.m.steegmans@law.leidenuniv.nl





Universiteit Leiden Tax Law and Economics

