Adding a small piece of information to the housing market price model – how the story changes?

### Ari Laitala, Lotta Laitala



http://midwestproducer.com/app/blog/wp-ontent/uploads/2013/05/20130524\_115639.jpg

#### About the presenter

- Not a practising valuer
- Project manager in Sykli environmental school of Finland
  - E.g "Interactive videos for the greener future" (Nordplus) just starting
- Editor in chief in Land Use Magazine Finland
- Some teaching activities anyway, mainly about quantitative methods
- Interests
  - Value of the energy efficiency to the buildings including
    - Renewable energy solutions
    - Elasticity of demand (electricity, district heating)
    - Energy surplus effect
    - AND decarbonization of the buildings, life cycle approach
- Other interest
  - Artificial intelligence, virtual reality

#### Data and research question...

- Housing market transaction data
  - <u>https://asuntojen.hintatiedot.fi/haku/?l=2&submit=In</u>
     <u>+English</u>
- Consisting 3275 housing market transactions in Helsinki
  - Only apartments and row houses are included
  - Covering appr. the period 11/2020-12/2019
  - Sample, collected by real estate brokerages (rough estimation, 75 % of all brokerage based transactions)
  - Private transaction excluded (rough estimation, 25 % are private)

#### ... Data and research question

- As a new variable (2020 =>) knowledge of the plot ownership is included
  - Typically, all the block of flats and row houses (buildings) are owned by the housing company
  - Plots are owned by the city or the housing company itself
- RQ: How the quality of the price model changes, if we include the information of the plot ownership into the model

#### Asuntojen.hintatiedot.fi

Instructions | Information about the service | Frequently asked questions | Feedback | Print | Suomeksi | På Svenska |

					1						Next pag	ge
Home sales Helsinki												
District ৰ	Apartment	Type of building ◄	m² ∢	Debt-free price ◄	€/m <sup>2</sup> ∢	Construction year ∢	Floor ◀	Elevator ◄	Condition ◀	Plot ∢	Energy class ◄	
Studio apartn	nent											
Kallio	1h+kk+kph	kt	21,00	191000	9095	1937	3/6	yes	average	own	E <sub>2007</sub>	
Kontula	1h, k	kt	25,50	108000	4235	1965	1/7	yes	average	rent	G <sub>2013</sub>	
Ullanlinna	1h, avok, kph, vh	kt	27,00	285000	10556	1929	1/6	yes	good	own	F <sub>2013</sub>	
Hakaniemi	1h, k, kph/wc,	kt	64,00	395000	6172	1938	3/7	yes	poor	own	E <sub>2013</sub>	
Alppila	1h+k+alk	kt	35,00	246000	7029	1931	6/7	yes	average	own	F <sub>2013</sub>	
Kruununhaka	1h, avok, kph	kt	29,00	283000	9759	1908	5/6	yes	poor	own	F <sub>2013</sub>	
Taka-töölö	1h, k, kph	kt	37,00	293000	7919	1927	6/7	no	average	own	D	
Sörnäinen	1h+k	kt	40,00	280000	7000	1928	3/6	yes	average	own	E <sub>2013</sub>	
Taka-töölö	1h+kk+kph	kt	26,50	202000	7623	1940	3/6	yes	poor	own	G <sub>2013</sub>	
Töölö	1h. kk. koh	kt	31.00	238000	7677	1933	3/5	ves	averade	own	F2012	•
					1						Next pag	ge

Source: Central Federation of Finnish Real Estat

## Let's assume that house prices differences are affected by (in a short period)

- Type of an object
  - D\_Flat, D\_RowHouse
- Number of rooms
  - D\_1r, D\_2r, D\_3r, D\_4r, D\_5r, D\_6r+
- Size (real)
  - SQM
- Condition
  - D\_CondPOOR, D\_CondAVERAGE, D\_CondGOOD
- Floor
  - D\_PossibleView (Floors >6)
- Age (real)
  - Age as a continuous variable
  - And we also perform a set of dummy variables (decade dummies) Age\_categ
    - D\_-1899, D\_1900-1909,...,D\_2010-2019, D\_2020-

## Let's assume that house prices differences are affected by (in a short period)

- Location (ZIP)
  - We have only ZIP-codes!
  - Are <u>not</u> coded into location dummies
- Distances
  - From every (ZIP) location there is calculated distance to Helsinki Business District (kilometers) (Distance00100)
  - Another data was exploited here (Statics Finland Paavo data, which includes coordinates of center points of all ZIP-code areas in Finland (Pythagoras was used, so distances are just direct distances)
- Elevator
  - D\_Elevator
- Apartment sauna
  - D\_Sauna
- Energy performance certificates
  - D\_EPC\_A, D\_EPC\_B, D\_EPC\_C, D\_EPC\_D, D\_EPC\_E, D\_EPC\_F, D\_EPC\_G

#### Some descriptive statistics of the size: Own plots vs. rented plots

Size (m2)						Size (m2)	)				
RENTED	Min	Max	Median	Average	Std.dev. (S)	OWN	Min	Max	Median	Average	Std.de v. (S)
D_1r	21,0	96,5	31,0	32,2	8,3	D_1r	14,0	75,5	30,0	31,7	9,1
D_2r	31,5	119,0	51,0	51,0	8,3	D_2r	29,0	127,0	51,0	52,5	11,9
D_3r	51,0	150,0	73,5	73,3	9,9	D_3r	47,0	180,0	73,0	76,7	16,9
D_4r	68,0	192,5	93,0	96,8	19,0	D_4r	62,0	218,0	99,8	106,5	26,1
D_5r	85,5	203,0	112,8	120,5	30,8	D_5r	88,0	282,0	128,5	141,1	39,3
D_6+r	119,0	242,0	160,0	173,7	62,6	D_6+r	104,0	280,0	186,0	189,8	44,6
All	21,0	242,0	63,0	66,2	25,4	All	14,0	282,0	59,0	66,6	35,6

#### Some descriptive statistics of the prices: Own plots vs. rented plots

Price (i possibl to hous compa	ncluding e debts sing ny )					Price (i possibl housin compa	ncluding e debts to g ny )				
RENTE					Std.dev.						Std.dev.
D	Min	Max	Median	Average	(S)	OWN	Min	Max	Median	Average	(S)
D_1r	90 000	455 000	179 375	182 187	51 797	D_1r	92 000	694 000	224 000	238 911	83 204
D_2r	95 000	414 000	200 750	209 903	58 799	D_2r	85 000	1 039 000	290 500	312 722	134 534
D 3r	100 000	660 000	245 000	256 328	83 869	D 3r	108 000	2 249 000	384 000	418 164	206 933
_ D 4r	118 000	852 000	303 000	325 021	125 458	D 4r	115 000	2 245 720	438 500	513 661	289 350
	400 640	050.000	202 500	202.042	400 705		100.000			COE 740	
D_5r	139 610	850 000	392 500	393 043	133 /05	D_5r	180 000	2 350 000	577500	685 /13	398 212
D_6+r	450 000	654 500	650 000	584 833	116 791	D_6+r	330 000	2 000 000	696 500	801 827	390 995
AII	90 000	852 000	227 000	247 319	99 069	All	85 000	2 350 000	313 000	371 483	227 373

#### Some descriptive statistics of the prices: Own plots vs. rented plots

Distance						Dist	ance					
RENTED	Min	Max	Median	Average	Std.dev. (S)	OW	'N N	vlin	Max	Median	Average	Std.dev. (S)
D_1r	1,0	11,3	5,2	5,9	2,5	D_	1r	0,0	13,7	2,5	3,4	2,8
D_2r	0,0	13,4	7,0	7,3	3,0	D_	2r	0,0	13,7	3,4	4,9	3,9
D_3r	0,0	13,7	9,1	8,5	3,0	D_	3r	0,0	13,7	5,2	5,9	4,0
D_4r	0,0	13,4	9,1	8,5	3,1	D_	4r	0,0	17,8	6,9	6,7	4,0
D_5r	1,0	13,1	8,8	8,2	3,2	D_	5r	0,0	17,8	6,2	6,5	4,4
D_6+r	5,2	8,6	5,2	6,3	2,0	D_	6+r	0,0	17,8	6,8	7,2	4,1
All	0,0	13,7	8,0	7,8	3,1	All		0,0	17,8	3,5	5,1	3,9

#### Some visualizations of the sample: EPCs



### Correlations between the transaction prices and prices per square meter (yellow rented)



### And before the calculation we do one technical variable transformation

- It is a property of the MRA modelling that results are better when there are more <u>linear</u> relationships between the dependent and the independent variables
- Standard solution => we take a logarithm, e.g In(Total\_Price)
- For example ln(267000) = 12,495
  - If the model gives the estimation 12,500 we calculate  $e^{12,5} \approx 268337$

#### Example (after transforming the dependent variable) Total price vs. SQM



#### About the method

- Multiple linear regression analysis (MRA)
  - JASP version 0.8.0.1
  - Estimation method: Ordinary Least Squares (OLS)
    - Hereby e.g. average of the residuals is zero
      - (difference between the observed value and the estimated value)
  - <u>Stepwise algorithm</u>
    - See the JASP handbook , p. 77
    - <u>https://jasp-stats.org/wp-</u> <u>content/uploads/2020/11/Statistical-Analysis-in-JASP-</u> <u>A-Students-Guide-v14-Nov2020.pdf</u>

Model withou	ıt		Mc	odel with	
Plot_rented			Plo	ot_rented	
R = 0,918			R =	0,928	
(Adj. R2 = 84,	,1 %)		(Ad	lj. R2 = 86	,0 %)
(Adj. R2 = 84, intercept SQM Distance00100 D_CondGOOD D_1970-1979 D_2010-2019 D_2010-2019 D_2000-2009 D_1r D_1960-1969 D_2r D_1900-1909 D_2r D_1900-1909 D_EPC_C D_Elevator D_Sauna D_CondPOOR Age D_2020- D_6+r	,1 %) 12.223 0.009 -0.072 0.112 -0.187 0.321 0.243 -0.222 -0.105 -0.079 0.091 0.064 0.042 0.062 -0.102 0.003 0.205 -0.189	2 Also the order (importance) of the variables has changed!	(Ad 4 interc SQM Distai D_Co D_19 Plot_r D_19 D_17 D_20 D_20 D_20 D_21 D_20 D_20 D_20 D_21 D_20 D_21 D_20 D_21 D_20 D_21 D_20 D_21 D_20 D_21 D_20 D_21 D_20 D_20 D_21 D_20 D_20 D_21 D_20 D_20 D_21 D_20	ij. R2 = 86 cept nce00100 ondGOOD 70-1979 rented 60-1969 10-2019 00-2009 00-1909 20- 000-1909 20- auna 30-1939 evator ondPOOR r 2C. C	,0 %) 12.302 0.009 -0.068 0.108 -0.198 -0.159 -0.115 -0.246 0.300 0.246 -0.087 0.068 0.257 0.003 0.049 -0.111 0.039 -0.094 -0.202 0.054
D_EPC_G D_1930-1939	-0.034 -0.076		D_19	0_0 20-1929 90-1999	-0.081
D_1920-1929	-0.066		D_EP	PC_G	-0.029
D_5r	-0.044		D_Po	ssibleView	0.041
D_PossibleView	0.036	BVC 2019 Laitala	D_5r		-0.046

#### Model with Plot\_rented

24	intercept	12.302	0.026		471.860	<.001
	SQM	0.009	0.000	0.562	46.014	< .001
	Distance00100	-0.068	0.001	-0.533	-56.610	< .001
	D_CondGOOD	0.108	0.007	0.106	14.867	< .001
	D_1970-1979	-0.198	0.012	-0.126	-16.419	< .001
	Plot_rented	-0.159	0.008	-0.152	-20.909	< .001
	D_1960-1969	-0.115	0.010	-0.088	-11.211	< .001
	D_1r	-0.246	0.014	-0.185	-17.685	< .001
	D_2010-2019	0.300	0.020	0.181	15.066	< .001
	D_2000-2009	0.246	0.018	0.131	14.069	< .001
	D_2r	-0.087	0.009	-0.085	-9.240	< .001
	D_1900-1909	0.068	0.027	0.019	2.515	0.012
	D_2020-	0.257	0.024	0.095	10.570	< .001
	Age	0.003	0.000	0.173	10.718	< .001
	D_Sauna	0.049	0.009	0.041	5.227	< .001
	D_1930-1939	-0.111	0.015	-0.058	-7.363	< .001
	D_Elevator	0.039	0.008	0.039	5.037	< .001
	D_CondPOOR	-0.094	0.019	-0.034	-4.981	< .001
	D_6+r	-0.202	0.038	-0.041	-5.366	< .001
	D_EPC_C	0.054	0.013	0.032	4.193	< .001
	D_1920-1929	-0.081	0.016	-0.041	-5.031	< .001
	D_1990-1999	0.068	0.019	0.028	3.641	< .001
	D_EPC_G	-0.029	0.011	-0.019	-2.763	0.006
	D_PossibleView	0.041	0.016	0.017	2.547	0.011
	D_5r	-0.046	0.020	-0.018	-2.316	0.021

#### Model without Plot\_rented

Residuals vs. Dependent 🔻



Residuals vs. Predicted



BVC 2019 Laitala

### Model with Plot\_rented

Residuals vs. Dependent 🔻







### Model without Plot\_rented

Standardized Residuals Histogram V



#### Q-Q Plot Standardized Residuals



### Model with Plot\_rented



Q-Q Plot Standardized Residuals



#### We try to answer...

- Plot ownership <u>seems</u> to be highly effecting and very significant variable to take account in the price models <u>and hereby in valuation</u>
  - Correlations between the observed price and price predictions increase
  - (No visually observable changes in the residuals)
- Still uncertainties in data (which may explain something)
  - No exact time stamp
  - No exact location
  - Private transactions excluded
  - No knowledge about the housing companies economic situations, they
    may vary a lot
- Results are in the line with the emerging knowledge, presented by Risto Peltola in the BVC 2020

#### We try to answer...

- Further (qualitative) considerations are needed. For example, there are basically two kind of rent agreements
  - Very old and very cheap
  - New and somewhat expensive ( $\approx$ 5 % yield)
- Benefits of the own plot => you avoid the rent but you have to pay real estate tax (≈ 1 %)
- But it seems that this is worth of further research and to take into account in the valuation practice. How?

#### SUOMEN YMPÄRISTÖOPISTO SYKLI



# Thank you!- questions & comments?

#### ari.laitala@sykli.fi

