VU Research Portal

Constructing high-resolution housing price indices for the Netherlands

Claassens, J.; Koomen, E.

2017

Link to publication in VU Research Portal

citation for published version (APA)
Claassens, J., & Koomen, E. (2017). *Constructing high-resolution housing price indices for the Netherlands: Spinlab Research Memorandum SL-14*. Vrije Universiteit Amsterdam.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 22. May. 2021

Constructing high-resolution housing price indices for the Netherlands

Jip Claassens
Eric Koomen

December 22, 2017

SPIN LAB SPATIAL INFORMATION LABORATORY

VU VRIJE UNIVERSITEIT AMSTERDAM

COLOPHON

TITLE

Constructing high-resolution housing price indices for the Netherlands

Spinlab Research Memorandum SL-14

AUTHORS

Jip Claassens, Spatial Information Laboratory (SPINlab), Vrije Universiteit Amsterdam.

Eric Koomen, Spatial Information Laboratory (SPINlab), Vrije Universiteit Amsterdam.

CONTACT

Vrije Universiteit Amsterdam

Faculty of Economics and Business Administration

Department of Spatial Economics/ Spatial Information Laboratory (SPINlab)

De Boelelaan 1105

1081 HV Amsterdam

The Netherlands

Phone: +31 20 5986095

Email: e.koomen@vu.nl

Website: https://spinlab.vu.nl/

This report was prepared for PBL Netherlands Environmental Assessment Agency, Den Haag.

Contents

1. Introduction

Following their initial urban transformation study (van Duinen et al., 2016), PBL Netherlands Environmental Assessment Agency is now preparing a spatially-explicit, national study into the possibilities and effects of different urbanisation scenarios, including different types of densification (brownfield redevelopment and demolishing of residences followed by new construction) and urban expansion (large-scale residential development outside existing urban areas). In addition to land development costs, potential revenues of housing development are an important input variable for the new PBL study. These revenues will differ substantially across space and will, for example, typically be higher for inner-city development than for urban expansions of former green-space. The study PBL is currently setting up, will help to evaluate whether such higher revenues will compensate for the higher development costs that can also be expected at inner-city locations.

The main research objective of this subproject is to quantify the spatial variation in revenues that can be expected from different types of residential development. Based on this information the relative profitability of different urban development strategies can be assessed in subsequent simulations in a land-use modelling environment. The current report tries to address which factors matter for the added-value of locations in housing prices, looking at factors such as proximity to jobs, distance to train stations, the presence of amenities, or the amount of nature in the area.

In order to quantify these effects, a hedonic pricing model is constructed to explain housing price variation in the Netherlands. Based on these outcomes, detailed maps are created that show the potential revenue of different types of housing at the level of individual (1 hectare) grid cells. These high resolution price indices are created separately for single and multi-family housing, presenting results for a typical housing type for each of these types based on region-average characteristics.

The report briefly discusses how the obtained results can be applied in subsequent simulations of the potential profitability of different urban development strategies. It concludes with a short summary of the results and some suggestions for further research.

2. Data

This research makes use of a dataset with all housing transactions handled by the Dutch association of real estate brokers (NVM) between 2000 and 2016 and covers approximately 70% of all housing transactions in that period. The dataset was kindly provided by NVM and for each transaction contains information about the price, transaction characteristics, many structural housing characteristics and its exact location. The dataset is enriched with the Dutch Basic Registration of Addresses and Buildings (BAG) from Kadaster, which contains all buildings in the Netherlands and for example its building year. The housing transaction dataset contains some typos or other improbable values and therefore has to be cleaned. The rules mentioned in Table 1 are applied to clean the dataset.

Table 1: Steps taken to clean the NVM-dataset

Observations are dropped if:	
Price	> 2,500,000 euro
Price	< 25,000 euro
Size	> 250 m^2
Size	< 25 m^2
Price / m2	> 5,000 euro
Price / m2	< 500 euro
Number of rooms	> 25
Number of rooms	= 0
Located outside the Dutch border	
Lot size	< 10 m^2 AND no apartment
Lot size	> 99,999
Apartment type	= 0 AND is apartment
4-digit postcode area	Missing
Municipality code	Missing
Lot size	Missing AND no apartment
Income	Missing
Distance to 100,000 jobs	Missing

Several additional spatial datasets are used in the explanatory analysis. The distance datasets are taken from Broitman & Koomen (2015). Local land use shares are calculated as the number of hectares of that land-use type in 2012 divided by the total number of hectares in a municipality (CBS, 2012). The Urban Attractivity Index describes the presence of four types of urban amenities: historic buildings and monuments; cultural facilities; shops; hotels, restaurants and other catering establishments. The numbers of units per amenity type are counted per 500 m grid cell, rescaled to a maximum value of 0.25 per type, added up and averaged out over a 2.5 km radius (Broitman and Koomen, 2015). The municipal income averages in 2012[1] are taken from Statline (Statistics Netherlands, 2012). The descriptive statistics for single-family housing are shown in Table 2 and for multi-family housing in Table 3.

[1]Selecting the categories: '*Gemiddeld inkomen van personen naar regio in 2012*' and '*gemiddeld persoonlijk inkomen*' and using the municipal division of 2012.

Table 2: Descriptive statistics for single-family housing sold between 2000 and 2016.

Variable	**Unit**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Transaction price	euro	1,462,050	248,200	121,409	25,000	1,250,000
Ln transaction price		1,462,050	12.33	0.43	10.13	14.04
Floor space	m2	1,462,050	129.14	34.74	28	250
Ln floor space		1,462,050	4.83	0.26	3.33	5.52
Lot size	m2	1,462,050	475.87	2394.91	11	99999
Ln lot size		1,462,050	5.45	0.82	2.40	11.51
Number of rooms		1,462,050	4.88	1.19	1	25
Number of bathrooms		1,462,050	0.94	0.43	0	8
Central heating present	dummy	1,462,050	0.92	0.27	0	1
Complete isolation present	dummy	1,462,050	0.21	0.41	0	1
Private parking present	dummy	1,462,050	0.40	0.49	0	1
Garden present	dummy	1,462,050	0.97	0.18	0	1
Score for maintenance outside	score	1,462,050	0.76	0.13	0	1
Score for maintenance inside	score	1,462,050	0.75	0.14	0	1
Building year	year	1,403,806	1966	37	1005	2020
Leasehold present	dummy	1,462,050	0.03	0.17	0	1
Buyer’s costs[2] present	dummy	1,462,050	0.96	0.19	0	1
Transaction year	year	1,462,050	2008	5	2000	2016
Share of urban green land-use type in municipality	share	1,462,050	0.04	0.04	0.00	0.27
Share of residential land-use type in municipality	share	1,462,050	0.14	0.11	0.00	0.56
Share of nature land-use type in municipality	share	1,462,050	0.12	0.13	0.00	0.94
Share of agriculture land-use type in municipality	share	1,462,050	0.46	0.23	0.00	0.94
Distance to motorway ramp	meter	1,462,050	3742	3341	1	40138
Ln distance to motorway ramp		1,462,050	7.88	0.86	0.00	10.60
Distance to 100,000 jobs	meter	1,462,050	13029	6625	244	62591
Ln distance to 100,000 jobs		1,462,050	9.34	0.55	5.50	11.04
Distance to train stations	meter	1,462,050	5042	5535	25	45766
Ln distance to train stations		1,462,050	8.04	1.00	3.22	10.73
Urban Attractivity Index	score	1,462,050	4.72	8.90	0	87
Average income in municipality	1000 euro	1,462,050	30.84	3.54	24.00	53.50
Ln average income in municipality		1,462,050	3.42	0.11	3.18	3.98
Building construction year unknown	dummy	1,462,050	0.04	0.20	0	1
Building construction year before 1945	dummy	1,462,050	0.21	0.41	0	1
Building construction year between 1945-1960	dummy	1,462,050	0.07	0.26	0	1
Building construction year between 1960-1970	dummy	1,462,050	0.11	0.32	0	1
Building construction year between 1970-1980	dummy	1,462,050	0.17	0.38	0	1
Building construction year between 1980-1990	dummy	1,462,050	0.15	0.36	0	1
Building construction year between 1990-2000	dummy	1,462,050	0.15	0.36	0	1
Building construction year after 2000	dummy	1,462,050	0.08	0.28	0	1
Apartment house type	dummy	1,462,050	0.00	0.00	0	0
Terraced house type	dummy	1,462,050	0.62	0.49	0	1
Semi-detached house type	dummy	1,462,050	0.21	0.41	0	1
Detached house type	dummy	1,462,050	0.18	0.38	0	1

[2] Purchasing costs payable by the purchaser (*kosten koper* in Dutch).

Table 3: Descriptive statistics for multi-family housing sold between 2000 and 2016.

Variable	Unit	Obs	Mean	Std. Dev.	Min	Max
Transaction price	euro	645,307	181,971	92,193	25,000	1,200,000
Ln transaction price		645,307	12.01	0.43	10.13	14.00
Floor space	m2	645,307	85.98	26.60	26	247
Ln floor space		645,307	4.41	0.30	3.26	5.51
Number of rooms		645,307	3.22	0.99	1	23
Number of bathrooms		645,307	0.87	0.39	0	6
Central heating present	dummy	645,307	0.88	0.32	0	1
Complete isolation present	dummy	645,307	0.20	0.40	0	1
Private parking present	dummy	645,307	0.12	0.32	0	1
Score for maintenance outside	score	645,307	0.78	0.11	0	1
Score for maintenance inside	score	645,307	0.78	0.14	0	1
Building year	year	534,916	1965	56	1005	2020
Leasehold present	dummy	645,307	0.15	0.36	0	1
Buyer's costs present	dummy	645,307	0.92	0.26	0	1
Transaction year	year	645,307	2008	5	2000	2016
Share of urban green land-use type in municipality	share	645,307	0.08	0.05	0.00	0.27
Share of residential land-use type in municipality	share	645,307	0.22	0.10	0.00	0.56
Share of nature land-use type in municipality	share	645,307	0.09	0.12	0.00	0.94
Share of agriculture land-use type in municipality	share	645,307	0.27	0.20	0.00	0.94
Distance to motorway ramp	meter	645,307	2687	2178	1	40138
Ln distance to motorway ramp		645,307	7.61	0.80	0.00	10.60
Distance to 100,000 jobs	meter	645,307	7663	4937	244	59393
Ln distance to 100,000 jobs		645,307	8.74	0.68	5.50	10.99
Distance to train stations	meter	645,307	2667	3031	25	43180
Ln distance to train stations		645,307	7.51	0.86	3.22	10.67
Urban Attractivity Index	score	645,307	16.07	16.49	0	87
Average income in municipality	1000 euro	645,307	31.70	3.06	24	53.5
Ln average income in municipality		645,307	3.45	0.09	3.18	3.98
Building construction year unknown	dummy	645,307	0.17	0.38	0	1
Building construction year before 1945	dummy	645,307	0.16	0.37	0	1
Building construction year between 1945-1960	dummy	645,307	0.09	0.28	0	1
Building construction year between 1960-1970	dummy	645,307	0.16	0.37	0	1
Building construction year between 1970-1980	dummy	645,307	0.11	0.31	0	1
Building construction year between 1980-1990	dummy	645,307	0.10	0.30	0	1
Building construction year between 1990-2000	dummy	645,307	0.10	0.30	0	1
Building construction year after 2000	dummy	645,307	0.11	0.31	0	1
Apartment type: Downstairs	dummy	645,307	0.13	0.34	0	1
Apartment type: Upstairs	dummy	645,307	0.25	0.43	0	1
Apartment type: Maisonette	dummy	645,307	0.08	0.27	0	1
Apartment type: Porch	dummy	645,307	0.33	0.47	0	1
Apartment type: Gallery	dummy	645,307	0.19	0.40	0	1
Apartment type: Service	dummy	645,307	0.00	0.06	0	1
Apartment type: Downstairs and upstairs	dummy	645,307	0.01	0.08	0	1

3. Methodology

In this research three methods are used to calculate a hedonic price index using data on housing transactions between 2000 and 2016. A hedonic price analysis is a method to monetise specific measured characteristics of a priced object by estimating the implicit or hedonic price of that characteristic using a regression framework with many observations (Rosen, 1974). For example, if two adjacent houses are identical except for a garage and the one with the garage costs 10,000 euro more. It could be assumed that the price of that garage is 10,000 euro. Using the same rationale, if two identical houses only differ in location, one can derive the location premium.

3.1 Specification 1: using region fixed effects at municipality level

In the first specification, the transaction price is explained by structural characteristics of the house[3] combined with transaction characteristics, transaction year dummies[4] and municipality dummies[5]:

$$\ln(P_{itr}) = \alpha + \beta S_i + \delta dT + \gamma_r dR + \varepsilon \qquad \text{(Eq. 1)}$$

Wherein P_{itr} is the transaction price of house i, at time t in region r, S_i a vector of structural housing characteristics, dT a set of transaction year dummies, dR a region-specific dummy variable, α the constant, and ε the error term. This equation provides, amongst others, coefficients for each region that capture spatial variation in prices.

The region-specific coefficient in this equation can be explained in a separate regression using several spatial variables:

$$\gamma_r = \theta + \pi L_r + \sigma \qquad \text{(Eq. 2)}$$

Wherein γ_r is a vector of to be explained region coefficients, L a vector of spatial variables, θ the constant and σ the error term. The spatial variables are, for example, distance to 100,000 jobs, distance to the nearest train station or the share of land-use type nature in a municipality. Furthermore, we approximate the quality of the living environment in the region with the average income in the municipality. As a measure for the amount of amenities, the Urban Attractivity Index is used. In order to choose a proper specification, all

---

[3] These structural component includes dummies for house type, the largest group is chosen as the reference category (terraced housing for single-family housing and porch apartments for multi-family housing).

[4] For the transaction year dummies is the year 2000 chosen as a reference category.

[5] The municipality with the lowest average price in the full dataset, is chosen as reference category (GM0765: Pekela in the eastern part of the province of Groningen).

different combinations of sets of variables are created and checked for their significance and robustness of the estimated coefficients.

The benefit of the second equation is that it provides a description of the contribution of different spatial variables in explaining regional variation in house prices. In combination with data sets describing the local variation of these variables we can use the outcomes of Eq. 2 to simulate spatial variation within the regions (in this case municipalities) that are used to define the regional price component. So, if house price differences between regions can be explained from, for example, regional differences in amenity levels, we expect that the variation within the regions themselves will also depend on local variation in amenity levels. Intuitively this makes sense: if Amsterdam has higher house process than its neighbouring municipalities because of its higher, average amenity levels, the local variation within Amsterdam is also likely to follow differences in amenity levels. This approach allows us to generate house prices based on structural and location-specific characteristics for each hectare in the Netherlands. These prices result from Equation 3, which is the combination of Equation 1 and 2:

$$P_{itr} = e^{\alpha} * e^{\beta S_i} * e^{\delta dT} * e^{(\theta + \pi L_r + \sigma) dR} * e^{\varepsilon} \qquad \text{(Eq. 3)}$$

We apply Equation 3 to each location in the Netherlands to obtain the potential, local revenues of housing development. This operation is performed in the GeoDMS modelling environment that allows for fast structured calculations using very large datasets.

3.2 Specification 2: using region fixed effects at 4-digit postcode level

The second specification is nearly the same as the first, except for the scale of the included regions. In this case, 4-digit postcode areas[6] are used instead of municipalities. The assumption is that by using smaller regions the to be explained variation is smaller because it shows less heterogeneity and is, therefore, better explained.

3.3 Specification 3: without region fixed effects

In the third specification, the price is explained in a more straightforward method. It uses the same structural characteristics and transaction year dummies, but no region-specific dummies are included as fixed effects. Instead, spatial variables are directly included in the specification (Equation 4) and subsequently the revenues for each location are calculated using the GeoDMS environment.

$$P_{itr} = e^{\alpha + \beta S_i + \delta d\_T + \pi L_r + \varepsilon} \qquad \text{(Eq. 4)}$$

[6] The 4-digit postcode area with the lowest average price is chosen as reference category (region 8507 in Friesland).

4. Results

In this chapter, the results of the research are presented. In the first subsection, the single-family housing results are described with extensive explanations. In the second subsection, we limit ourselves to only showing the results for multi-family houses and provide a less extensive explanation. Note that we use the term houses here to describe individual housing units. Technically several units may comprise one larger house or building, but we analyse prices at the level of individual units that are sold separately on the market.

4.1 Single-family housing

The results of the regression analysis for single-family houses are shown in Table 4. First, we discuss the initial regressions (specification 1 and 2) where only the structural and transaction characteristics are included with municipal (1) dummies and 4-digit postcode (2) dummies. Specification 3 includes the spatial variables. The estimated region coefficients from specification 1 and 2 are explained using the spatial variables listed in Table 5. Table 10 in Appendix 1 lists the estimated coefficients of the transaction year and construction year dummies. As reference categories we selected 'terraced housing' as housing type and 'build after 2000' as building year.

The results for all three specifications show similar and expected coefficients, except for leasehold (*Erfpacht* in Dutch). In specification 3, the coefficient is suddenly positive. However, this can be explained by the fact that leasehold is almost exclusively used in the municipality of Amsterdam that has above average prices. This effect is captured when using region fixed effect but is not captured when we only use spatial variables.

Table 5 shows similar coefficients for both the municipal region dummies and the 4-digit region dummies. Both with the same explanatory power.

Table 4: Regression results for single-family houses

Variable	**(1)** **Ln price**	**(2)** **Ln price**	**(3)** **Ln price**
Ln floor space	0.647***	0.576***	0.651***
	(0.000833)	(0.000753)	(0.000943)
Ln lot size	0.136***	0.154***	0.143***
	(0.000296)	(0.000276)	(0.000342)
Number of rooms	0.0176***	0.0148***	0.0162***
	(0.000159)	(0.000141)	(0.000180)
Number of bathrooms	0.0353***	0.0258***	0.0363***
	(0.000384)	(0.000339)	(0.000438)
Central heating present	0.0553***	0.0567***	0.0583***
	(0.000611)	(0.000539)	(0.000698)
Score for maintenance inside	0.281***	0.298***	0.276***
	(0.00212)	(0.00186)	(0.00243)
Score for maintenance outside	0.160***	0.139***	0.166***
	(0.00229)	(0.00201)	(0.00262)
Complete isolation present	0.0104***	0.0178***	0.0181***
	(0.000496)	(0.000441)	(0.000563)
Private parking present	0.0767***	0.0712***	0.0628***
	(0.000393)	(0.000348)	(0.000448)
Garden present	0.0410***	0.0337***	0.0501***
	(0.000972)	(0.000860)	(0.00111)
Buyer's costs present	-0.0284***	-0.0399***	-0.0499***
	(0.00116)	(0.00104)	(0.00132)
Leasehold present	-0.116***	-0.0932***	0.00496***
	(0.00109)	(0.00113)	(0.00107)
Semi-detached house type	0.0957***	0.0935***	0.0826***
	(0.000472)	(0.000424)	(0.000533)
Detached house type	0.183***	0.194***	0.176***
	(0.000624)	(0.000563)	(0.000702)
Share of nature land-use type in municipality			0.0235***
			(0.00146)
Share of agriculture land-use type in municipality			-0.0887***
			(0.000955)
Ln distance to 100,000 jobs			-0.121***
			(0.000424)
Urban Attractivity Index			0.00501***
			(2.53e-05)
Ln average income in municipality			1.346***
			(0.00183)
Constant	7.072***	7.064***	4.248***
	(0.00667)	(0.0208)	(0.00925)
Observations	1,462,050	1,462,050	1,462,050
R-squared	0.812	0.857	0.753
Transaction year fixed effects	Yes	Yes	Yes
Municipality fixed effects	Yes	No	No
4-digit postcode fixed effects	No	Yes	No
House type dummies	Yes	Yes	Yes
Building year category dummies	Yes	Yes	Yes
Transaction year >= 2000	Yes	Yes	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Regression results of the explanation of the region-coefficients of single-family houses

Variable	(1) Mun. coeff.	(2) PC4 coeff.
Ln average income in municipality	1.340***	1.545***
	(0.0681)	(0.0307)
Average Urban Attractivity Index in region	0.0169***	0.00745***
	(0.00295)	(0.000417)
Ln distance to 100,000 jobs	-0.104***	-0.149***
	(0.0221)	(0.00800)
Share of agriculture land-use type in municipality	-0.156***	-0.231***
	(0.0379)	(0.0153)
Share of residential land-use type in municipality	-0.469***	-0.271***
	(0.118)	(0.0468)
Constant	-12.13***	-13.50***
	(0.836)	(0.354)
Observations	414	3,683
R-squared	0.691	0.698

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To implement Equations 3 and 4, we use the average values for all the structural housing characteristics in the corresponding region[7], as well as the local (grid cell) values for the spatial variables, and we use the 2016 year coefficient to express the resulting housing price in 2016-euros. In the following figures, the results for the structural (Equation 1 without region dummies) and spatial (Equation 2) components are split and the last figure shows them combined (Equations 3 and 4). Each step is done separately for municipal fixed effects and 4-digit postcode fixed effects.

The first figure shows the price for only the physical structures based on the region-specific values for its structural characteristics. So it combines the national estimates of the housing price contribution of different structural characteristics with regional variation in the distribution of these characteristics. A striking feature of these results are the higher housing values for more peripheral locations. This is explained from the fact that houses in these regions are, for example, larger on average. As expected, house price variation based on structural characteristics is much larger when looking at 4-digit postcode regions. This can be observed from the wider range of values in the legend (Figure 1 at bottom). The relative of prevalence of lower values (in green) seems to suggest that the higher values occur in fewer regions than the lower values.

Figure 2 shows the computed values for the region-specific coefficients that includes all spatial variation and that can be interpreted as region-specific multipliers[8]. The municipality and 4-digit postcode results show very similar patterns. Combining the structural and spatial component results in the map in Figure 3. Figure 4 shows the results for alternative specification 3. The result is very similar to the 4-digit postcode region map in Figure 3, although the latter shows a slightly higher maximum value which seems less plausible.

[7] If there are no observations in a 4-digit postcode area, the municipal average is taken.

[8] Note that the computed values have a higher spatial resolution, but smoother general patterns than the initial region-specific coefficients obtained in Eq. 1 (see Figure 11 in Appendix 2 for a graphical depiction of these coefficients).

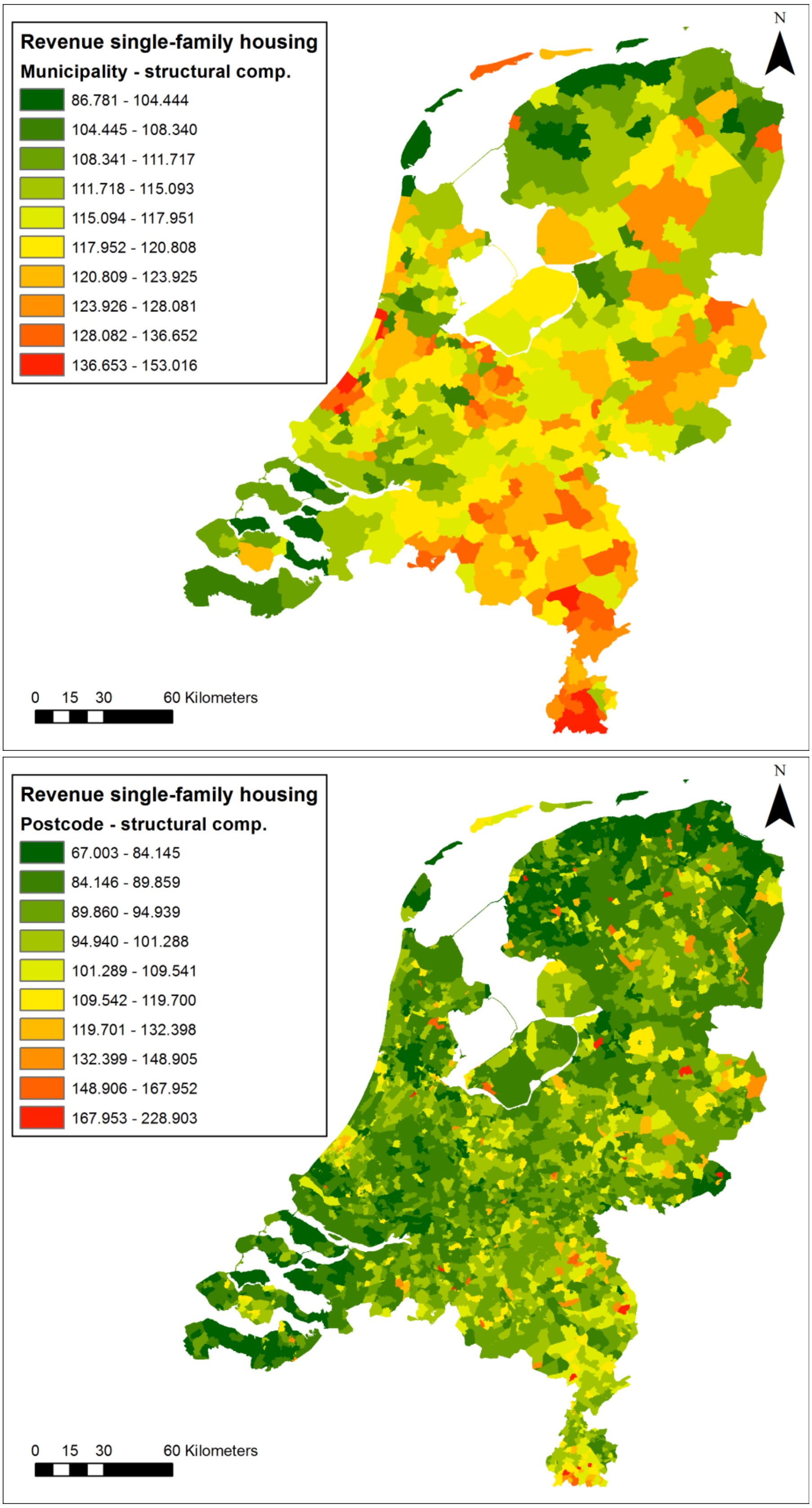


Figure 1: Revenue for single-family housing: structural component. Values are in euros and refer to terraced housing, indicating only the structural and transaction components of the transaction price using region-specific averages for structural characteristics. Results are shown for municipalities (top; specification 1) and 4-digit postcode regions (bottom; specification 2).

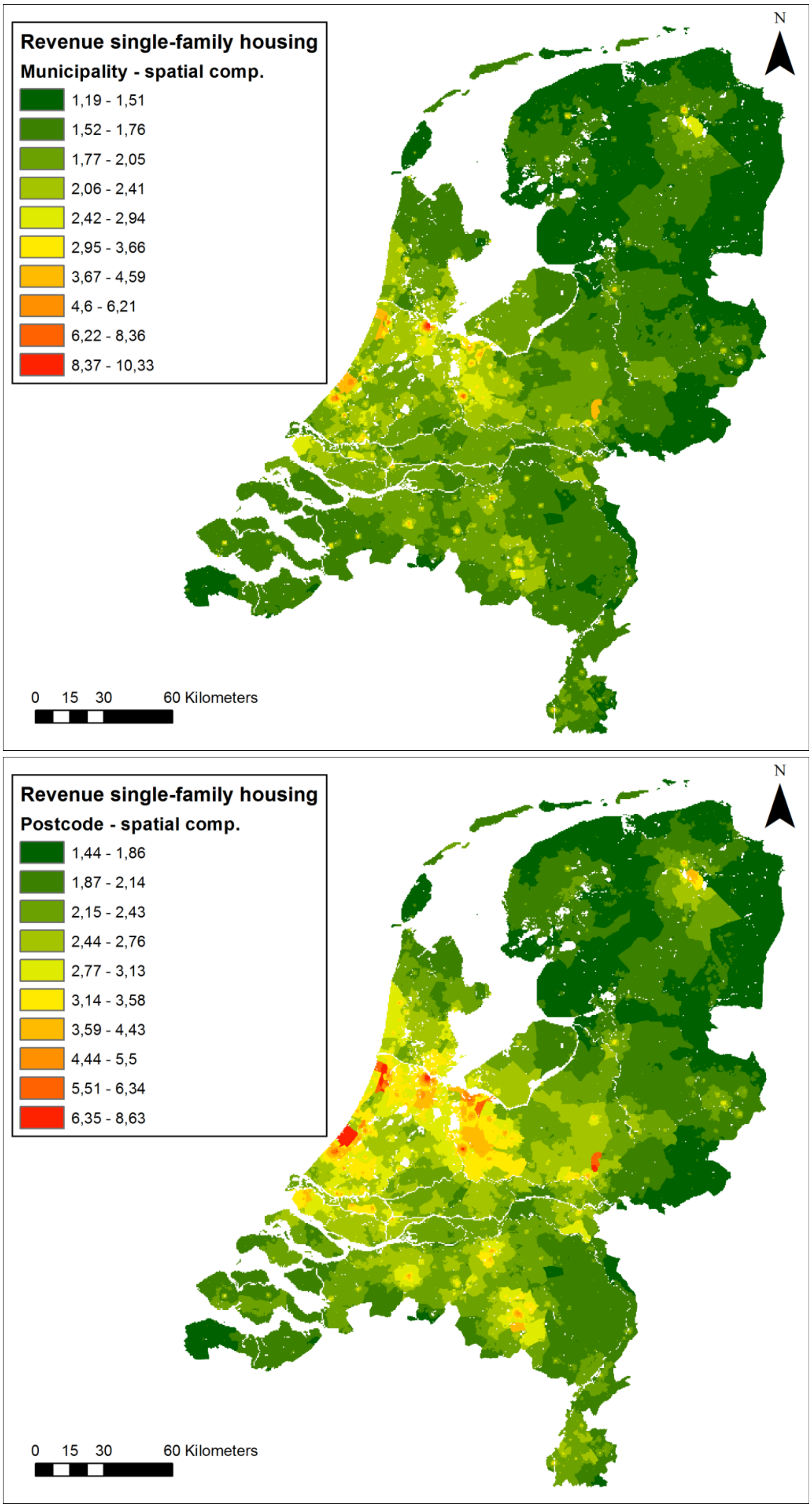
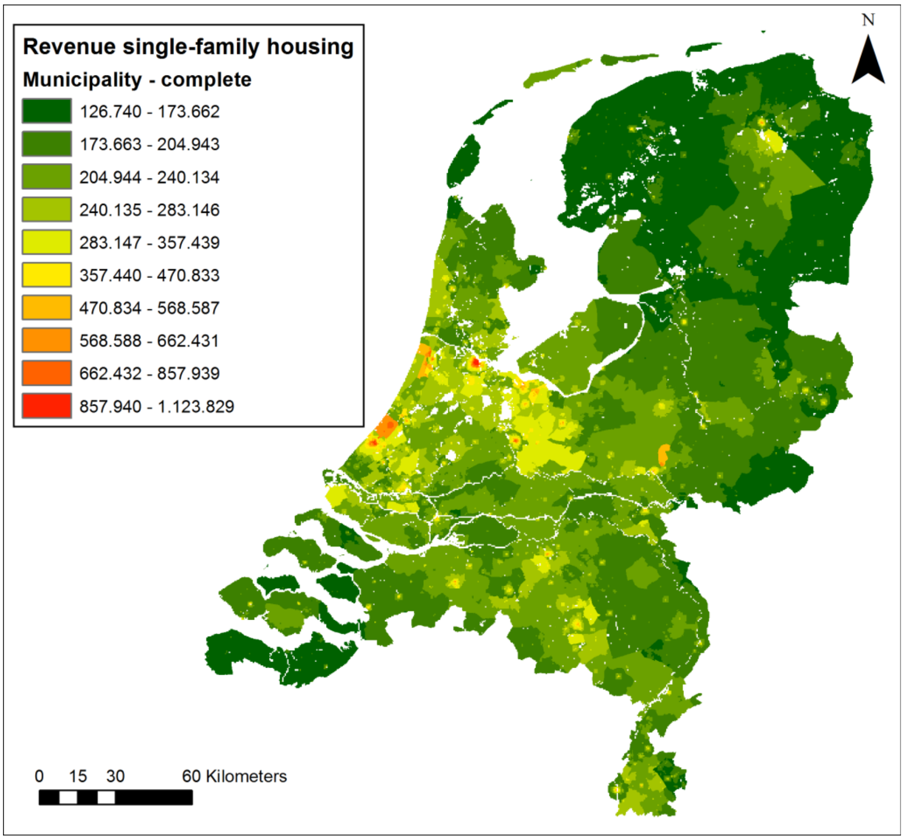


Figure 2: Revenue for single-family housing: spatial component (Equation 2). Values refer to terraced housing and can be interpreted as a location multiplier. The structural component is multiplied by the values from this spatial component to obtain the potential revenues. Results are shown for municipalities (top; specification 1) and 4-digit postcode regions (bottom; specification 2).



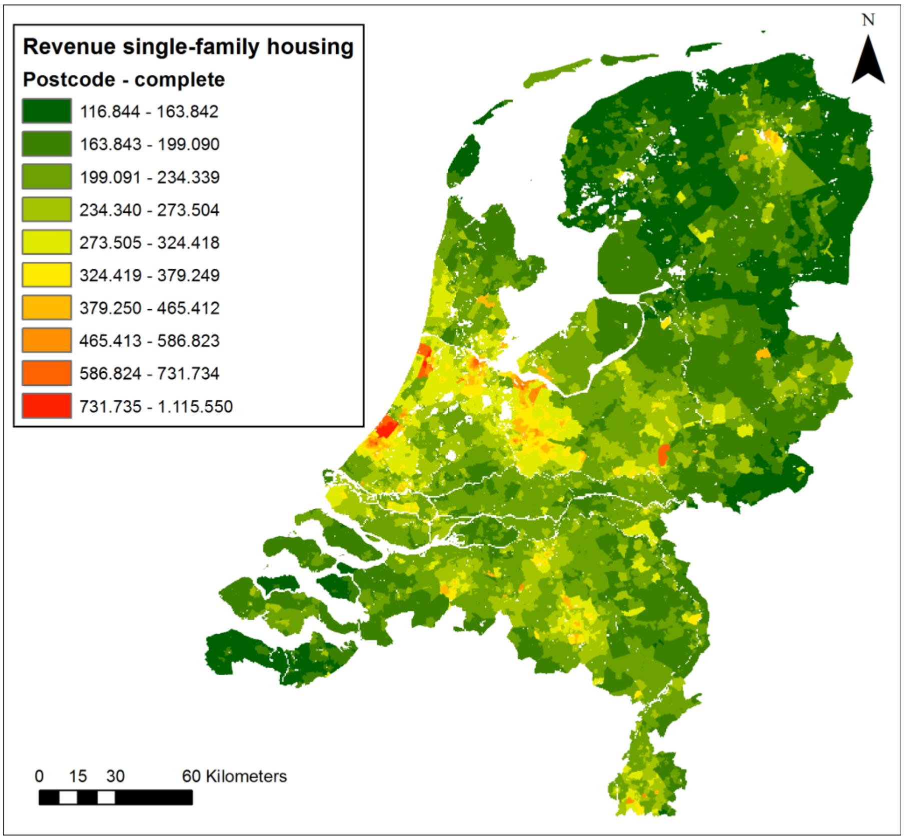


Figure 3: Combined potential revenue for single-family housing (equation 4). The structural component is multiplied by the spatial component to give potential revenues per location in euros. These revenues refer to terraced housing. Results are shown for municipalities (top; specification 1) and 4-digit postcode regions (bottom; specification 2).

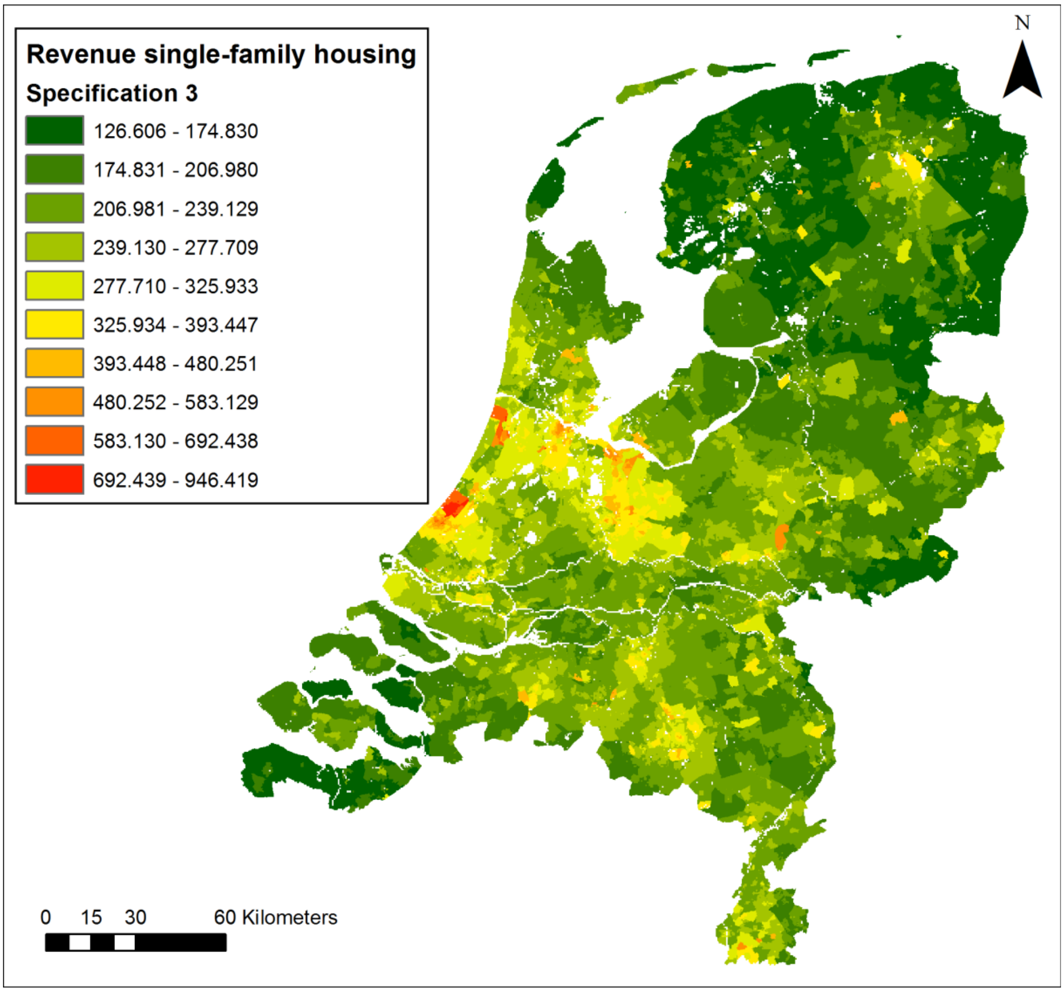


Figure 4: Revenue for single-family housing: specification 3. This figure shows the potential revenues in euros, when the spatial variables are included instead of region fixed effects. These revenues refer to terraced housing.

4.2 Multi-family housing

The results of the regression analysis for multi-family houses are shown in Table 6. First, we discuss the initial regressions (specifications 1 and 2) where only the structural and transaction characteristics are included with municipal (1) dummies and 4-digit postcode (2) dummies. Specification 3 includes the spatial variables. The estimated region coefficients from specification 1 and 2 are explained using a limited set of spatial variables (Table 7). As reference categories we selected 'porch apartments' as housing type and 'build after 2000' as reference year. Table 11 in Appendix 1 lists the estimated coefficients of the transaction year and construction year dummies.

Similar to the results of single-family houses, the estimated coefficients are as expected. Moreover, most coefficients show the same sign and magnitude. The number of rooms shows a negative sign for multi-family housing compared to a positive sign for single-family housing. This can possibly be explained from the fact that multi-family houses are usually smaller and by dividing that smaller space into more rooms the utility of that house may get lower. Another difference is that private parking shows a larger coefficient for multi-family housing, implying that this characteristic is more important for the price. This seems plausible because single-family houses are more likely to have a private parking spot (or at least more easy access to parking place) thus limiting its relative importance.

Table 6: Regression results for multi-family houses

Variable	**(1)** **Ln price**	**(2)** **Ln price**	**(3)** **Ln price**
Ln floor space	0.811***	0.732***	0.764***
	(0.00133)	(0.00112)	(0.00143)
Number of rooms	-0.00380***	0.0115***	-0.00432***
	(0.000395)	(0.000328)	(0.000425)
Number of bathrooms	0.0396***	0.0274***	0.0467***
	(0.000739)	(0.000601)	(0.000800)
Central heating present	0.0566***	0.0429***	0.0559***
	(0.000906)	(0.000743)	(0.000981)
Score for maintenance inside	0.367***	0.349***	0.372***
	(0.00267)	(0.00216)	(0.00290)
Score for maintenance outside	0.216***	0.147***	0.204***
	(0.00366)	(0.00297)	(0.00398)
Complete isolation present	0.0162***	0.0289***	0.0250***
	(0.000890)	(0.000734)	(0.000959)
Private parking present	0.122***	0.114***	0.116***
	(0.000955)	(0.000802)	(0.00102)
Buyer's costs present	0.0138***	-0.0488***	-0.00668***
	(0.00128)	(0.00108)	(0.00138)
Leasehold present	-0.0812***	-0.0236***	0.0236***
	(0.000923)	(0.000863)	(0.000889)
Share of nature land-use type in municipality			0.0204***
			(0.00174)
Share of agriculture land-use type in municipality			-0.0998***
			(0.00256)
Ln distance to 100,000 jobs			-0.00798***
			(0.000590)
Urban Attractivity Index			0.00625***
			(2.31e-05)
Ln average income in municipality			1.076***
			(0.00334)
Constant	7.198***	8.153***	4.213***
	(0.0406)	(0.0212)	(0.0150)
Observations	645,307	645,307	645,307
R-squared	0.745	0.835	0.697
Transaction year fixed effects	Yes	Yes	Yes
Municipality fixed effects	Yes	No	No
4-digit postcode fixed effects	No	Yes	No
Apartment type dummies	Yes	Yes	Yes
Building year category dummies	Yes	Yes	Yes
Transaction year >= 2000	Yes	Yes	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7 shows that the coefficients for the municipal dummies and the 4-digit postcode dummies are less similar than for single-family housing. In this case, the income coefficient is much lower, which also gives a much lower constant. Moreover, the explanatory power is much lower for the multi-family housing (approx. 0.45 compared to approx. 0.69).

Table 7: Regression results of the explanation of the region-coefficients of multi-family houses

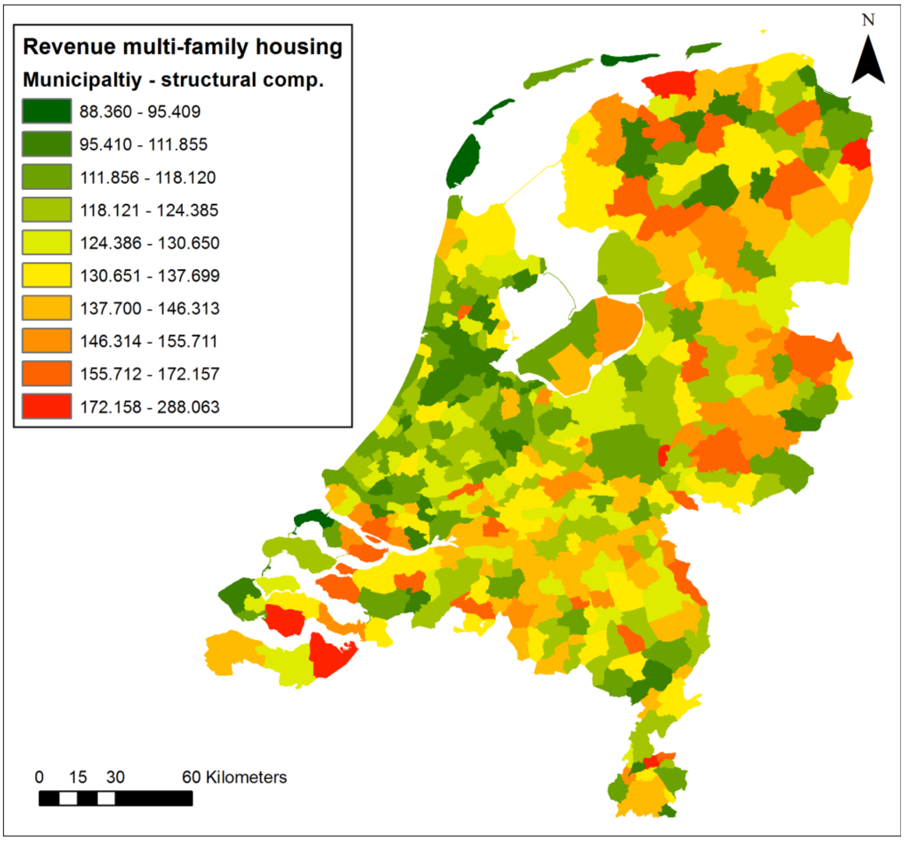
Variable	(1) Mun. coeff.	(2) PC4 coeff.
Ln average income in municipality	0.799***	1.248***
	(0.0602)	(0.0466)
Average Urban Attractivity Index in region	0.00634**	0.00633***
	(0.00257)	(0.000560)
Ln distance to 100,000 jobs	-0.0467**	-0.0435***
	(0.0191)	(0.0123)
Share of agriculture land-use type in municipality	-0.0908***	-0.0452*
	(0.0332)	(0.0253)
Share of residential land-use type in municipality	-0.397***	-0.313***
	(0.103)	(0.0694)
Constant	-7.290***	-12.53***
	(0.735)	(0.532)
Observations	392	1,640
R-squared	0.460	0.433

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Again Equation 3 and 4 are implemented and plotted in maps. The first two figures (Figure 5) show the revenues for only the physical structures given the region-specific values for its characteristics. Similar to single-family results we see higher housing values for more peripheral locations and larger variation at the more detailed 4-digit postcode level.

Figure 6 shows the computed value for the region-specific coefficients. It displays some difference between the municipality and 4-digit postcode results, but much less compared to the single-family results. The originally estimated region-specific coefficients from Equation 1 are listed Figure 12 in Appendix 2.

Figure 7 shows the total housing prices that result from the combination of structural and spatial components. It shows similar results for municipal and postcode regions. The results for specification 3 are depicted in Figure 8. As with the single-family results the spatial patterns similar to those in Figure 7.



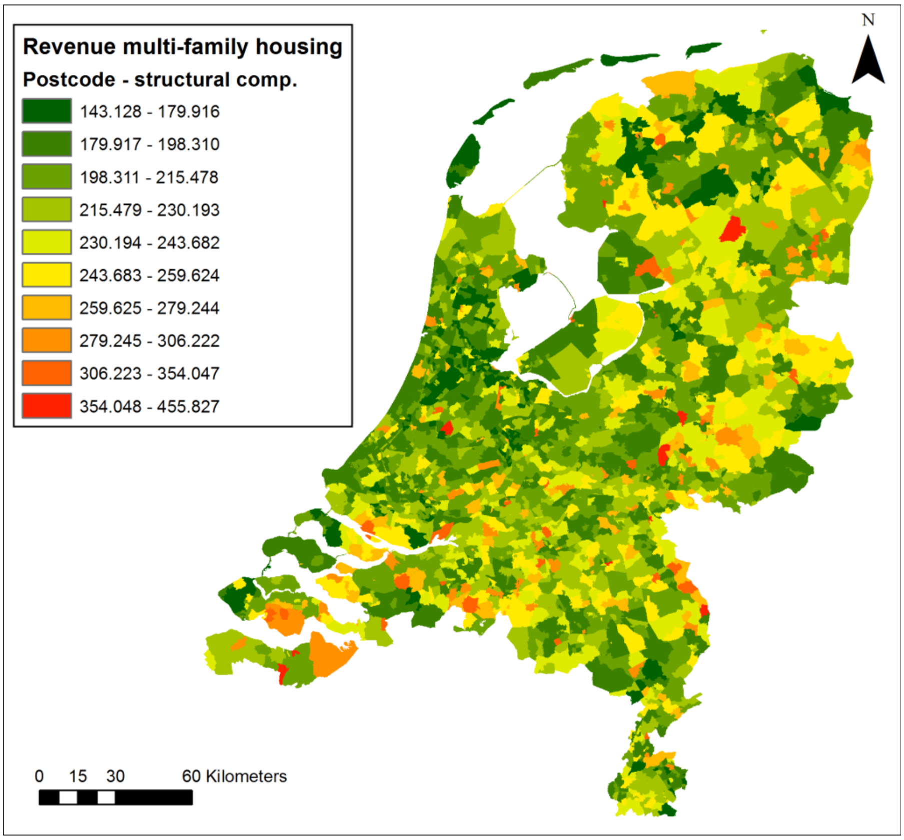
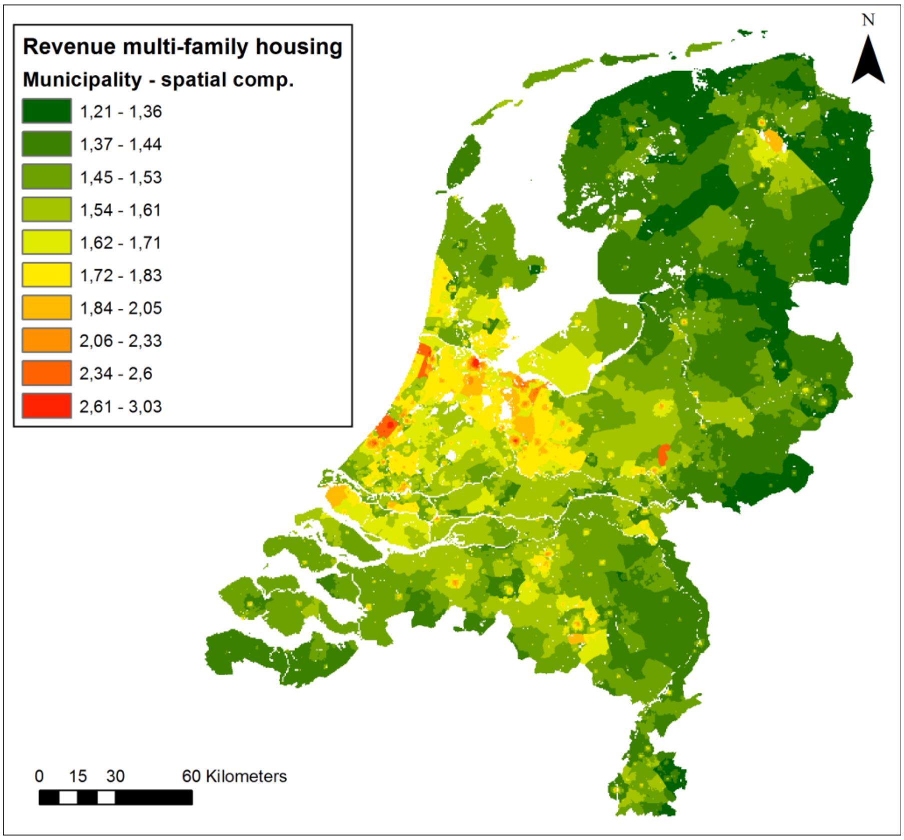


Figure 5: Revenue for multi-family housing: structural component. Values are in euros and refer to porch apartments, indicating only the structural and transaction components of the transaction price using region-specific averages for structural characteristics. Results are shown for municipalities (top; specification 1) and 4-digit postcode regions (bottom; specification 2).



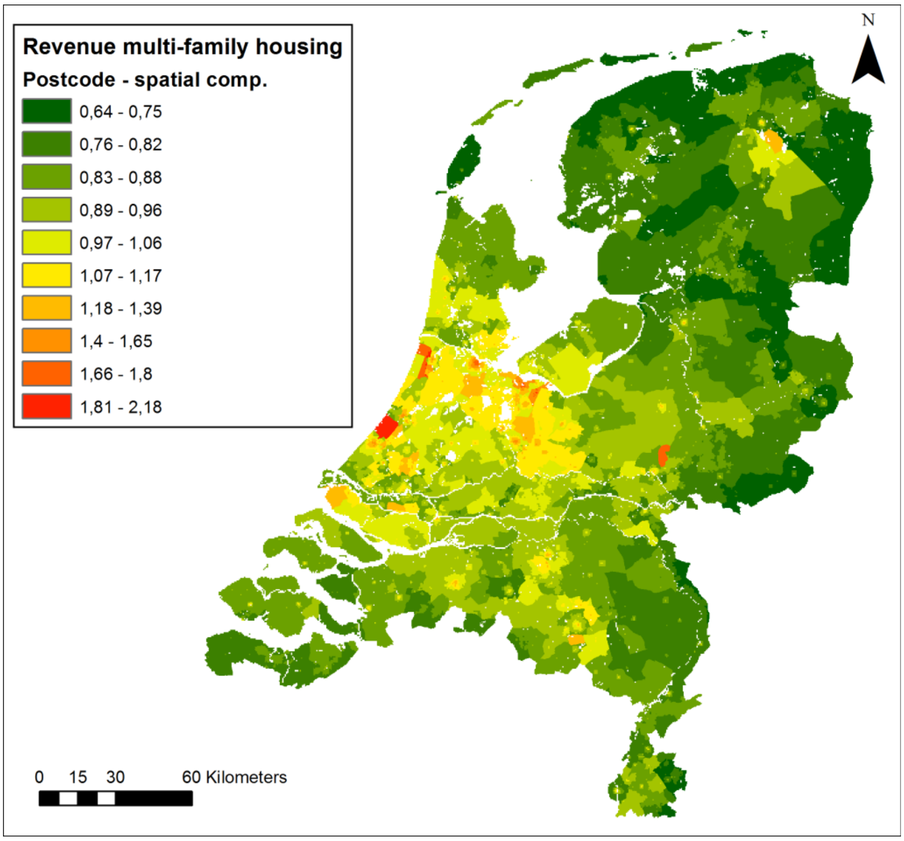


Figure 6: Revenue for multi-family housing: spatial component (Equation 2). Values are a location multiplier and refer to porch apartments. The structural component is multiplied by the values from this spatial component to obtain the potential revenues. Results are shown for municipalities (top; specification 1) and 4-digit postcode regions (bottom; specification 2).

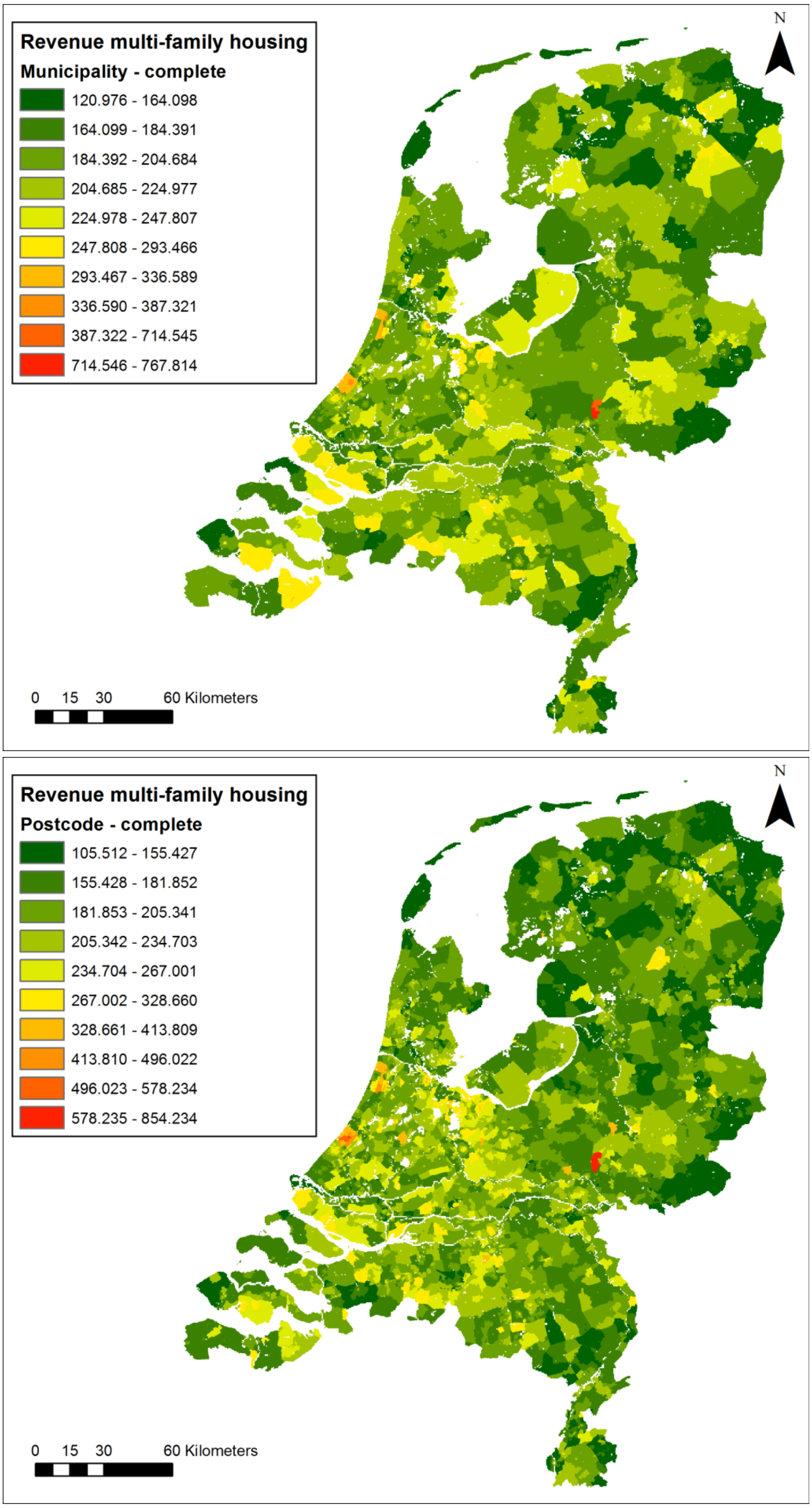


Figure 7: Revenue for multi-family housing (equation 4). The structural component is multiplied by the spatial component to give potential revenues per location in euros. These revenues refer to porch apartments. Results are shown for municipalities (top; specification 1) and 4-digit postcode regions (bottom; specification 2).

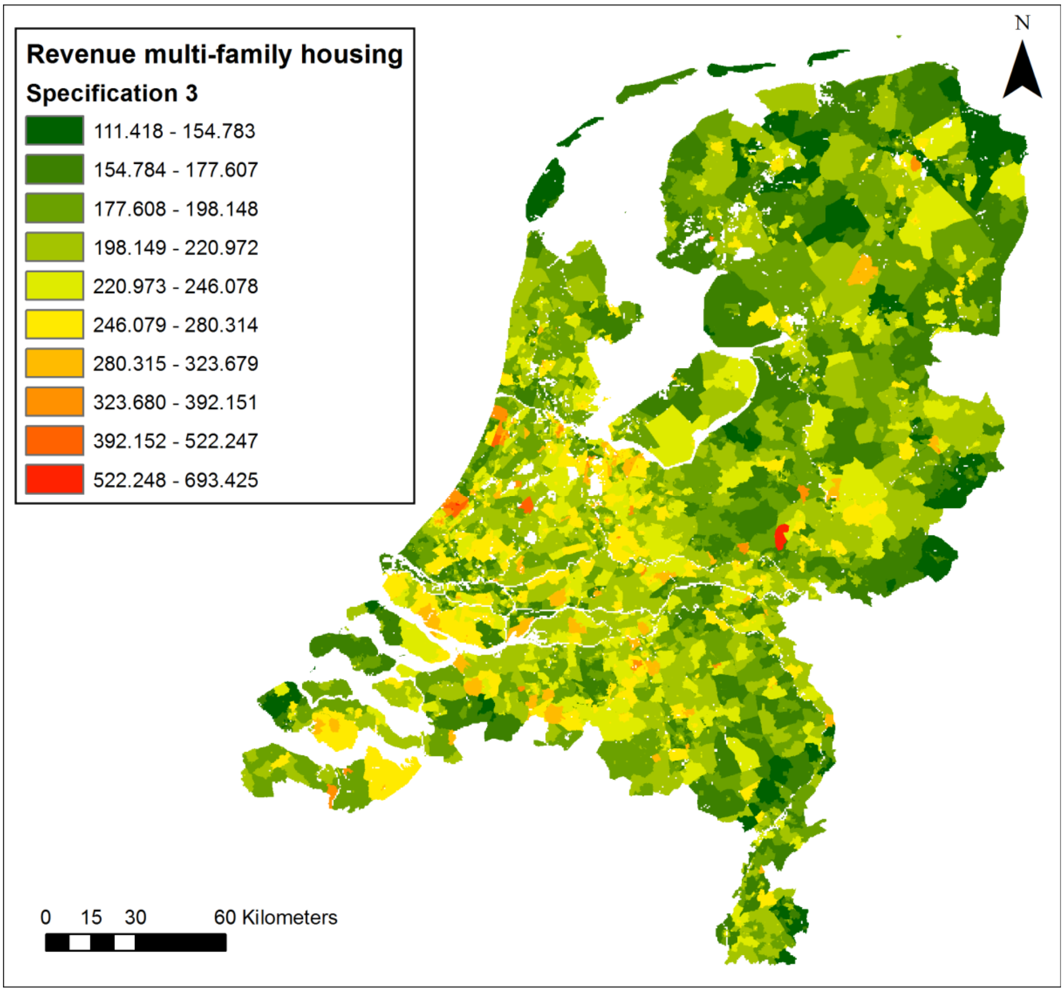


Figure 8: Revenue for multi-family housing: specification 3. This figure shows the potential revenues in euros, when the spatial variables are included instead of region fixed effects. These revenues refer to porch apartments.

5. Application of the results

The obtained high resolution housing price indices can be used to simulate local variation in the potential profitability of different urban development strategies. PBL has developed a spatially-explicit land-use modelling environment (Land Use Scanner) that can be used to simulate the competition between different types of housing. The price indices documented in this study can be used in combination with information on the cost of constructing these housing types to assess their net benefits and hence their profitability relative to other types of housing or alternative uses of the land (as discussed in more detail in Koomen et al., 2015).

To help define indicative urban development options, an exploratory analysis is performed to characterise Dutch housing environments at the spatial resolution of the Land Use Scanner model. We assumed density per hectare to be the most important measure to characterise spatial variation in housing environments in a limited number of distinct classes. For this exploratory analysis, we used the BAG snapshot of January 1st, 2017 and selected only those grid cells (with a size of one hectare) that contained at least one housing unit. We categorised those cells into categories that roughly correspond with percentiles[9]. Furthermore, we were interested in the composition of these cells in terms of housing types and presence of non-residential functions and buildings. Using the BAG-data we therefore also calculated the fraction of single-family houses in that cell (the remainder being multi-family housing units), as well as the fraction of non-housing functions and the fraction of non-housing buildings. The latter two, of course, are closely related.

Table 8: Housing composition of the cells containing at least 1 housing unit classified into percentiles based on the 2017 BAG-dataset.

Percentile	Density	Cells	Housing units	Fraction single-family	Fraction non-housing functions	Fraction non-housing buildings
P0-30	1 - 4	290,608	478,473	0.95	0.18	0.16
P30-40	5 - 7	40,402	238,367	0.93	0.16	0.15
P40-50	8 - 12	47,653	470,945	0.92	0.16	0.15
P50-60	13 - 17	38,316	571,603	0.90	0.16	0.14
P60-70	18 - 23	38,861	793,050	0.87	0.15	0.14
P70-80	24 - 31	42,027	1,147,470	0.84	0.15	0.14
P80-90	32 - 43	37,626	1,381,010	0.75	0.15	0.14
P90-99	44 - 117	33,781	2,132,120	0.36	0.17	0.16
P99-995	118 - 300	2,903	442,358	0.03	0.20	0.18
P995-100	301 - 731	46	17,810	0.00	0.25	0.25
Total		*572,223*	*7,673,206*	*0.66*	*0.17*	*0.16*

[9] We based this classification on percentiles reflecting the frequency distribution of cells with housing units ordered by the number of units they contain. As we have roughly 570,000 cells with housing units we would expect the top 10 percent to relate to the 57,000 cells that contain the highest number of housing units. The numbers of cells per percentile class in Table 8 deviate a bit from these expected values as we applied rounded density values as class boundaries.

Table 9 shows the same type of percentiles, but then only for those areas that were (re)developed between 2000 and 2012. These (re) developed areas are based on the CBS 1-hectare squares (as BAG data is unavailable for that period) and defined as those cells in which more than 10 housing units were added between 2000 and 2012, and that are bordered by at least 4 adjacent cells that also experienced the same minimum increase. The spatial variation in the resulting housing composition classes is illustrated for the Amsterdam region in Figure 10.

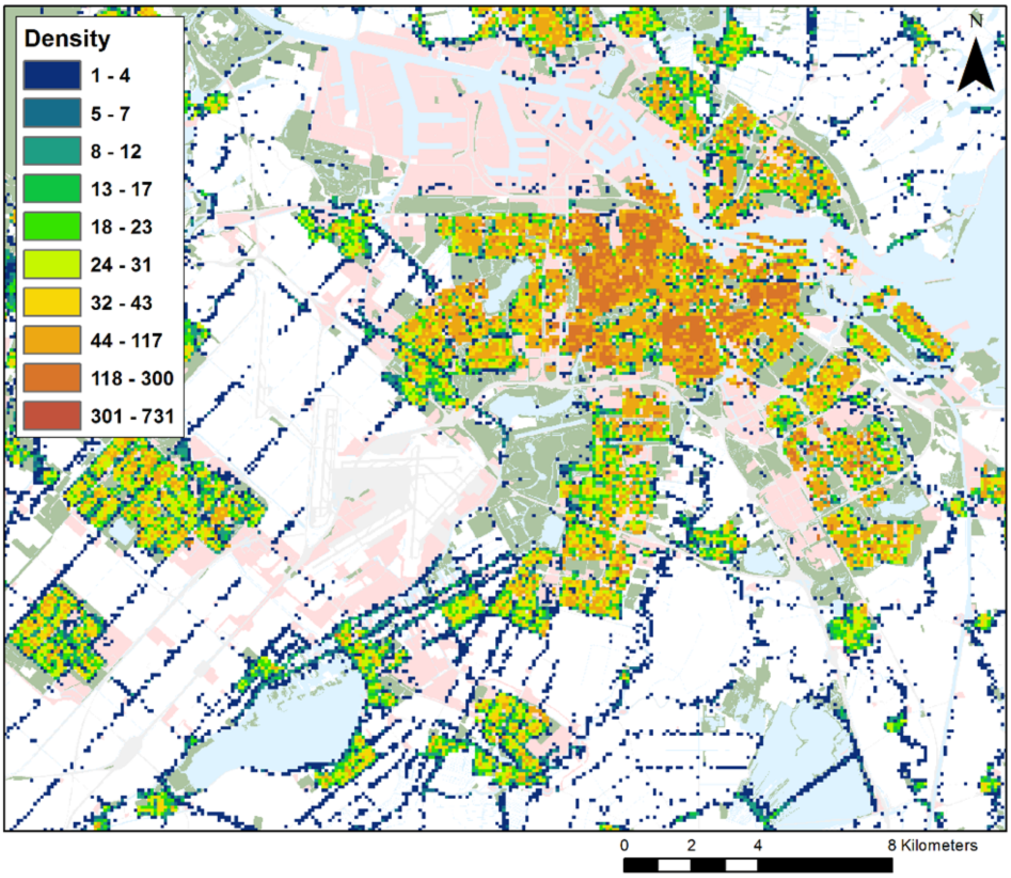
From these two tables, several conclusion can be drawn. First, a large share of the residential area consists of very low density housing. High densities are only found on a relatively small area, but these cover a large share of the total amount of housing units. From Figure 9 (at left), you can see that 50% of the cells that contain housing units, account for only 15% of the units in the Netherlands. Furthermore, the lower densities consist almost entirely of single-family housing, while the highest densities are almost exclusively comprised of multi-family housing. Remarkably, the fraction of non-housing functions and non-housing buildings is rather stable over the percentiles. Only in the high-density percentiles it is slightly higher. The large-scale development areas are much more homogeneous: the 50% cells with the lowest densities contains 32% of the housing units. In this case the fraction of non-housing functions and buildings is also close to zero. This homogeneity can, furthermore, be seen in Figure 9 (at right), where the line depicting the cumulative frequency distribution is much closer to the 45° line representing a complete homogenous distribution.

Table 9: Housing composition of the (re)development areas observed in the CBS-squares dataset. This selection of residential grid cells concerns groups of 5 or more adjacent cells in all of which the number of housing units increased by at least 10 between 2000 and 2012. The selected grid cells are characterized in the same way as in Table 8.

Percentile	Density	Average density	Cells	Houses	Fraction single-family	Fraction non-housing functions	Fraction non-housing buildings
P0-30	1 - 19	14	5,826	79,320	0.90	0.02	0.02
P30-50	20 - 27	23	3,577	83,374	0.85	0.03	0.02
P50-70	28 - 36	32	3,023	95,591	0.80	0.03	0.02
P70-90	37 - 60	45	2,914	131,766	0.59	0.03	0.03
P90-100	61 - 624	91	1,247	113,913	0.15	0.06	0.06
Total			*16,587*	*503,964*	*0.66*	*0.03*	*0.03*

Figure 9: Housing distribution in the Netherlands. The cumulative share of the number of houses per hectare is plotted against the share of cumulative land area.

The described housing unit classes can be used to define urban development options that bear resemblance to the actual composition of residential areas. For example, a typical development option (P30-P50) could consist of 23 housing units per hectare, containing 85% single-family housing units and 3% other functions like shops. The total revenues for this development option at a particular location in the Netherlands can be inferred from the results presented in Figure 4 and Figure 8. If we take the Amsterdam area by way of example, the estimated values for houses with regional-average characteristics are around 400,000 euro for a single-family house and 250,000 euro for a multi-family housing unit. So, a to-be-developed hectare could potentially result in 0.85 * 23 = 20 single-family houses and 0.15 * 23 = 3 multi-family housing units, which results in 20 * 400,000 + 3 * 250,000 = 8,750,000 euro in gross revenue. Likewise a denser development option (P90-P100) would result in 0.15 * 91 = 14 single family house and 0.85 * 91 = 77 multi-family houses and a total gross value of 24,850,000 euro.



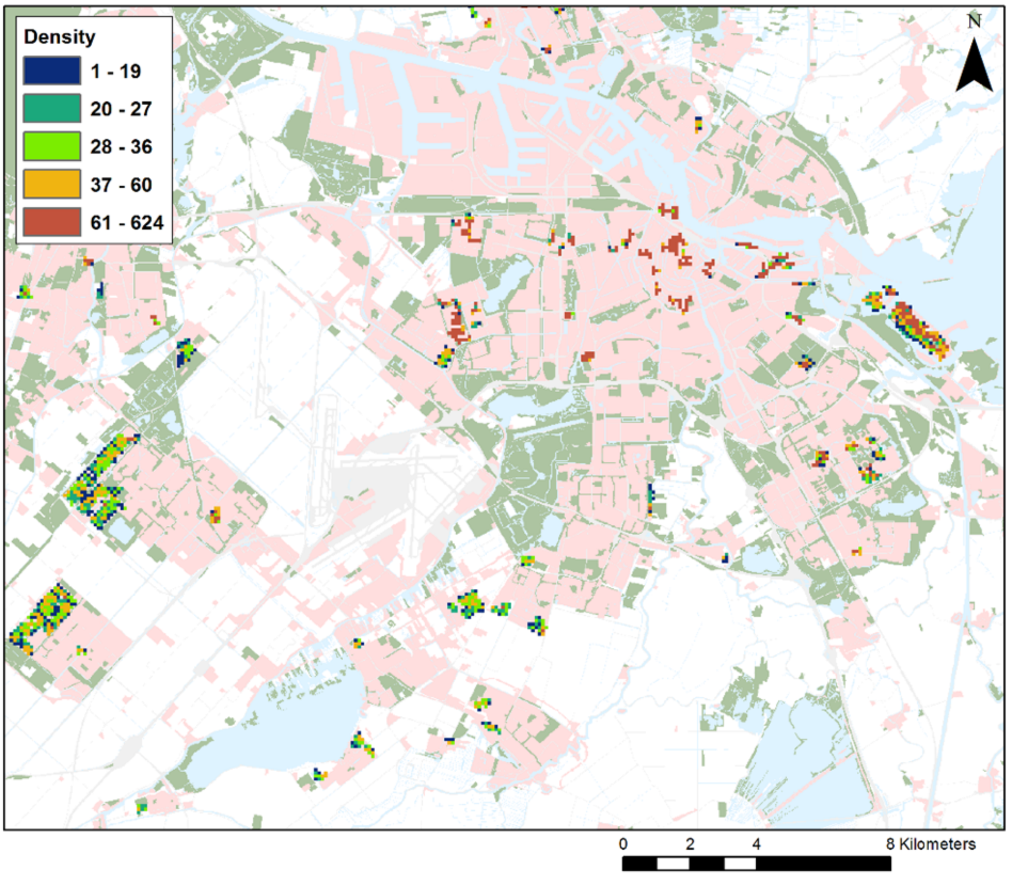


Figure 10: Housing composition classes based on all residential areas in the Netherlands (top) and larger (re)development areas only (bottom).

6. Conclusion

In this research, a hedonic pricing model is constructed to calculate a high-resolution housing price indices for the Netherlands. The results are used to create maps that show the potential revenue of different types of housing for each hectare in the country. In this report, we distinguished between single and multi-family housing and presented results for a typical housing type for each of them based on regional average conditions.

Three methods (specifications) were used, in the first municipal fixed effects were used, in the second 4-digit postcode fixed effects and in the third no region-specific fixed effects are included, but instead, spatial variables capture regional variation. In the first and second specification, the estimated region coefficients were explained in a separate regression using the same spatial variables that were also included in the third specification. The estimated coefficients from the regressions were subsequently used to calculate the potential revenue for single and multi-family housing units for each location in the Netherlands. Using region averages of the structural characteristics of the corresponding regions and the local values of the spatial variables. Specification 3 seems to show the most plausible range in housing values and appears to better capture spatial variation.

The results of this research offer a contribution to an overarching research project at PBL that develops a model to assess the local profitably of different housing development. The potential revenues calculated in this research are an important ingredient for this assessment. An obvious next step would be to compare the obtained revenues from this study with the development costs of different development options. These development options can be based on the housing composition classes of recently (re)developed areas documented in this study.

Obviously, the potential revenues for different urban development strategies that will result from the approach suggested in this report are stylised representations of potential future developments. They should only be used to sketch the relative profitability of hypothetical development options and not be seen as exact predictions of future changes. One limitation in the suggested approach relates to the definition of development strategies based on the average housing compositions presented in Section 5. The national averages used to describe potential housing densities per hectare presented may be less applicable to specific areas. In large urban areas, for example, the higher densities will prevail, while the lowest densities are characteristic of more rural areas. So, subsequent simulation efforts will have to strike a balance between a manageable number of development options (preferred from a modelling perspective) and realistic representation of regional variation. Ideally, development alternatives are defined at the regional (say housing market) level, based on the observed types of residential development that are characteristic of distinguished regions. But this may quickly result in a wide variety of development options to be included in the modelling framework that may obscure underlying narrative of the different urban development strategies.

Another concern relates to the fairly high (or low) housing prices for specific regions found in this study. Typical examples are the high values in affluent municipalities such as Wassenaar, Bloemendaal and Rozendaal (Gelderland) that result from a combination of a very specific housing stock (e.g. very large villas) and regional conditions (high average income). While these high values correspond to the current situation in these areas, they may be less relevant for future urban development options, as they are dependent on very specific conditions that are not easily replicated in future developments. Moreover, it is unlikely that many new houses can be added to these areas that would yield the same revenues. These concerns suggest that it is best to focus the definition of urban development options on combinations of housing types and densities that are commonly observed and to apply these on the majority of regions that have fairly standard characteristics.

Other improvements that could be pursued in future research include extending the current set of explanatory variables. Potential extensions include, for example, incorporating residential density using the Floor Space Index data PBL is currently preparing, proximity to natural amenities such as parks and water bodies or reference to negative externalities such as polluting industries, noise contours et cetera. Another tempting idea for the simulation framework is to make specific explanatory variables dynamic. The presence of amenities or shares of specific land-use types within regions can be updated following initial land-use simulations. In this way changing spatial conditions can be captured that can be used to calculate changes in the potential revenues of locations.

7. References

Broitman, D., Koomen, E., 2015. Residential density change: Densification and urban expansion. Computers, Environment and Urban Systems 54, 32-46.

CBS, 2012. Bestand Bodemgebruik. Centraal Bureau voor de Statistiek, Voorburg.

Koomen, E., Diogo, V., Dekkers, J.E.C., Rietveld, P., 2015. A utility-based suitability framework for integrated local scale land-use modelling. Computers, Environment and Urban Systems 50, 1-14.

Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy 82, 34-55.

Statistics Netherlands, 2012. Statline, the central database of Statistics Netherlands.

van Duinen, L., Rijken, B., Buitelaar, E., 2016. Transformatiepotentie: woningbouwmogelijkheden in de bestaande stad. PBL Planbureau voor de Leefomgeving, Den Haag, Nederland.

8. Appendices

Appendix 1

Table 10: Additional used dummy variables in the regression analysis of single-family houses

Variable	(1) Ln price	(2) Ln price	(3) Ln price
Building construction year unknown	-0.119***	-0.155***	-0.133***
	(0.00109)	(0.00101)	(0.00124)
Building construction year before 1945	-0.0329***	-0.0933***	-0.0852***
	(0.000757)	(0.000749)	(0.000882)
Building construction year between 1945-1960	-0.0871***	-0.127***	-0.115***
	(0.000887)	(0.000844)	(0.00101)
Building construction year between 1960-1970	-0.128***	-0.143***	-0.133***
	(0.000812)	(0.000783)	(0.000921)
Building construction year between 1970-1980	-0.122***	-0.120***	-0.125***
	(0.000756)	(0.000739)	(0.000856)
Building construction year between 1980-1990	-0.0825***	-0.0766***	-0.0827***
	(0.000736)	(0.000733)	(0.000836)
Building construction year between 1990-2000	-0.0194***	-0.0199***	-0.0197***
	(0.000693)	(0.000676)	(0.000788)
Building construction year after 2000	(reference)	(reference)	(reference)
Transaction year: 2000	(reference)	(reference)	(reference)
Transaction year: 2001	0.0762***	0.0775***	0.0759***
	(0.000911)	(0.000797)	(0.00105)
Transaction year: 2002	0.129***	0.131***	0.130***
	(0.000910)	(0.000797)	(0.00104)
Transaction year: 2003	0.157***	0.158***	0.158***
	(0.000913)	(0.000799)	(0.00105)
Transaction year: 2004	0.193***	0.194***	0.193***
	(0.000910)	(0.000797)	(0.00104)
Transaction year: 2005	0.229***	0.231***	0.230***
	(0.000887)	(0.000777)	(0.00102)
Transaction year: 2006	0.263***	0.266***	0.263***
	(0.000887)	(0.000777)	(0.00102)
Transaction year: 2007	0.305***	0.306***	0.305***
	(0.000891)	(0.000780)	(0.00102)
Transaction year: 2008	0.315***	0.318***	0.316***
	(0.000924)	(0.000809)	(0.00106)
Transaction year: 2009	0.273***	0.273***	0.272***
	(0.000991)	(0.000869)	(0.00114)
Transaction year: 2010	0.275***	0.273***	0.277***
	(0.000981)	(0.000859)	(0.00112)
Transaction year: 2011	0.257***	0.253***	0.259***
	(0.00101)	(0.000882)	(0.00116)
Transaction year: 2012	0.190***	0.187***	0.193***
	(0.000991)	(0.000869)	(0.00114)
Transaction year: 2013	0.151***	0.147***	0.156***
	(0.00100)	(0.000877)	(0.00115)
Transaction year: 2014	0.175***	0.171***	0.181***
	(0.000922)	(0.000808)	(0.00106)
Transaction year: 2015	0.212***	0.208***	0.216***
	(0.000892)	(0.000782)	(0.00102)
Transaction year: 2016	0.261***	0.259***	0.264***
	(0.000861)	(0.000755)	(0.000987)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

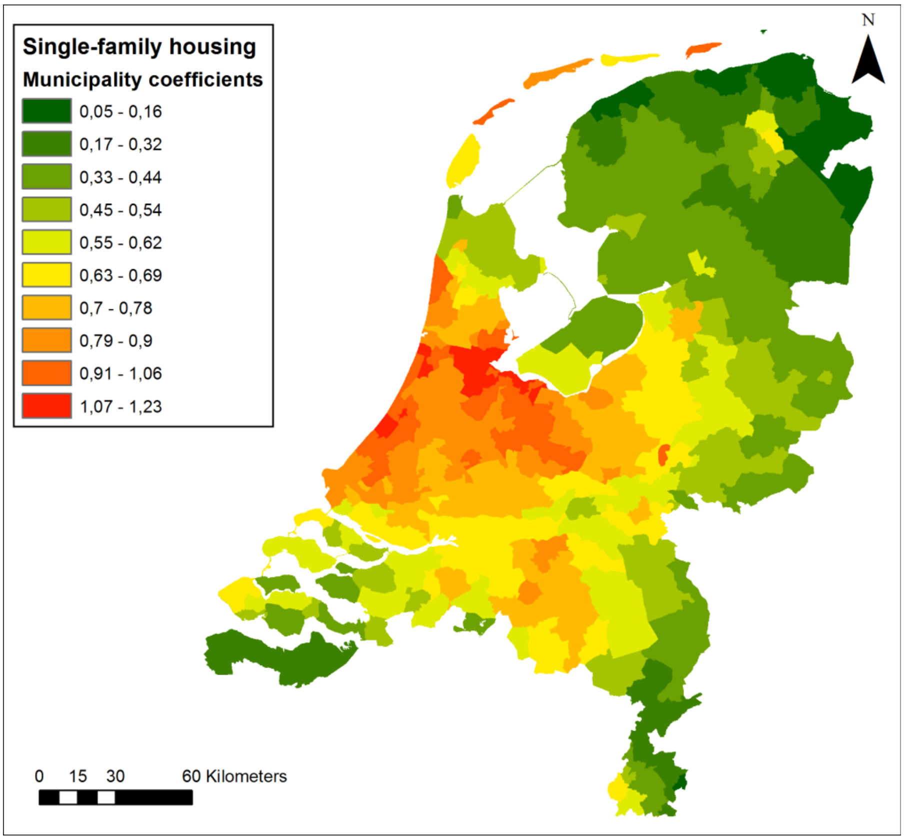
Table 11: Additional used dummy variables in the regression analysis of multi-family housing units

Variable	(1) Ln price	(2) Ln price	(3) Ln price
Building construction year unknown	0.00690***	-0.0781***	0.0228***
	(0.00126)	(0.00111)	(0.00134)
Building construction year before 1945	-0.0826***	-0.0969***	-0.209***
	(0.00137)	(0.00122)	(0.00145)
Building construction year between 1945-1960	-0.137***	-0.152***	-0.201***
	(0.00147)	(0.00131)	(0.00155)
Building construction year between 1960-1970	-0.207***	-0.189***	-0.188***
	(0.00132)	(0.00123)	(0.00140)
Building construction year between 1970-1980	-0.141***	-0.139***	-0.147***
	(0.00138)	(0.00129)	(0.00146)
Building construction year between 1980-1990	-0.0889***	-0.0940***	-0.126***
	(0.00135)	(0.00123)	(0.00144)
Building construction year between 1990-2000	0.0240***	0.0116***	0.0118***
	(0.00124)	(0.00111)	(0.00133)
Building construction year after 2000	(reference)	(reference)	(reference)
Apartment type: Downstairs	0.0739***	0.0594***	0.103***
	(0.000968)	(0.000805)	(0.00103)
Apartment type: Upstairs	0.0157***	-0.0149***	0.0556***
	(0.000845)	(0.000705)	(0.000874)
Apartment type: Downstairs and upstairs	0.0807***	0.0585***	0.0974***
	(0.00330)	(0.00270)	(0.00358)
Apartment type: Maisonette	-0.0644***	-0.0612***	-0.0473***
	(0.00112)	(0.000947)	(0.00120)
Apartment type: Service	-0.336***	-0.303***	-0.341***
	(0.00427)	(0.00359)	(0.00458)
Apartment type: Gallery	-0.0272***	-0.0234***	-0.0303***
	(0.000797)	(0.000683)	(0.000854)
Apartment type: Porch	(reference)	(reference)	(reference)
Transaction year: 2000	(reference)	(reference)	(reference)
Transaction year: 2001	0.0956***	0.0959***	0.0923***
	(0.00171)	(0.00138)	(0.00187)
Transaction year: 2002	0.157***	0.159***	0.151***
	(0.00170)	(0.00137)	(0.00185)
Transaction year: 2003	0.182***	0.183***	0.175***
	(0.00168)	(0.00136)	(0.00183)
Transaction year: 2004	0.208***	0.210***	0.202***
	(0.00165)	(0.00133)	(0.00180)
Transaction year: 2005	0.242***	0.247***	0.241***
	(0.00162)	(0.00131)	(0.00176)
Transaction year: 2006	0.281***	0.283***	0.278***
	(0.00162)	(0.00130)	(0.00176)
Transaction year: 2007	0.329***	0.332***	0.328***
	(0.00162)	(0.00130)	(0.00176)
Transaction year: 2008	0.347***	0.354***	0.346***
	(0.00166)	(0.00134)	(0.00180)
Transaction year: 2009	0.316***	0.316***	0.318***
	(0.00174)	(0.00141)	(0.00189)
Transaction year: 2010	0.311***	0.309***	0.313***
	(0.00174)	(0.00141)	(0.00190)
Transaction year: 2011	0.294***	0.288***	0.295***
	(0.00179)	(0.00144)	(0.00194)
Transaction year: 2012	0.231***	0.225***	0.237***
	(0.00179)	(0.00145)	(0.00195)
Transaction year: 2013	0.187***	0.178***	0.195***

	(0.00179)	(0.00145)	(0.00195)
Transaction year: 2014	0.217***	0.211***	0.229***
	(0.00166)	(0.00134)	(0.00180)
Transaction year: 2015	0.253***	0.252***	0.265***
	(0.00161)	(0.00130)	(0.00175)
Transaction year: 2016	0.304***	0.310***	0.309***
	(0.00158)	(0.00128)	(0.00172)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix 2



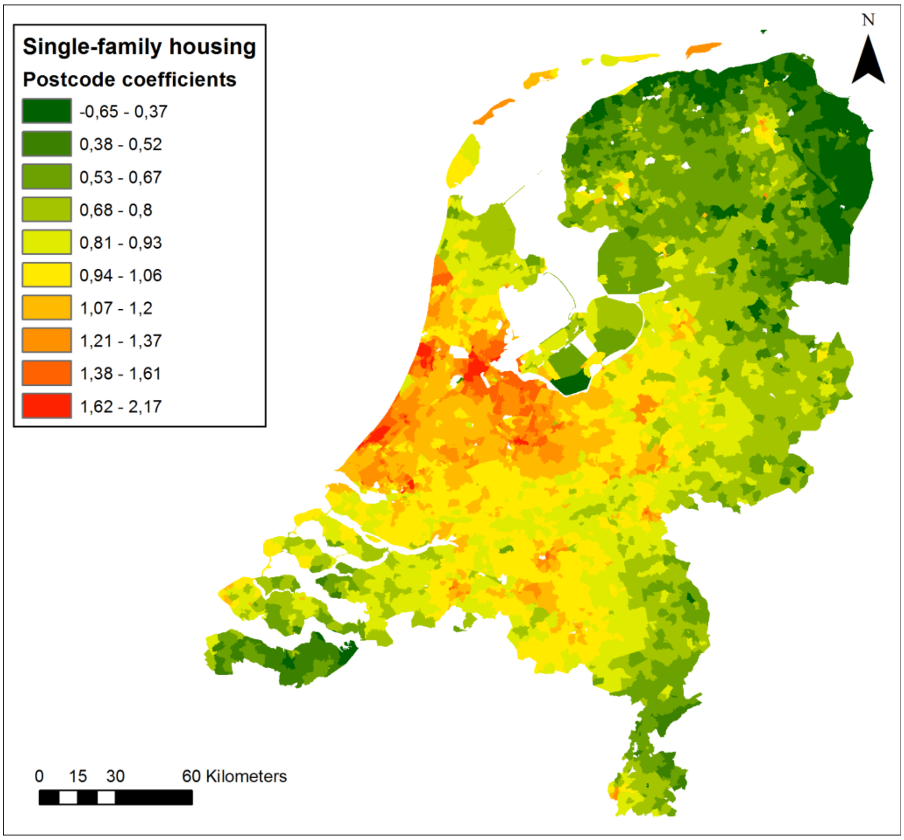
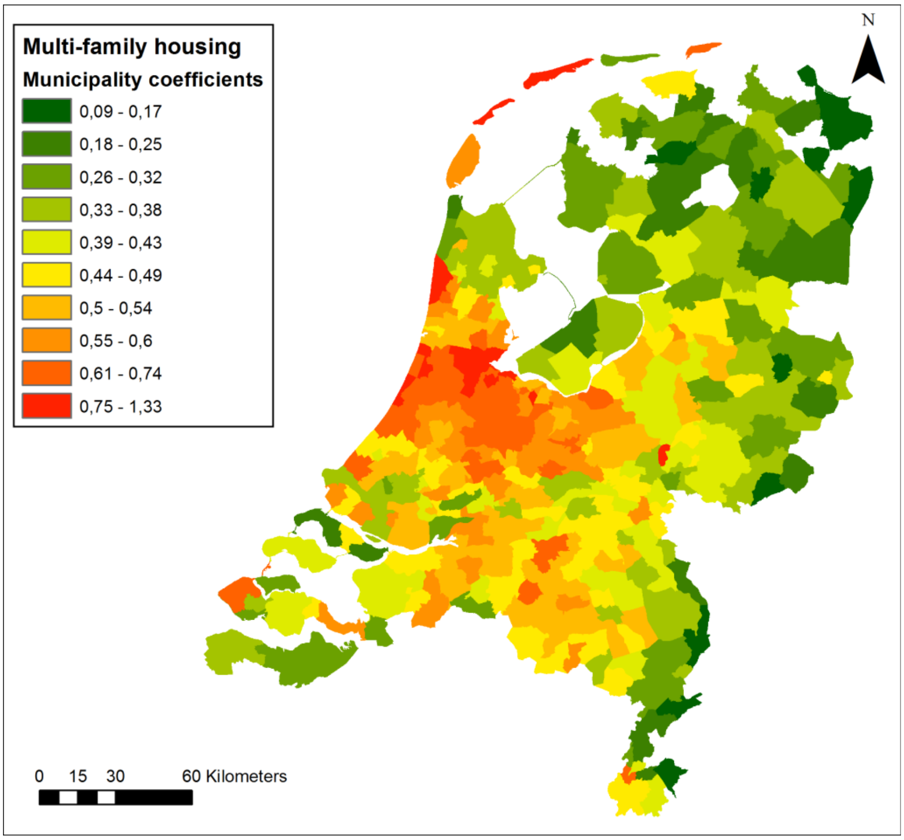


Figure 11: Single-family region-specific coefficients as estimated in equation 1. Undefined areas indicate coefficients that were not statistically significant at a 5% level. The reference category is the municipality of Pekela in the eastern part of Groningen and the 4-digit postcode area 8507 in Friesland.



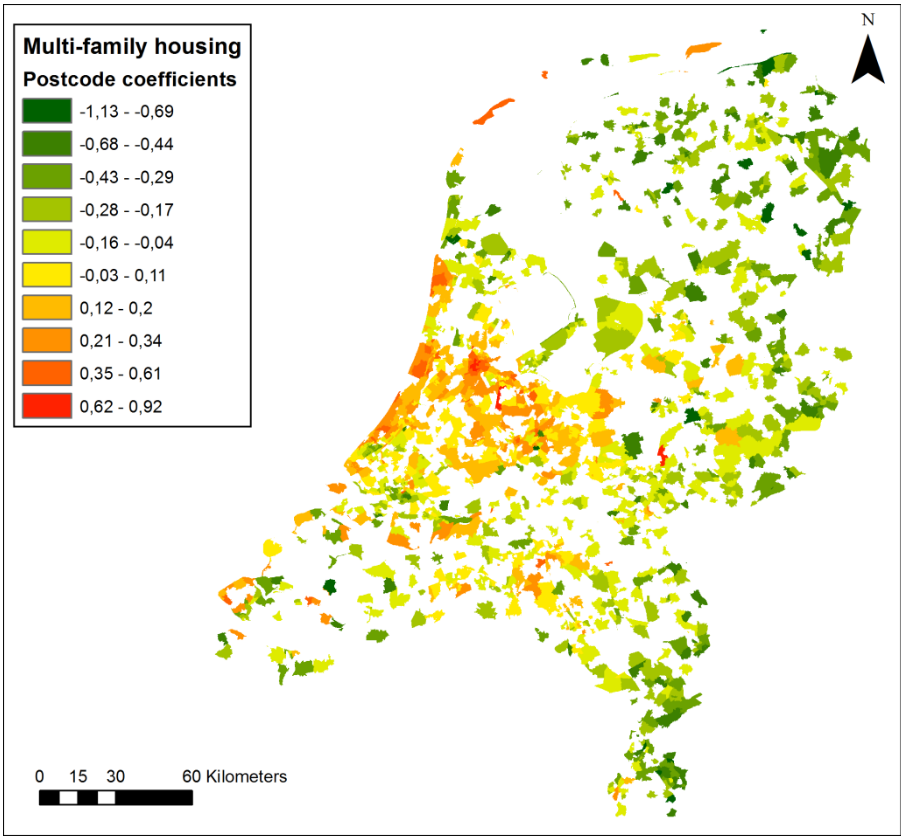


Figure 12: Multi-family region-specific coefficients as estimated in equation 1. Undefined areas indicate coefficients that were not statistically significant at a 5% level. The reference category is the municipality of Pekela in the eastern part of Groningen and the 4-digit postcode area 8507 in Friesland.