



HOUSING RENTAL SUBMARKETS IN HEDONIC REGRESSION: ECONOMETRIC ARGUMENTS AND PRACTICAL APPLICATION

Marko Kryvobokov and Sébastien Pradella
Centre for Sustainable Housing Studies of Wallonia, Belgium

ERES 2019, ESSEC Business School, Cergy-Pontoise, 4 July 2019

Introduction

Geographical housing submarkets

- Geographically and socially heterogeneous areas
- Economically meaningful and disaggregated spatial units (*Keskin & Watkins 2017*)
- Each submarket has its own supply and demand functions (*Rothenberg et al. 1991, Goodman & Thibodeau 1998, Watkins 2001*)
- Within a submarket, identical properties are closely substitutable (*Pryce 2013*)
- The value of a hypothetical standardised housing unit, (i.e. dwelling with the same attributes), changes if it is located in different submarkets

Geographical housing submarkets

in terms of hedonic regression

(Schnare & Struyk 1976, Watkins, 2001)

- Each submarket should be analysed with its own price equation
- These submodels should provide better results than the overall model

Geographical housing submarkets

in terms of hedonic regression

However

- A clear conceptual basis for submarkets still needs to be developed
- Example :

A difference between the concepts of *submarkets* and *neighbourhoods* = smaller areas within a market segment, where market influences are relatively constant (e.g. *Borst 2007*)

The *neighbourhoods* are modelled as dummy variables

Nevertheless, the *submarkets* are often modelled as dummies in an overall model

Two conceptual problems:

- 1) How to *delineate* geographical housing submarkets?
- 2) How to *use the delineated submarkets* in a hedonic price model?

The first problematic is known in the literature since 1950s and 1960s.

The second problematic is less discussed in the literature

- Should the submarkets always be modelled with separate hedonic equations?
- Is the set of such submodels always better than the overall model? According to which criteria?

The purpose of the study:

To search for *the best way to model* the identified submarkets

under the condition of a relatively small sample size

Applied to:

- A rental housing market
- At regional geographical level (the whole cities are regarded as submarkets)

Practical motivation of the study:

The development of *the official regional « **rent calculator** »* for dwellings based on the market principles

Housing tenancy decree (2018):

To estimate “the reference rent” for any dwelling rented on the market in the region

Homogenous zones should be delimited according to rent levels observed on the market

Literature

1) Delineation of submarkets

Literature does not provide a clear answer

whether **statistical methods**
(*principal components,*
clustering ...)
are better than

expert opinion methods

(e.g. *Des Rosiers 1991,*
Bourassa et al. 1999,
Borst & McCluskey 2008)

(e.g. *Bourassa et al.*
2003, Keskin & Watkins
2017)

2) Use of the identified submarkets

What is better:

- submarket dummies
(e.g. *Fletcher et al. 2000, Bourassa et al. 2003, 2007*)

or

- a set of submarket-specific submodels
(e.g. *Watkins, 2001; Goodman and Thibodeau, 2003, 2007*)?

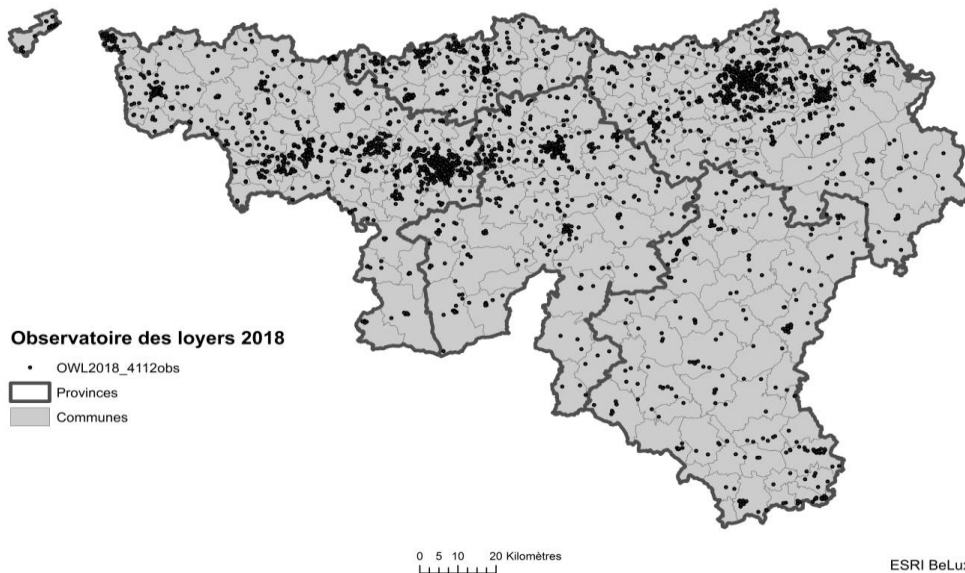
Leishman et al. (2013)

- Search for the best way to model the submarkets once they have been identified
- Three strategies of submarkets modelling are compared:
 - a citywide « benchmark » model
 - a series of submarket-specific submodel
(*a scientific and professional standard*)
 - multi-level models

Data

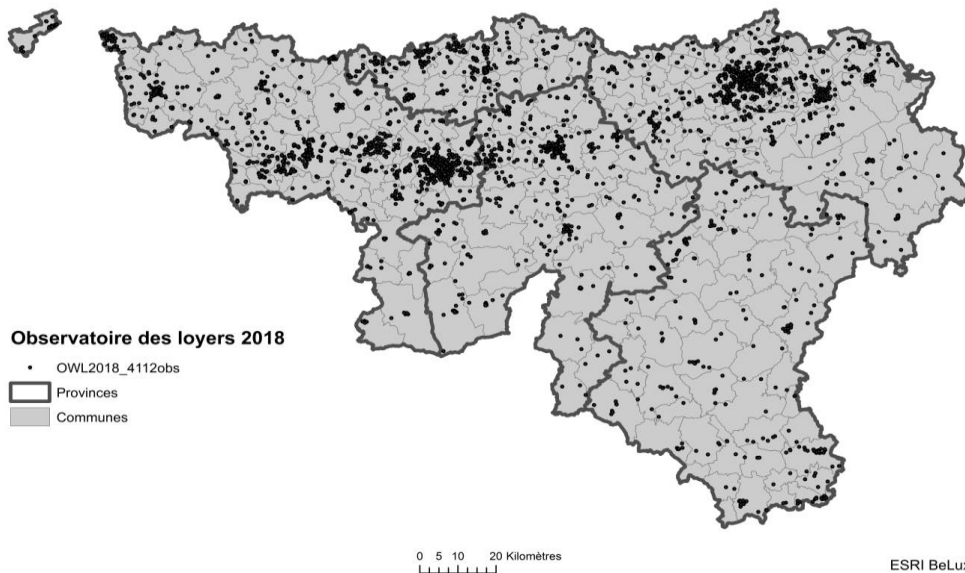
Regional Housing Rent Survey in Wallonia 2018

- Face-to-face (77%) and telephone (23%) interviews
- 4.112 dwellings = households private tenants (sampling rate 1%)
- Representative sample of the regional rental market (*geography, building type and building age*)
- Rent + housing attributes + address + tenancy agreement



Regional Housing Rent Survey in Wallonia 2018

- Face-to-face (77%) and telephone (23%) interviews
- 4.112 dwellings = households private tenants (sampling rate 1%)
- Representative sample of the regional rental market (*geography, building type and building age*)
- Rent + housing attributes + address + tenancy agreement



Alternative data sources:

- Register of tenancy agreements : **very incomplete**
- Real estate agencies : **non representative**

Data

Variable	Mean
Rent without charges, €	616.35
More than 9 years in the dwelling with the same tenancy agreement (dummy)	0.035
Home sharing (dummy)	0.009
Row-house or semi-detached house (dummy)	0.351
Detached house (dummy)	0.181
Apartment (dummy)	0.436
Studio (dummy)	0.032
Construction period before 1919 without renovation since 2008 (dummy)	0.109
Construction period before 1919 with renovation since 2008 (dummy)	0.041
Construction period from 1919 to 1945 (dummy)	0.251
Construction period after 1990 (dummy)	0.215
Number of bedrooms	2.07
Living area, m ²	81.33
Garage is included in rent (dummy)	0.35
Number of bathrooms / shower-rooms	1.06
Number of WC inside	1.18

Variable	Mean
Equipped kitchen (dummy)	0.471
Balcony or terrace (dummy)	0.370
Laundry, attic or other storage space (dummy)	0.683
Individual garden (dummy)	0.513
Swimming pool permanently installed (dummy)	0.006
Room for office use only (dummy)	0.132
Large windows (dummy)	0.135
Genuine parquet flooring (dummy)	0.209
Additional security equipment (dummy)	0.122
Ground floor (for apartments, dummy)	0.106
Energy performance certificate A (dummy)	0.015
Energy performance certificate B (dummy)	0.102
Energy performance certificate C (dummy)	0.135
Energy performance certificate D (dummy)	0.238
Energy performance certificate E (dummy)	0.233
Energy performance certificate F (dummy)	0.133
Energy performance certificate G (dummy)	0.144

Models

Models

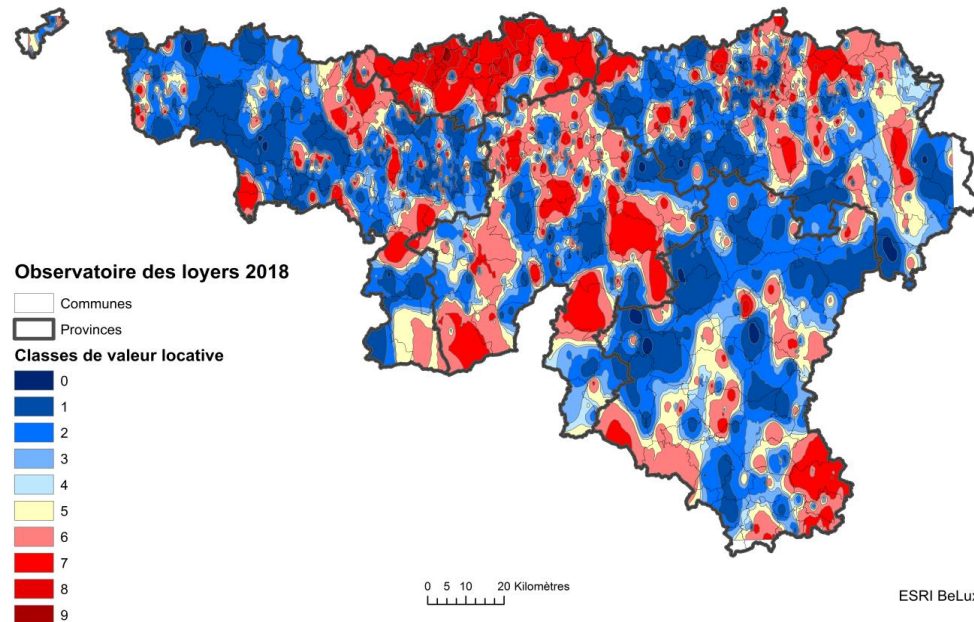
Initial OLS model
(without
geographical
variables)

Variable, parameter	Initial	
	Coef.	Signif.
Constant	5.491	0.000
MoreThan9years	-0.172	0.000
HomeSharing	0.066	0.027
RowOrSemiDetachedHouse	-0.055	0.000
DetachedHouse	-0.032	0.014
Studio	-0.083	0.000
Before1919withoutRenovation	-0.043	0.000
Before1919withRenovation	-0.025	0.075
1919_1945	-0.023	0.001
After1990	0.016	0.041
LnNbBedrooms	0.337	0.000
LnLivingArea	0.093	0.000
Garage	0.074	0.000
LnNbBathrooms	0.404	0.000
LnNbWC	0.223	0.000
EquippedKitchen	0.059	0.000
BalconyOrTerrace	0.018	0.005
LaundryOrAttic	0.038	0.000
IndividualGarden	0.072	0.000
SwimingPool	0.062	0.093
RoomOffice	0.043	0.000
LargeWindows	0.070	0.000
GenuineParquetFlooring	0.056	0.000
SecurityEquipement	0.036	0.000
GroundFloor	-0.032	0.003
EPC_F	-0.036	0.000
EPC_G	-0.061	0.000
N	4,112	
R ² adj.	0.5999	
Max VIF	3.55	
Standard error of the estimate	0.1738	
Mean standard error of the prevision	0.0135	
Moran's I in residuals	0.137 (0.000)	
Predictions within 10%	46.4%	
Predictions within 20%	77.7%	

Delineation of submarkets

The « **location value** » technique :

a ratio of *the observed rent* to *the predicted rent* from the regional model
(without geographical attributes)



The potential methodological problem: omitted variables might not necessarily be only location attributes.

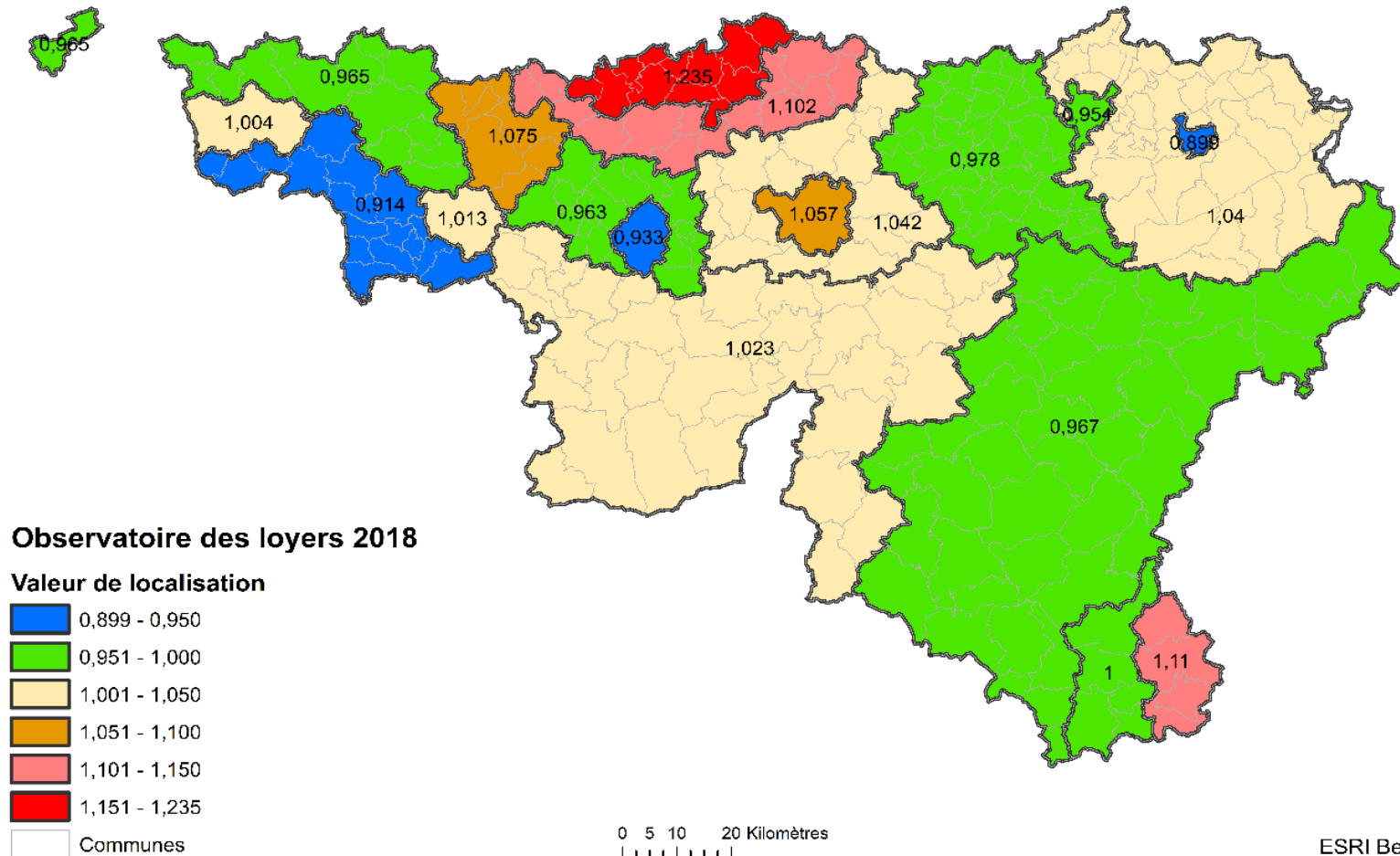
However, the model includes about thirty non-location variables.

The technique is a good proxy for « location value ».

Delineation of submarkets

Grouping 262 municipalities into **19 rental submarkets**

Combination of a clustering with an expert approach



Models

Variable, parameter	Initial		With submarkets' dummies		With « location values »	
	Coef.	Signif.	Coef.	Signif.	Coef.	Signif.
Constant	5.491	0.000	5.430	0.000	5.396	0.000
MoreThan9years	-0.172	0.000	-0.165	0.000	-0.165	0.000
HomeSharing	0.066	0.027	0.090	0.001	0.091	0.001
RowOrSemiDetachedHouse	-0.055	0.000	-0.049	0.000	-0.048	0.000
Internal variables
...
Charleroi_city	-	-	-0.121	0.000	-	-
Liège_city	-	-	-0.093	0.000	-	-
Namur_city	-	-	N/S	-	-	-
Mons_city	-	-	-0.050	0.001	-	-
Tournai_city	-	-	-0.055	0.002	-	-
Verviers_city	-	-	-0.154	0.000	-	-
BrabantWalloonNorth	-	-	0.175	0.000	-	-
BrabantWalloonSouth	-	-	0.063	0.000	-	-
HainautNorth	-	-	0.033	0.067	-	-
HainautNorthWest	-	-	-0.071	0.000	-	-
HainautSouthWest	-	-	-0.132	0.000	-	-
HainautEast	-	-	0.073	0.000	-	-
ProvNamurN_ProvLiègeNW	-	-	N/S	-	-	-
ProvNamurS_HainautS	-	-	N/S	-	-	-
ProvLiègeWest	-	-	-0.058	0.000	-	-
LuxembourgSouthEast	-	-	0.088	0.000	-	-
LuxembourgSouth	-	-	N/S	-	-	-
LuxembourgN_ProvLiègeSE	-	-	-0.078	0.000	-	-
LOCATION VALUE	-	-	-	-	1.034	0.000
N	4,112		4,112		4,112	
R ² adj.	0.5999		0.6665		0.6672	
Max VIF	3.55		3.68		3.55	
Standard error of the estimate	0.1738		0.1586		0.1585	
Mean standard error of the prevision	0.0135		0.0162		0.0126	
Moran's I in residuals	0.137 (0.000)		0.013 (0.000)		0.015 (0.000)	
Predictions within 10%	46.4%		50.9%		51.3%	
Predictions within 20%	77.7%		81.8%		81.8%	

Set of 19 submodels

Reduction in the weighted standard error **16.3%**

More *previsions* within the 10% interval: **56.9%**

More *previsions* within the 20% interval: **86.6%**

Moran's I is decreased to **0.9%** ($p=0.004$)



Formally, **according to econometric indicators**, the set of submodels is better than the overalls models, **but ...**

Set of 19 submodels

What about these 19 submodels?

Submarket	n	R ² adj.	Total number of variables	Number of significant variables (at the 10% level)	Living area significant
Charleroi_city	315	47.9%	25	7	Non
Liège_city	418	66.6%	25	14	Oui
Namur_city	190	70.8%	26	10	Oui
Mons_city	162	45.9%	24	10	Non
Tournai_city	105	65.4%	23	10	Non
Verviers_city	97	60.6%	25	5	Oui
BrabantWalloonNorth	239	71.8%	26	7	Oui
BrabantWalloonSouth	157	64.4%	25	10	Oui
HainautNorth	94	80.8%	24	8	Oui
HainautNorthWest	181	54.6%	24	8	Non
HainautSouthWest	207	53.2%	24	10	Non
HainautEast	323	64.3%	26	12	Oui
ProvNamurN_ProvLiègeNW	200	60.6%	26	8	Oui
ProvNamurS_HainautS	336	61.4%	25	13	Oui
ProvLiègeWest	264	56.5%	26	10	Oui
ProvLiègeCentre	516	71.1%	26	16	Oui
LuxembourgSouthEast	65	73.6%	22	4	Non
LuxembourgSouth	28	67.4%	22	3	Non
LuxembourgN_ProvLiègeSE	215	68.8%	25	11	Non

Set of 19 submodels

The problem of non-significant crucial variables (little discussed in the literature, e.g. *Watkins 2001*)

Submarket	n	R ² adj.	Total number of variables	Number of significant variables (at the 10% level)	Living area significant
Charleroi_city	315	47.9%	25	7	Non
Liège_city	418	66.6%	25	14	Oui
Namur_city	190	70.8%	26	10	Oui
Mons_city	162	45.9%	24	10	Non
Tournai_city	105	65.4%	23	10	Non
Verviers_city	97	60.6%	25	5	Oui
BrabantWalloonNorth	239	71.8%	26	7	Oui
BrabantWalloonSouth	157	64.4%	25	10	Oui
HainautNorth	94	80.8%	24	8	Oui
HainautNorthWest	181	54.6%	24	8	Non
HainautSouthWest	207	53.2%	24	10	Non
HainautEast	323	64.3%	26	12	Oui
ProvNamurN_ProvLiègeNW	200	60.6%	26	8	Oui
ProvNamurS_HainautS	336	61.4%	25	13	Oui
ProvLiègeWest	264	56.5%	26	10	Oui
ProvLiègeCentre	516	71.1%	26	16	Oui
LuxembourgSouthEast	65	73.6%	22	4	Non
LuxembourgSouth	28	67.4%	22	3	Non
LuxembourgN_ProvLiègeSE	215	68.8%	25	11	Non

Set of 19 submodels

Submodel example: Mons

Variable, parameter	Coefficient	Significance
Constant	5.647	0.000
MoreThan9years	N/S	-
RowOrSemiDetachedHouse	N/S	-
DetachedHouse	-0.116	0.086
Studio	-0.203	0.058
Before1919withoutRenovation	N/S	-
Before1919withRenovation	N/S	-
1919_1945	-0.069	0.053
After1990	N/S	-
LnNbBedrooms	N/S	-
LnLivingArea	N/S	-
Garage	0.079	0.032
LnNbBathrooms	0.640	0.025
LnNbWC	0.311	0.097
EquippedKitchen	0.096	0.007
BalconyOrTerrace	N/S	-
LaundryOrAttic	N/S	-
IndividualGarden	0.118	0.002
RoomOffice	N/S	-
LargeWindows	0.082	0.087
GenuineParquetFlooring	0.094	0.016
SecurityEquipement	N/S	-
GroundFloor	N/S	-
EPC_F	N/S	-
EPC_G	N/S	-
n	162	
R ² adj.	0.4592	
Max VIF	3.98	

GWR and multi-level models

Variable, parameter	GWR estimates aggregated to submarkets		Multi-level model	
	Submarket min	Percentage of submarkets with a variable significant in at least 5% cases	Fixed-effects estimate	Random effects variance
Constant	100.0%	100.0%	5.407	0.104
MoreThan9years	4.7%	100.0%	-0.150	0.005
HomeSharing	31.1%	100.0%	N/S	N/S
RowOrSemiDetachedHouse	0.0%	68.4%	-0.048	N/S
DetachedHouse	0.0%	52.9%	-0.028	N/S
Studio	0.0%	33.3%	-0.091	0.009
Before1919withoutRenovation	0.0%	44.4%	-0.033	N/S
Before1919withRenovation	0.0%	46.2%	N/S	N/S
1919_1945	0.0%	57.9%	N/S	N/S
After1990	0.0%	21.1%	N/S	0
LnNbBadrooms	0.0%	94.7%	0.302	N/S
LnLivingArea	0.0%	98.5%	0.134	0.007
Garage	0.6%	84.2%	0.065	0
LnNbBathrooms	0.0%	73.7%	0.297	0
LnNbWC	0.0%	63.2%	0.206	N/S
EquippedKitchen	0.0%	89.5%	0.055	N/S
BalconyOrTerrace	0.0%	42.1%	0.015	N/S
LaundryOrAttic	0.0%	63.2%	0.024	0.001
IndividualGarden	0.0%	78.9%	0.055	N/S
SwimingPool	0.0%	0.0%	N/S	N/S
RoomOffice	0.0%	78.9%	0.039	N/S
LargeWindows	0.0%	82.4%	0.033	N/S
GenuineParquetFlooring	0.0%	73.7%	0.053	N/S
SecurityEquipement	0.0%	52.6%	0.033	0
GroundFloor	0.0%	41.2%	-0.027	N/S
EPC_F	0.0%	63.2%	-0.039	0
EPC_G	0.0%	84.2%	-0.060	N/S
N	4,112		4,112	
Submarkets	19		19	
Moran's I in residuals	0.025 (0.000)		0.010 (0.002)	
Predictions within 10%	52.8%		54.9%	
Predictions within 20%	84.2%		86.4%	

GWR

0.6963
0.017 (0.000)
57.1%
87.1%

Model comparison

Model	Number of insignificant variables	Adj. R ²	Moran's I in residuals and its significance	Predictions within 10%	Predictions within 20%
OLS regional initial	0	0.5999	0.137 (0.000)	46.4%	77.7%
OLS regional with submarkets' dummies	3 internal + 4 submarkets	0.6665	0.013 (0.000)	50.9%	81.8%
OLS regional with "location values"	3 internal	0.6672	0.015 (0.000)	51.3%	81.8%
OLS submodels	10 internal to 23 internal	-	0.009 (0.004)	56.9%	86.6%
GWR regional	1 internal*	0.6963	0.017 (0.000)	57.1%	87.1%
GWR regional, estimates aggregated by submarkets	0 internal to 17 internal*	-	0.025 (0.000)	52.8%	84.2%
Multi-level model	5 internal (fixed effects) + 22 internal** (random effects)	-	0.010 (0.002)	54.9%	86.4%

* – in the GWR, the number of variables with no significant estimates is reported (the cases without a particular variable in a particular submarket are not counted)

** – in the multi-level model, the variables with zero variances are counted as well

Conclusion

- The paper seeks an appropriate econometric method to create the market-based “rent calculator” in the Walloon region in Belgium. According to legislation, homogenous zones should be delimited.
- Nineteen geographical submarkets are delineated with the combination of statistical methods and expert approach. The study seeks the best way to use the created submarkets in hedonic regression under the condition of a relatively small sample size.
- The best **econometric** outcomes are found with:
 - regional GWR (but after aggregation to submarkets, its superiority is lost)
 - a set of OLS submodels

- But crucial structural attributes, such as *living area* or *building age*, are insignificant in many submarkets. It is unacceptable for the users of the “rent calculator”
- The remaining question is to what extent this result is a consequence of:
 - a small size of sub-samples
 - the lack of variability of variables at a local scale
 - the lack of transparency, especially in impoverished areas

- In the overall models with geographical elements, this problem (almost) does not exist
- The practical advantages of the OLS model with “location values” are more important than the relative econometric superiority of the GWR and multi-level alternatives
- The region-wide model with “location values” is preferable to that with submarket binaries:
 - several econometric indicators are slightly better
 - the “location values” force the model to include all submarkets (even minor changes are not lost)
 - there is a smoothing affect in the boundaries between submarkets

Thank you for your attention

Delineation of submarkets

Grouping of 262 municipalities in rental submarkets:

- For the municipalities with at least 10 observations, *an average location value* is calculated
- These municipalities are grouped into ten clusters with the Ward method
- Each of the six major Walloon cities formed a submarket
- The remaining municipalities have been grouped with their neighbouring « value influence centres » and their clusters
- The rules applied: min 30 observations, the max difference at the border 15%, the administrative division and the economic classification of the municipalities are taken into account

Combination of the clustering
with an expert approach

Result : 19 submarkets

