

RETAX

Het schatten van de verkoopprijs van woningen, verdere verfijningen

Frank Vastmans

Voor meer informatie over deze publicatie: frank.vastmans@kuleuven.be

© 2020 RETAX

RETAX is een Strategisch Basis Onderzoek (SBO, S005718N) gefinancierd door het Fonds voor Wetenschappelijk Onderzoek (FWO).

Niets uit deze uitgave mag worden verveelvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze ook, zonder voorafgaande schriftelijke toestemming van de uitgever.

No part of this book may be reproduced in any form, by mimeograph, film or any other means, without permission in writing from the publisher.

Deze publicatie is ook beschikbaar via www.retax.be

INHOUD

| | |
|--|-----------|
| Inleiding | 4 |
| 1. Locatie en Geo-Info | 6 |
| 1.1 Geo-informatie & Databronnen | 7 |
| 1.2 Resultaten | 7 |
| 1.2.2 De KNN-methode (nabije buren) en geografische informatie: substituten of complementen? | 11 |
| 2. Kwaliteitskenmerken op basis van de EPC-databank | 12 |
| 2.1 Resultaten Schildermans en Damen | 12 |
| 2.2 De EPC -score en de onderliggende databanken | 12 |
| 2.3 Databronnen en appartementen | 14 |
| 2.3.1 Het koppelen van gegevens van appartementen | 14 |
| 2.3.2 Gemeenschappelijke delen | 14 |
| 2.3.3 Koppelverkoop bij appartementen | 15 |
| 2.3.4 Niet-residentieel | 15 |
| 2.4 Bespreking EPC binnen België | 15 |
| 2.5 Energieprestatie-kenmerken van woningen in Vlaanderen | 19 |
| 2.5.1 Energetische variabelen | 19 |
| 2.5.2 Algemene woningkenmerken vanuit de EPC-databank | 19 |
| 2.5.3 Transactie info | 20 |
| 2.6 De aanwezigheid van een EPC | 20 |
| 3. Evolutie van woonvoorkeuren doorheen de tijd | 23 |
| Conclusie | 27 |
| Bijlagen | 29 |
| Referenties | 30 |

INLEIDING

Het kadastraal inkomen (KI) is een schatting van het jaarlijkse netto huurinkomen van het onroerend goed. De schattingen waarop het KI gebaseerd is dateren echter van 1975, maar de huurwaarde van een onroerend goed kan doorheen de tijd drastisch veranderen. Boogaerts et al. (2020) toonden aan dat woningen met een vergelijkbare verkoop- of huurprijs sterk verschillende KI's kunnen hebben. Een vervanging van het kadastrale inkomen door een belastbare basis van de actuele waarde dringt zich op.

Schildermans en Damen (2021) gingen daartoe na welke statistische modellen het meest accuraat zijn in het schatten van de verkoopprijs van onroerende goederen op basis van de data die momenteel voor de overheid beschikbaar zijn. Ze deden dit voor huizen, appartementen en bouwgronden afzonderlijk. Een methode die goede resultaten gaf voor huizen en appartementen (relatief goede predicties met als voordeel ook interpreteerbare coëfficiënten) was een uitbreiding van de loglineaire hedonische prijsmethode met de K-Nearest-Neighbor methode (KNN), waarbij informatie van de dichtstbijgelegen observaties het locatie-effect bepalen. In wat volgt verwijzen we naar KNN-methode om te duiden dat we een loglineaire prijsmethode schatten met de KNN-uitbreiding.

In dit rapport worden verkoopprijzen van woonhuizen geschat via de KNN-methode zoals gebruikt bij Schildermans en Damen, maar met licht verschillende modellen en verfijningen. Daar waar Schildermans en Damen de performantie vergeleken tussen verschillende methodes¹ wordt er hier dus nagegaan wat de verschillende opties zijn van modelleren binnen de loglineaire hedonische prijsanalyse uitgebreid met de KNN-methode. We testen onze analyses op de verkopen in Vlaanderen waarvoor een EPC beschikbaar is (in 2.6 wordt uitgelegd waarom). En bovendien verwijderen we observaties waarvan de schattingsfout uitzonderlijk hoog is² omdat deze uitschieters een vertekend effect kunnen hebben op de schatting van coëfficiënten, in lijn met de best practice richtlijnen van Eurostat (2017). Als criterium om de verschillende modellen te vergelijken gebruiken we de R^2 die de verklaarde variantie weergeeft.

De waarde van een woning wordt in grote lijnen bepaald door drie klasse van variabelen: de kenmerken van de woning op zich, de locatie van de woning, en het tijdstip van de waardebepaling. Deze drie elementen zijn ook de standaardelementen die we terugvinden in een hedonische prijsanalyse. In dit rapport wordt voor elk van deze drie elementen nagegaan hoe de analyse van Schildermans en Damen (2021) verfijnd kan worden, waarbij de nadruk ligt op een bespreking van de voor- en nadelen van dergelijke verfijningen. Het opzet is immers niet altijd zozeer om de beste predictie te krijgen, maar ook een model te ontwikkelen dat eenvoudig aanpasbaar is, overdraagbaar, gebruiksvriendelijk en interpreteerbaar.

Mayer et al (2019) wezen erop dat het voordeel van de betere predictie dient afgewogen te worden ten opzichte van het feit dat ze meer volatiele resultaten geven. *“There is a clear trade-off across methods depending on whether the goal is to improve accuracy or avoid volatility.”* Dit gegeven lijkt op het eerste zicht contradictorisch omdat een betere predictie vaak leidt tot minder volatiliteit. Maar complexere modellen bestaan vaak uit een set van verschillende deelpredictiemodellen. Het algemeen model selecteert dus het submodel dat het best schat voor een bepaalde set van observaties. Dit leidt

¹ Een belangrijk element om de performatie van verschillende methoden te vergelijken is out-of-sample testen. Dit element speelt in dit rapport bijna geen rol aangezien telkens dezelfde methode gebruikt wordt en enkel een paar variabelen worden toegevoegd, die zeer beperkt in aantal zijn in vergelijking met het aantal observaties. Bij Schildermans en Damen (2021) was cross-validatie via out-of-sample testen wel belangrijk aangezien complexere methoden een uitgebreide set van verschillende modelleringen testen wat leidt tot het gevaar van overfitting.

² studentised residuals buiten het interval [-2.2]

ertoe dat bij het toevoegen van nieuwe observaties doorheen de tijd, deze modellen ook in grotere mate andere opties zullen kiezen waardoor de prijsschattingen van dezelfde woningen dus sterker fluctueren aangezien ze zijn gebaseerd op een ander model. Het is echter belangrijk dat de schattingen van de belastbare basis stabiele resultaten doorheen de tijd geeft. Sterke fluctuaties doorheen de tijd zijn daarom te vermijden.

Het algemene loglineaire predictiemodel kan geschreven worden in volgende vorm:

$$\log(y_i) = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip} + e_i$$

Aan de linkerkant van deze vergelijking staat y_i , die de numerieke variabele weergeeft die we gaan trachten te schatten. Het subscript i wijst erop dat het in dit geval gaat over de uitkomst voor observatie i of in de context van dit rapport de verkoopprijs van woning i . Aan de rechterkant van de vergelijking vinden we de predictor-variabelen terug. b_0 is het zogenaamde intercept. Deze variabele wordt geschat en is gelijk voor alle observaties, vandaar wordt het ook wel de constante genoemd. Daarna volgen een hele reeks combinaties van b 's en x 's. De x 's zijn hier de verschillende verklarende variabelen die gebruikt zullen worden om de uitkomst te schatten. De b 's zijn de coëfficiënten van deze variabelen en bepalen in welke mate de variabelen de prijs bepalen. De b 's hebben in deze vergelijking geen subscript i omdat zij verondersteld worden gelijk te zijn voor alle woningen. Er bestaan echter ook modellen waarbij de coëfficiënten niet gelijk zijn voor alle woningen. Zo kan het zijn dat de waarde van een garage groter is in stedelijk gebied, dus verschilt naargelang de locatie. Het kan ook zijn dat de waarde van een coëfficiënt verandert doorheen de tijd. Zo kunnen we nagaan indien sinds COVID-19 woonvoorkeuren veranderen en de coëfficiënten – als indicatie van woonvoorkeur- voor grotere woningen sterker gestegen zijn ten opzichte van kleinere woningen.

Meer algemeen kan bovenstaande vergelijking uitgeschreven worden in de belangrijkste deelaspecten die de waarde van de woning bepalen, namelijk grootte, kwaliteit, tijdstip en locatie.

$$\log(\text{prijs}_i) = b_0 + b_g G_i \text{grootte} + b_k K_i \text{kwaliteit} + b_t t_i \text{tijdsstip} + b_l L_i \text{locatie} + e_i$$

Elk van deze deelaspecten wordt verder onderzocht.

In het eerste hoofdstuk bespreken we verdere verfijningen qua **locatie**. bv. door het toevoegen van geografische informatie, waar de KNN methode geen gebruik van maakte. De hamvraag in dit hoofdstuk is dus of geografische informatie expliciet mee opgenomen dient te worden, of dat de KNN-methode volstaat om de locatie-effecten in kaart te brengen.

In het tweede hoofdstuk wordt verder gekeken naar de **kwaliteitskenmerken** van de woning. In welke mate verbeteren de predicties door het toevoegen van extra variabelen (zoals energetische kenmerken op basis van de EPC databank). Daarnaast wordt er ook gekeken hoe variabelen gemodelleerd kunnen worden. Hierbij wordt er nagegaan in welke mate de set woningkenmerken zoals die door AAPD gekend zijn volstaan als variabelen.

In het derde hoofdstuk kijken we naar de **evolutie doorheen de tijd**. Hierbij kijken we niet enkel naar de algemene hedonische prijsindex, die de algemene prijsevolutie van een woning weergeeft, maar gaan we na hoe de geschatte coëfficiënten voor diverse woningkenmerken anders evolueren doorheen de tijd. Deze analyse illustreert hoe de voorkeuren (willingness to pay) voor woningkenmerken en locaties veranderen wat o.a. belangrijk is om de frequentie van updates van het model te bepalen.

1. LOCATIE EN GEO-INFO

De locatie van een woning heeft een belangrijk aandeel in de verklaring van een woningprijs. Het is dan ook belangrijk om de waarde van deze locatie goed in te schatten. Ruwweg kan het locatie-effect op twee manieren bepaald worden: op basis van de verkoopprijzen binnen een buurt, of op basis van geografische kenmerken van de locatie zelf, waarbij elk geografische kenmerk afzonderlijk (aanwezigheid van groen, voorzieningen, erfgoed....) een waardering krijgt.

De eerste manier gaat op basis van de verkoopprijzen bepalen in welke locaties de woningprijzen duurder zijn dan elders. Een vaak gebruikte methode is het toevoegen van gemeenteddummies aan het hedonische prijsmodel. De coëfficiënten van deze gemeenteddummies geven dan weer in welke mate eenzelfde woning in prijs verschilt afhankelijk van de gemeente waarin de woning gelegen is. Schildermans en Damen (2021) wezen er al op dat deze geografische indeling niet zomaar kan verkleind worden om tot meer regionaal gedifferentieerde schattingen te komen. Zo zijn er vaak niet voldoende observaties beschikbaar binnen een statistische sector, en al zeker niet op straatniveau. Maar indien er voldoende observaties beschikbaar zijn binnen een bepaalde statistische sector, dan kan men uiteraard best een dummy-variabele toevoegen die het locatie-effect bepaalt binnen die statistische sector, een methode die Vastmans (2019) voor de hedonische huurprijsanalyse. Er wordt een locatie-effect geschat voor een zo klein mogelijke gebiedsindeling op voorwaarde dat er voldoende observaties zijn. De gebiedsindeling is dus soms een gemeente, soms een statistische sector, of gebiedsindelingen daar tussenin, afhankelijk van het aantal observaties.

De K-Nearest Neighbor (KNN) methode hanteert een gelijkaardig idee, alleen op continue schaal en flexibele wijze, waarbij voor elke adres afzonderlijk het locatie-effect bepaald wordt op basis van de k meest nabij gelegen verkopen, en dit dan volgens een gewogen gemiddelde, waarbij de dichterbijgelegen verkopen meer bepalend zijn. Deze methode zorgt er dus automatisch voor dat er voldoende observaties zijn (namelijk k -aantal) wat er toe leidt dat in buurten met veel verkopen het locatie-effect gebaseerd zal zijn op een zeer lokale markt in een perimeter van enkele honderden meters, terwijl dit in rurale gebieden dan weer breder zal zijn. Het basisidee is bovendien eenvoudig: Als het model in een eerste stap vindt dat alle woningen in de buurt $x\%$ te laag geschat zijn, dan wordt de predictie met $x\%$ verhoogt (out-of-sample). Welke buurtkenmerken dit verklaren is niet belangrijk, enkel de extra predictiekracht telt.

De tweede benadering vertrekt net wel van de buurtkenmerken. Zo kan de afstand tot een treinstation in het hedonisch prijsmodel mee opgenomen worden. Het grote voordeel is dat dit effect gebiedsdekkend gemeten wordt, en bovendien ook gebruikt kan worden om de waardering van de afstand tot een treinstation te bepalen in gebieden waar nog geen of onvoldoende verkopen waren om een effect te bepalen. Men kan dus het effect van de nabijheid tot een treinstation bepalen in één regio op basis van verkopen in andere regio's. Dit laatste voordeel heeft de KNN-methode niet. Dit voordeel vormt echter ook een nadeel. Niet elk treinstation is hetzelfde. En zelfs al zou het treinstation hetzelfde zijn, dan nog kan men verwachten dat de waardering ervan regionaal verschilt. Damen (2017) formuleerde dit als volgt in een onderzoek naar de effecten van de erfgoedkarakteristieken van woningen en hun omgeving op de marktprijzen van woningen in Vlaanderen: *"The use of this model requires caution and background knowledge, given that it is a rough average which is therefore not applicable in a specific case. In other words, simply applying the results of this general model could result in a strong distortion of reality."* En om het effect van elk geografische kenmerk regionaal verschillend te meten zijn er vaak te weinig observaties. Bovendien kan men verwachten dat de KNN-methode reeds in grote mate deze locatie-effecten capteert, en wel op een manier die de som van al deze onderliggende geografische kenmerken en hun onderlingende samenhang capteert voor die locatie als geheel.

1.1 Geo-informatie & Databronnen

Er is een toename aan geografische informatie, die veelal gebiedsdekkend beschikbaar is. En aangezien de waarde van een onroerend goed in grote mate door de locatie bepaald wordt zal dit er ook toe leiden dat meer geografische informatie verwerkt kan worden. Via de KNN- methode wordt het locatie-effect bepaald door de eerdere verkopen. Het KNN-locatie-effect wordt dus bepaald door hoeveel er in het verleden meer of minder betaald werd in een bepaalde locatie en maakt geen gebruik van geografische informatie. Daartoe gaan we in dit hoofdstuk na hoe bijkomende geografische informatie de verklaringskracht van het model kan verbeteren.

Het verwerken van geografische informatie bleek echter minder evident dan verwacht. In België worden twee verschillende referentie-coördinatensystemen gebruikt: Lambert 1972 en Lambert 2008. Het kadaster is gedigitaliseerd en beschikbaar in Lambert 2008, maar hier zitten nog onnauwkeurigheden in. Het Belgische Nationaal Geografisch Instituut (NGI) gebruikt eveneens Lambert 2008. Dit nieuw stelsel levert onder meer voordeel op voor organisaties die hun data-inwinningsprocessen op GPS-metingen baseren. Het betekent dat de zeer hoge nauwkeurigheid van GPS-metingen bewaard kan blijven bij de voorstelling van de resultaten op een kaart. Bij de overgang naar het Lambert 72-coördinatenstelsel ondergaan GPS-metingen een datumtransformatie met een bijbehorende (kleine) afwijking in x- en y-coördinaten op de kaart als gevolg. De meeste kaarten zijn echter enkel beschikbaar in Lambert 1972³. Dit zorgt ervoor dat het geografisch kaartmateriaal twee verschillende coördinatiesystemen bevat: de kadasterdata in Lambert 2008 en het overig kaartmateriaal in Lambert 1972. Op die manier kan men wel nog accuraat genoeg x,y coördinaten met elkaar vergelijken, maar zijn andere GIS-toepassingen niet mogelijk aangezien men niet kan meten of een punt op een lijn ligt, en dergelijke meer. Een perceel in Lambert 2008 coördinaten kan met andere woorden aan een straat liggen, in de straat liggen of net een paar centimeter van de straat verwijderd zijn. Op die manier kan men niet zomaar berekenen wat de perceelbreedte aan de straatkant is, aangezien hiervoor kaartmateriaal gecombineerd dient te worden met verschillende coördinatiesystemen.

1.2 Resultaten

De hypothese is dat het toevoegen van additionele geografische variabelen, die de waarde van de buurt kunnen verklaren, de predicties van het model niet sterk gaan verbeteren omdat de waarde van de buurt reeds in grote mate door KNN gecapteerd is.

Onderstaande tabel toont dat dit inderdaad zo is. De toevoeging van de nearest neighbors (KNN) leidt tot de grootste toename van de verklaarde variantie (R^2). Het toevoegen van bijkomende variabelen heeft slechts een beperkte incrementele meerwaarde. We bespreken kort de verschillende kenmerken die we hebben toegevoegd.

³ <https://www.geopunt.be/voor-experts/lambert-2008>, Aanbeveling gebruik coördinatensystemen (geopunt.be)

Tabel 1 Toename verklaarde variantie (R²) van geografische info in vergelijking met basismodel met KNN (predictie huisprijzen in Vlaanderen 2012-2021)

| TYPE verfijning | R ² + | specifiek |
|--|------------------|--|
| Baseline zonder KNN | 79% | Totale R ² (inclusief EPC-variabelen) |
| Baseline met KNN | + 3.04% | KNN voegt veruit meeste predictiekracht toe. |
| bouwwolume | + 0.11% | Geografische info (DHMV+GRB) maar geen locatie-effect |
| straten | + 0.08% | Type straat, breedte, éénrichtingsverkeer,... |
| overstromingsgevoelig | + 0.07% | -1.3% volledig perceel effectief overstromingsgevoelig (potentieel overstromingsgevoelig -0.4%) |
| knooppuntwaarden & voorzieningenniveau | + 0.07% | voorzieningen belangrijker dan knooppunten |
| buren | + 0.03% | +0.3% indien huizen in de buurt groter zijn - 0.8% indien de huizen in de buurt kleiner zijn. |
| Studentenhuis | + 0.00% | +7.4% voor een huis in Leuven voor studentenhuisvesting (tov van eenzelfde regulier huis) |

Bron: AAPD verkoopprijzen van huizen, KNN-effecten Schildermans en Damen (2021), eigen regressie

1.2.1.1 Bouwvolume

In het volgend hoofdstuk vinden we dat het beschermd bouwvolume volgens de EPC-databank een extra meerwaarde geeft aan het model. De EPC-gegevens zijn echter niet voor heel de woningvoorraad beschikbaar. Op basis van geografische informatie kan eveneens een bouwvolume berekend worden. Hiertoe werd het Digitaal Hoogtemodel voor Vlaanderen (DHMV)⁴ gebruikt en het Grootschalig Referentiebestand Vlaanderen. Door beiden te combineren kan een bouwvolume geschat worden. We vinden dat dit bouwvolume extra voorspelkracht heeft bovenop de bestaande woningkenmerken. Deze capteert de combinatie van de oppervlakte en de hoogte van de kamers. De gemiddelde hoogte in een woning is dan ook een belangrijke variabele die niet in het kadaster beschikbaar is. Maar mogelijk biedt het bouwvolume ook een verdere verfijning van de oppervlakte. Men kan verwachten dat voor een deel van de huizen in de woningvoorraad de bestaande ruimte is vergroot via een uitbouw waarvoor niet altijd een vergunning is aangevraagd hoewel dit wettelijk wel zou moeten. Daarnaast zijn er ook bijbouwen zoals veranda's waarvoor niet altijd een vergunning dient aangevraagd te worden. Het is alvast opmerkelijk dat deze variabele de verklaarde variantie het meest doet toenemen. Op zich is het geen geografische variabele die de waarde van een locatie bepaald. Het is een woningkenmerk dat wordt afgeleid op basis van geografische informatie.

1.2.1.2 Straten

Het Adressenregister⁵ voorziet de adresseerbare objecten in het Gebouwenregister van adressen en de wegen in het Wegenregister van straatnamen. Op basis hiervan kan een link gelegd worden tussen de gebouwen en percelen enerzijds en de kenmerken van de wegen. Het wegenregister bevat informatie over de wegbreedte, het aantal rijstroken, éénrichtingsverkeer,.... Door deze informatie te integreren verbeterde de R² met 0,08%.

⁴ <https://overheid.vlaanderen.be/informatie-vlaanderen/producten-diensten/digitaal-hoogtemodel-dhmv>

⁵ <https://www.vlaanderen.be/digitaal-vlaanderen/onze-oplossingen/gebouwen-en-adressenregister>

1.2.1.3 Overstromingsgebied

De kaart van de overstromingsgevoelige gebieden wordt gebruikt in het kader van de watertoets en voor de informatieplicht in het kader van de verkoop of verhuur van overstromingsgevoelig vastgoed. Deze informatie (al dan niet gekend via de informatieplicht bij verkoop) heeft een duidelijke prijsimpact.

Woningen gelegen in effectief overstromingsgevoelige gebieden hebben gemiddeld een 1,3% lagere verkoopprijs. Dit zijn de recent overstromde gebieden (ROG), gecorrigeerd op basis van de hoogteligging, aangevuld met de gemodelleerde overstromingsgebieden (MOG) met middelgrote kans (d.w.z. een herhalingsperiode van 100 jaar)⁶. Mogelijk overstromingsgevoelige gebieden leiden tot een 0,4% lagere verkoopprijs voor de woningen die in deze gebieden gelegen zijn. Dit zijn de van nature overstroombare gebieden (NOG) met uitzondering van de zones die al geruime tijd (sinds de jaren '70 of eerder) bebouwd zijn.

Daarnaast bakende de Federale regering risicozones voor overstromingen af in het kader van de natuurrampenverzekering. De risicozones bestaan uit de omhullende contouren van de gemodelleerde overstromingsgebieden (MOG) met een terugkeerperiode van 25 jaar en de recent overstroomde gebieden (ROG) die gedurende de laatste 10 jaar meer dan tweemaal overstroomd zijn geweest. Deze kleinere zones zijn niet onderzocht, maar hebben mogelijk wel een nog grotere prijsimpact.

1.2.1.4 Knooppuntwaarden en voorzieningenniveau

Geografische informatie omtrent de knooppuntwaarden en de nabijheid van voorzieningen vertonen een sterk verband met de algemene liggingseffecten. De knooppuntwaarde wordt onder meer bepaald door de transportmodus (trein, metro, tram, bus), de frequentie van het vervoersaanbod, de vervoerscapaciteit, aansluitings- en (multimodale) overstapmogelijkheden en aansluiting op fiets- en wandelnetwerken. Voor het voorzieningenniveau wordt een combinatie gemaakt van basisvoorzieningen, regionale en metropolitane voorzieningen.

Bij de knooppunten worden wel niet de autosnelwegen mee opgenomen. Helgers (2016) vond nochtans dat deze zeer belangrijk zijn. Huizen zijn hoger gewaardeerd indien de reistijd tot op- en afritten van autosnelwegen korter is. Tegelijkertijd verkiest men ook dat de huizen zelf niet dichtbij een autosnelweg liggen. De verschillende indicatoren m.b.t. publiek transport waren evenzeer positief gecorreleerd met woningprijzen, maar in mindere mate. Bovendien vond hij dat de resultaten konden verschillen naargelang het woningtype (huis/appartement) en het eigenaarsstatuut (eigenaar/huurder).

Aangezien enkel de knooppunten opgenomen werden (dus variabelen mbt tot openbaar vervoer) is de meerwaarde hiervan beperkter in het model. Tesaamen met voorzieningenniveaus stijgt de R^2 met 0,07%, waarvan de voorzieningen belangrijker zijn dan de knooppunten. Op zich is dit resultaat dus niet verwonderlijk. Knooppunten spelen op een breder regionaal niveau een belangrijke rol, wat in ons model reeds in de gemeentedummies en de nearest neighbors gecapteerd is.

1.2.1.5 Buren, option value en bouwvoorschriften

Helgers (2016) onderzocht hoe de woningkenmerken van de buurt de woningprijs beïnvloeden. Hij vond dat naast de oppervlakte van de woning zelf, ook de oppervlakte van de woningen in de buurt een gelijkaardig effect op de woningprijs hadden. Hoe groter de woningen in de buurt, hoe hoger de waarde van de woning. In deze analyse vonden we een gelijkaardig effect: Indien de huizen in de buurt groter zijn dan de verkochte woning, dan gaat dit samen met een 0,3% hogere woningprijs. Zijn de

⁶ <https://www.integraalwaterbeleid.be/nl/beleidsinstrumenten/watertoets/overstromingskaarten>

huizen in de buurt kleiner, dan leidt dit tot een -0,8% lagere woningprijs. Door het opnemen van dit buurteffect verbeterde de verklaarde variantie met 0,03%-punt.

Het is bovendien interessant om stil te staan bij de mogelijke verklaringen van dit resultaat, en dan de vooral de *option value*. Clapp e.a. (2013) geeft aan dat de option value een positieve functie is van niet-gerealiseerd bouwpotentieel van een stuk grond, gemeten door het maximaal toegelaten bouwvolume te vergelijken met het bestaand bouwvolume.

Het maximaal bouwvolume van een woning is echter niet digitaal beschikbaar. Het wordt bepaald door de bouwvoorschriften van de gemeente en deze kunnen bovendien variëren doorheen de tijd, slechts tendele vastgelegd zijn, etc. Meer in het algemeen, en zeker ook voor bouwgronden (welke in dit rapport niet genalyseerd worden), zou het interessant zijn om informatie digitaal en gebiedsdekkend beschikbaar te hebben van de bouwvoorschriften en de mogelijke maximale bouw mogelijkheden.

Er kan echter een benadering uitgewerkt worden. Aangezien de naburige burens veelal gelijkaardige bouwvoorschriften hebben, kan men op deze manier een indicatie krijgen van de *option value*. Indien de huizen in de buurt groter zijn dan het huis in kwestie, wijst dit mogelijk op extra bouwpotentieel van de woning zelf, terwijl een huis dat groter is dan dit van de burens geen extra bouwpotentieel heeft. Daarnaast is het zo dat huishoudens soms zonder vergunning diverse werken hebben uitgevoerd zoals een bijbouw of het bewoonbaar inrichten van de zolder. Het zou er dus ook op kunnen wijzen dat het werkelijk bouwvolume nu reeds groter is dan het officieel vergund bouwvolume met bijhorende administratieve woningkenmerken.

1.2.1.6 Studentenhuisen

De toename van hoger onderwijs heeft ervoor gezorgd dat vele jongeren hun woningzoektocht op latere leeftijd in mindere mate starten vanuit de ouderlijke woonplaats, en in grotere mate vanuit de studentenstad. Dit is één van de verklaringen van de shift van de vraag van jonge kopende starters naar meer centraal gelegen woningen en de daarbij horende prijsstijgingen in centrumsteden (Vastmans en Dreesen 2021).

Maar wereldwijd is in vele steden studentenhuisvesting op zich een uitdaging (Rifaad et al 2021). De vraag is dan ook indien men op een gelijkaardige manier de marktwaarde kan bepalen van de woningen voor studentenhuisvesting als deze voor de algemene woningvoorraad. In het kadaster is echter geen aparte indeling voor studentenhuisvesting voorzien. Studentenkamers binnen een appartementsgebouw worden soms opgenomen onder 104 en 114 (diverse lokalen, apart gekadastreerd binnen een gebouw), maar soms ook als wooneenheid. Maar naast de typische studentenhomes (purpose built student accommodation, PBSA), zijn ook de typische studentenhuisen met 'studentenkoten' (Housing in Multiple Occupation, HMO). In Leuven kan hier onderscheid gemaakt worden tussen een gezinswoning en een studentenhuis. Verschillende studentenhuisen waren oorspronkelijk huizen voor gezinnen, maar hebben in het verleden het statuut van studentenhuis verworven. Momenteel kunnen huizen binnen de reguliere woningvoorraad niet meer eenvoudig "herbestemd" worden naar een studentenhuis. Het aanbod van nieuwe studentenvoorzieningen komt voornamelijk vanuit grotere PBSA ontwikkelingen. Dit zorgt ervoor dat de vroegere herbestemming naar studentenhuis leidt tot een aparte categorie van huizen, de studentenhuisen met hun studentenkoten. Doordat deze markt nu gescheiden is van de reguliere markt, gedreven door een ander soort vraag, kan men veronderstellen dat het prijsmechanisme ook anders is. Deze studentenhuisen worden inderdaad tegen een premie van 7,4% verkocht.

1.2.2 De KNN-methode (nabije buren) en geografische informatie: substituten of complementen?

Bij de bespreking van de diverse buurtvariabelen op basis van geografische informatie hebben we één element niet besproken: geografische informatie kan als een substituut werken van de KNN of als een complement. Bij de variabele knooppunten en voorzieningen werken beiden zeer sterk als substituten aangezien knooppunten en voorzieningen ruimtelijk zeer gelijkaardig werken als de nearest neighbors, namelijk op continue schaal. Hoe verder weg een woning zich bevindt van een voorziening, hoe lager de woningwaarde. Eenzelfde effect wordt gemeten via de KNN-methode.

Maar geografische informatie kan ook complementair werken aan de KNN-methode. Zo zal het men via gegevens over de straat in eenzelfde buurt een onderscheid kunnen maken indien deze woning aan een steenweg ligt, of in de woonbuurt honderd meter verder. Op basis van de nearest neighbors kan men dit verschil niet maken omdat deze een gelijkmatige (continue) verandering in de woningwaarde veronderstellen. Zo ook kan aan de ene kant van de straat de woning overstromingsgevoelig zijn, en aan de andere kant niet. Geografische informatie maakt dit lokaal onderscheid, maar de nearest neighbors niet. Ook bij het buurt-effect zit een asymmetrie: het buurteffect geldt niet voor alle woningen, enkel de kleinere zijn meer waard, de grotere net minder. De informatie omtrent studentenhuzen werkt eveneens complementair.

De KNN-methode kan dus nog op verschillende manieren verfijnd worden. Om te tonen hoe bovenstaande opdeling tussen substituten of complimenten verwerkt kan worden, hernemen we de regressie-vergelijking uit de inleiding:

$$\log(\text{prijs}_i) = b_0 + b_g G_i \text{ grootte} + b_k K_i \text{ kwaliteit} + b_t t_i \text{ tijdsstip} + b_l L_i \text{ locatie} + e_i$$

De nearest neighbours kunnen bepaald worden op basis van de residuals e_i , het deel dat in een eerste stap niet verklaard kon worden, zoals Schildermans en Damen (2021) hebben toegepast, en dat uitgaat van het feit dat de extra geo-locatie variabelen ($L_i \text{ locatie}$) als complementen beschouwd worden. Maar het effect van de nearest neighbours kan eveneens bepaald worden op basis van $b_l L_i \text{ locatie} + e_i$, waarbij de locatie-variabelen die als substituten beschouwd konden worden van de KNN-methode mee in het locatie-effect opgenomen worden. De nearest neighbours hebben dan gelijkaardige locatie-effecten, niet alleen gelijkaardige residuals. In het RETAX-onderzoeksrapport waar de huurprijzen geanalyseerd worden (Vastmans te verschijnen) presteerde de KNN-methode iets beter wanneer de gemeentedummies mee opgenomen werden en zo het volledige locatie-effect, en niet enkel de residual i van de naburige buren gebruikt werd. Indien de naburige observaties binnen eenzelfde gemeente of andere discrete regionale indeling liggen zullen beide methoden gelijkaardige resultaten geven. Maar aan de grenzen van de gemeenten lijkt het dus beter om rekening te houden met het verschillend gemeente-effect van een nearest neighbour. Indien men enkel de residual neemt van een naburige observatie in een andere en duurdere gemeente zal deze mogelijks negatief zijn, aangezien de observatie van die nearest neighbor uit een duurdere gemeente dicht bij een goedkopere gemeente ligt. Indien men de error van deze observatie vervolgens neemt om de liggingscoëfficiënt van een adres in de goedkopere gemeente, maar aan de grens met die duurdere gemeente, te bepalen, dan zal dit leiden tot een lagere voorspelling. Maar veelal zal men net verwachten dat de observaties dichtbij een duurdere gemeente een hogere voorspelling krijgen. Door het totale locatie-effect van die naburige observatie in een duurdere gemeente te nemen (dus inclusief het discreet locatie-effect van de gemeente) stelt zich dit probleem niet, waardoor de resultaten verbeteren. Het illustreert alvast dat de predicties niet enkel afhangen van de gebruikte algemene methode, maar dat ook de diverse opties binnen een methode bepalend zijn voor de accuraatheid. Verder onderzoek naar andere benaderingen binnen de KNN-methode kunnen dit verder verhelderen. Het lijkt ook aannemelijk dat dit gemeente-effect slechts ten dele dient mee opgenomen te worden omdat de fiscaliteit tussen gemeenten verschilt, waardoor het locatie-effect kan verschillen aan weerskanten van de gemeentegrenzen.

2. KWALITEITSKENMERKEN OP BASIS VAN DE EPC-DATABANK

In dit deel bespreken we, na een korte samenvatting van de resultaten van de EPC-analyse van Schildermans en Damen (2021) de mogelijkheden om energie-variabelen op te nemen in het model. Informatie uit EPC databanken, die gewestelijk verschillen, is de grootste databron met gegevens over de woningkwaliteit. Daartoe bespreken we eerst deze gegevensbron. Vervolgens bespreken we de moeilijkheid om de EPC-score van appartementen aan de data te koppelen en mee in het model te brengen. De resultaten worden in drie delen opgesplitst. Vooreerst bespreken we de impact van de EPC-score op de verkoopprijs aan de hand van de resultaten van Reusens, Vastmans en Damen (2022) voor de drie gewesten en doorheen de tijd. Vervolgens gaan we na welke bijkomende energieprestatie-kenmerken van woningen een meerwaarde bieden voor het Vlaams Gewest, en in een laatste stap gaan we na waarom verkopen waarvoor geen EPC beschikbaar is, veelal “slechte” observaties zijn en dus best uit de analyse gelaten worden, ook voor analyses waarbij het EPC-kengetal zelf niet mee opgenomen wordt in het model.

2.1 Resultaten Schildermans en Damen

Schildermans en Damen (2021) vonden dat het toevoegen van de EPC-score aan het model de performantie van het model aanzienlijk verbetert: de MAE (mean absolute error of procentuele fout) daalt van 0.18 naar 0.16 voor Vlaanderen. De MAE van 0.16, uitgedrukt in het natuurlijk logaritme, komt overeen met een gemiddelde absolute schattingsfout van 37.710 euro voor een gemiddelde verkoopprijs gelijk aan 243.947 euro.

Ze testten ook hoe het model presteerde indien niet de EPC-score, die de energieprestatie per m² weergeeft, maar het geschatte theoretische energiegebruik toegevoegd werd aan het model. Deze bestaat uit de som van het geschatte energieverbruik voor ruimteverwarming, energieverbruik voor sanitair warm water, energieverbruik voor hulpenergie en energieverbruik voor koeling. Hiervan wordt dan de energiebijdrage door PV panelen en energiebijdrage door elektriciteitsproductie van een warmtekrachtkoppeling afgetrokken. Dit leidde niet tot betere resultaten. Ook de combinatie van beide variabelen (EPC-score en geschat energieverbruik) leidde niet tot een verbetering van de schattingen.

2.2 De EPC -score en de onderliggende databanken

De EPC-score, of het EPC-kengetal, is gebaseerd op een uitgebreide set van energetische kenmerken van de woning die de energiedeskundige opmeet en die in de EPC-databank zijn opgeslagen. De EPC-databank bevat dus veel meer informatie dan de EPC-score. Zo worden er zelfs de kenmerken van elke raamopening in opgeslagen. We spreken in het vervolg van de energieprestaties van de woning om alle energetische kenmerken van de woning weer te geven zoals dubbel glas, en waarvan de informatie is opgeslagen in de EPC-databank. Het EPC-kengetal, de EPC-score, is hiervan het samenvattend berekend resultaat dat het theoretisch berekende energieverbruik weergeeft in kWh per jaar per m² bruikbare vloeroppervlakte. Het energieprestatiecertificaat (hierna EPC genoemd) is ingevoerd in het Vlaams Gewest op 1 november 2008 voor verkoop en 1 januari 2009 voor verhuur van woningen en appartementen.

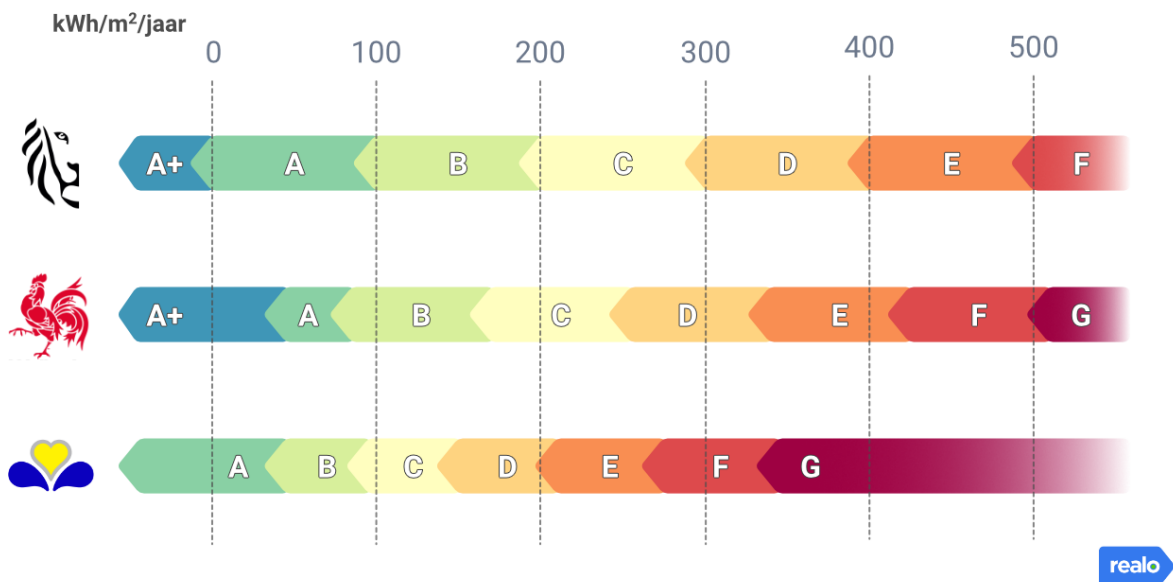
De EPB-certificaten voor nieuwbouw zijn vanaf 2006 beschikbaar voor Vlaanderen. In een EPB-berekening worden de materialen en kenmerken van het gebouw nog meer gedetailleerd ingegeven. Bovendien dient een EPB-verslaggever te beschikken over een ingenieursdiploma of één van architectuur, een hogere vereiste dan dit voor EPC-verslaggever, waar iedereen na het volgen van een

opleiding een EPC-verslaggever kan worden. Dit zorgt ervoor dat de EPC-score die berekend is op basis van een EPB-verslag veelal een hogere kwaliteit heeft dan deze van een EPC-verslag. Maar het belangrijkste is dat zowel bij een EPB-verslag als een EPC-verslag eenzelfde theoretisch energieverbruik in kWh per jaar per m² bruikbare vloeroppervlakte berekend wordt en de EPC-score dus vergelijkbaar is.

Het Brussels Gewest en Wallonië zijn sinds 2011 gestart met het opstellen van EPC's en EPB's.

Hoewel de EPC-score dus voor alle gewesten en type woningen (EPB versus EPC) eenzelfde kengetal weergeven en vergeleken kan worden, zijn er dan ook minimale verschillen. Niet alleen tussen EPC en EPB, maar ook tussen de gewesten, die elk hun eigen protocollen en software hebben voor de berekening van het EPC. Bemerkt dat het vergelijken van energielabels wel verschilt, zoals onderstaande figuur toont.

Figuur 1 EPC-labels per gewest in België



Bron: realo

In 2.5 gaan we dieper in op de onderliggende energetische kenmerken van woningen voor Vlaanderen door het gebruik van de hele EPC databank. Hoewel ook voor de EPB gegevens van nieuwbouw een EPC-kengetal beschikbaar is, gaan we deze gegevens niet gebruiken in deze verfijning aangezien de EPB databank een andere structuur heeft dan de EPC databank en dus geen gelijkaardige variabelen geconstrueerd kunnen worden. Het Vlaams Energie- en Klimaatagentschap (VEKA) voorziet in toenemende mate meer gedetailleerde uniforme data die wel vergelijkbaar zijn, en deze zullen in de toekomst in grotere mate gebruikt kunnen worden, zoals het geaggregeerd energieverlies vanwege muurisolatie, dakisolatie, en zo meer. Dit is *work in progress* en onderdeel van het energiedataplatform. In de toekomst gaat dit dus beter bruikbaar zijn. Een verdere afstemming tussen de berekeningen van het EPC tussen de gewesten staat ook op de agenda.

Wij gebruiken een paar van deze standaardvariabelen die wel voor de hele periode beschikbaar zijn en vullen deze aan met variabelen die we zelf aangemaakt hebben op basis van een datadump (ruwe brondata) van de EPC databank. Dit laat toe om een uitgebreide set van variabelen te creëren maar slechts enkele variabelen kunnen geselecteerd worden omdat andere variabelen niet consistent voor alle perioden aangemaakt konden worden doordat er verschillende versies van de software en inspectieprotocollen zijn (2013, 2017, 2019, 2021), waarbij een vergelijking niet altijd mogelijk was.

2.3 Databronnen en appartementen

EPC data zijn een belangrijke databron als indicator van de kwaliteit van de woning maar zijn enkel beschikbaar voor nieuwbouw, verkochte en verhuurde woningen. Op die manier kunnen ze gebruikt worden in de hedonische prijsanalyses, maar niet om de hele woningvoorraad te waarderen. Voor appartementen zijn er bij het gebruik van EPC-gegevens nog andere uitdagingen die we hier kort bespreken.

2.3.1 Het koppelen van gegevens van appartementen

Het koppelen van gegevens is voornamelijk bij appartementen een probleem omdat er geen officiële busnummers bestaan. Een appartement wordt in het kadaster aangeduid met een partitienummer. Bij de officiële verkoop van een appartement wordt de verkoopprijs van dat appartement gelinkt aan de woningvoorraad via het partitienummer van het appartement binnen het appartementsgebouw, waardoor de bijhorende kenmerken gekend zijn. Voor het koppelen van de EPC-data aan de verkoopprijzen van appartementen is deze optie niet beschikbaar. Aangezien er (historisch) geen officiële busnummers bestaan is het niet mogelijk om de juiste EPC-data te koppelen aan de verkoop. Indien men dus aan de verkoop van een appartement het EPC van een appartement in een appartementsgebouw koppelt waarvan de uitgiftedatum het dichtst voor de datum van verkoop ligt, is het niet zeker indien dit EPC wel correct toegewezen is aan het juiste appartement.

Dit probleem zal in de toekomst echter verdwijnen. Zo werkt men binnen Vlaanderen aan een gebouwenregister⁷ waarbij men zowel aandacht heeft voor het bepalen van appartementen binnen een appartementsgebouw en het opstellen van unieke busnummers als het afstemmen en integreren van het grootschalig referentiebestand (GRB) met de kadastrale plannen. Bovendien is het recent zo dat een EPC slechts kan opgemaakt worden voor een gebouweenheid als de eenheid beschikbaar is in het Gebouwenregister. In het verleden kon het immers gebeuren dat een EPC verleend werd aan een appartement binnen een herenhuis dat officieel nog niet opgesplitst was in verschillende wooneenheden.

2.3.2 Gemeenschappelijke delen

Naast de appartementen binnen een appartementsgebouw zijn er ook gemeenschappelijke delen. Het EPC van de gemeenschappelijke delen van een appartementsgebouw (EPC GD) geeft aan de eigenaars inzicht over de energetische prestaties van deze gemeenschappelijk beheerde delen (= het gebouw) en geeft ook aan hoe de eigenaars die prestaties kunnen verbeteren.

Deze gemeenschappelijke delen beïnvloeden op verschillende manieren de prijs. De grootte van deze gemeenschappelijke ruimtes en de staat hiervan bepalen mee de prijs van het individueel appartement. De waarde van deze gemeenschappelijke elementen zijn niet voorhanden en bovendien ontbreken veelal de verdeelsleutels om de waarde van deze gemeenschappelijke elementen aan elk afzonderlijk privaatief vastgoed toe te wijzen. Wat betreft het EPC is het zo dat alle appartementen mee de gemeenschappelijke delen dienen te financieren. Het dak is ook een gemeenschappelijk deel. Zo zal een dakappartement waarvan het dak niet-geïsoleerd is een slechter EPC hebben, maar is de kost voor dakisolatie een gedeelde kost. Dit zorgt ervoor dat als men een appartement koopt, de energetische kenmerken van het hele appartementsgebouw belangrijk zijn, en dus ook de gemiddelde EPC-score van alle appartementen een betere indicator kan zijn dan het EPC van het appartement zelf. Een laatste element dat belangrijk is bij de koop van een appartement is het reservekapitaal. Dit dient door de vereniging van medeëigenaars (VME) aangelegd te worden om grotere uitgaven voor het

⁷ <https://www.vlaanderen.be/digitaal-vlaanderen/onze-oplossingen/gebouwen-en-adressenregister>

gebouw te dekken, zoals het vernieuwen van het dak. De grootte van het reservekapitaal bepaalt mee de koopprijs van het appartement omdat het reservekapitaal deel blijft van de medeïgendom en niet terugbetaald wordt aan de verkoper.

2.3.3 Koppelverkoop bij appartementen

Soms bestaat de prijs die men betaalt uit twee goederen. In dat geval is het niet mogelijk om de prijs op te splitsen in zijn onderliggende componenten, tenzij ze afzonderlijk bepaald zijn. Dergelijke koppelverkoop worden veelal uit de analyse geweerd.

Bij appartementen is het echter mogelijk om hier een uitzondering op te voorzien omdat het geheel als de verkoop van één woongelegenheden gezien kan worden. Dit is het geval indien een koppelverkoop bestaat uit de verkoop van een appartement, een apart gekadastreerde garage en een apart gekadastreerde berging.

Bij huizen is dit probleem beperkter en anders. Een koppelverkoop bestaat er vaker uit de verkoop met een perceel landbouwgrond dat als uitbreiding van de tuin is aangekocht maar dat officieel niet als deel van het woonperceel geregistreerd is, dus apart gekadastreerd is. Bij de verkoop van het huis – met uitgebreide tuin- krijgt men zo een verkoop van twee percelen.

2.3.4 Niet-residentieel

Eén van de doelen van het RETAX-onderzoek is het ontwikkelen en testen van statistische modellen om het volledige woningpark in één klik te waarderen. Het gaat hierbij over residentieel vastgoed, niet dit van niet-residentieel. Er bestaan echter diverse tussenvormen waardoor dit onderscheid niet altijd eenvoudig gemaakt kan worden. Op eenzelfde perceel kunnen immers zowel residentiële als niet-residentiële gebouwen gevestigd zijn. De AAPD werkt met planpercelen waarop meerdere patrimoniumpercelen (zelfde grondperceel maar met verschillende partities) kunnen gekoppeld zijn. Dus op een planperceel kan zowel een woning liggen, als een aanhorigheid (garage of schuur,...), handelshuis, magazijn,... . Op basis van deze data kan men dus afleiden op welke percelen een combinatie van beiden voorkomt. Zo zijn er niet-residentiële gebouwen die woongelegenheden hebben, en residentiële gebouwen die niet-residentiële functies hebben.

Voornamelijk bij residentiële appartementsgebouwen zijn er vaker niet-residentiële ruimtes (102 en 112, Exploitatie-eenheden voor beroepsdoeleinden, afzonderlijk gekadastreerd, zoals handelspanden, kantoren, uitbatingseenheden). We vinden dat 4% van de huizen een niet-residentieel goed op zijn perceel heeft liggen. Bij appartementen is dit 8% (meer details in Bijlage 1). In deze 8% zitten de exploitatie-eenheden voor beroepsdoeleinden (102 en 112) in vorig punt niet in. Dus deze moeten er bijkomend opgeteld worden.

De verkopen van residentiële gebouwen waarop ook niet-residentiële gebouwen op gelegen zijn, kunnen dus best uit de analyse gelaten worden aangezien niet duidelijk is in welke mate de grond gebruikt wordt door het residentieel of niet-residentieel vastgoed.

2.4 Bespreking EPC binnen België

In de NBB-studie van Reusens, Vastmans en Damen (2022) werden de EPC kengetallen aan de verkoopprijzen gekoppeld op basis van het adres dat in beide databestanden zit. Voor een huis vormt dit geen probleem aangezien het adres en het perceel in dit geval éénduidig gelinkt zijn aan het huis en ook de EPC-score. In een beperkt aantal gevallen heeft een huis reeds diverse EPC-scores. In dat geval wordt het EPC genomen dat net voor de verkoopdatum is uitgereikt. Voor appartementen is het koppelen van EPC's zoals aangegeven niet mogelijk. Er zijn daartoe twee benaderingen gebruikt die zich baseren op de koppeling op adresniveau (appartementsgebouw).

De eerste optie maakt net zoals bij huizen de veronderstelling dat het net voor de verkoopdatum laatst uitgereikte EPC, het EPC is dat bij het appartement hoort. Dit is echter zeker niet altijd het geval. Appartementen worden vaak verhuurd. Voor deze verhuring diende dan reeds eerder een EPC aangevraagd te zijn (en binnen de 9 jaar niet opnieuw aangevraagd te worden voor een verkoop). Indien het appartement later verkocht wordt, is de kans kleiner dat het laatste uitgereikte EPC binnen een appartementsgebouw ditgene is dat uitgereikt is voor de verkoop van het appartement. Bovendien zijn er sterke verschillen in EPC-kengetallen binnen een appartementsgebouw.

De tweede optie gaat ervan uit dat het gemiddelde kengetal van EPC's uitgereikt voor de verkoopdatum een beter beeld geeft van de energetische situatie. Dit gemiddelde is stabiel per appartementsgebouw dan enkel het laatst uitgereikte EPC. Bovendien is het zo dat de EPC-waarden tussen appartementen binnen eenzelfde appartementsgebouw sterk kunnen verschillen. Dit gemiddelde EPC geeft dus een beter beeld van de energiescore van het hele appartementsgebouw, wat zoals eerder aangehaald, ook prijsbepalend is omdat bv. dakisolatie een gemeenschappelijke kost is.

Op basis van de resultaten van Reusens et al (2022) kunnen we besluiten dat voor appartementen het meest recente EPC koppelen de beste resultaten geeft om de evolutie van het gemiddelde EPC van de verkochte appartementen weer te geven (zoals weergegeven in Figuur 2), maar dat het gemiddelde EPC van alle appartementen binnen een appartementsgebouw een betere predictor is om de (energetische) waarde van het appartement te schatten (zoals weergegeven in Figuur 3). De plots in Figuur 1 geven de evolutie van het gemiddeld EPC weer (Figuur 2). Deze in Figuur 2 geven als de waardering voor het EPC-kengetal (Figuur 3) Zowel de evolutie van het gemiddelde EPC als de waardering vergelijkt gelijkaardig in de drie gewesten⁸.

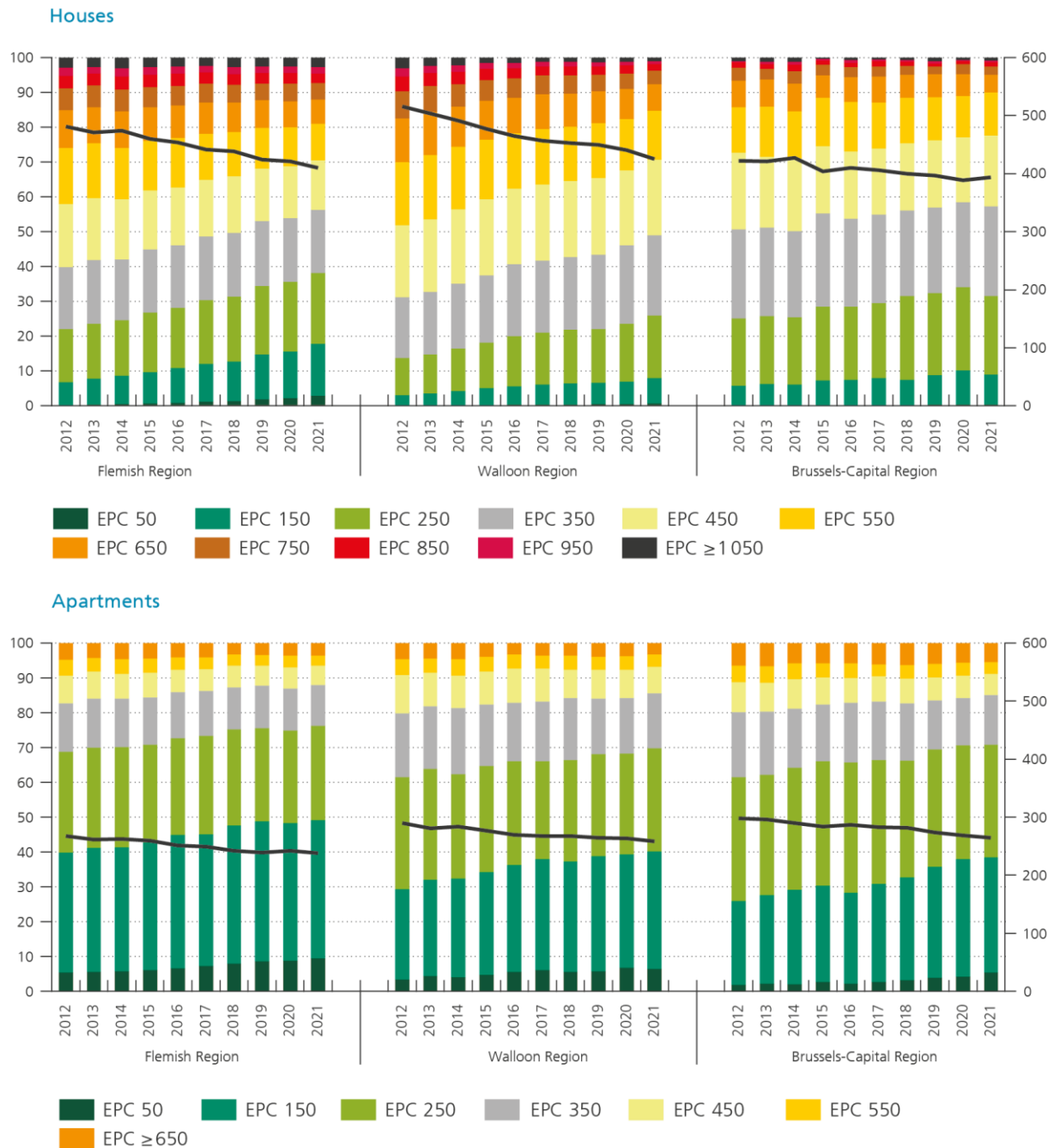
Bij de verbetering van een EPC met 100 (wat vaak overeenkomt met een energielabel dat 1 klasse beter is, bv. van D naar C), zien we dat de woningprijzen 4% à 7% hoger zijn (meer bij de betere energieklassen, en minder bij de slechtere energieklassen). Een waardering die bovendien licht toeneemt doorheen de tijd.

Bij de interpretatie dient men er rekening mee te houden dat het EPC in dit geval een belangrijke proxy is voor de algemene woningkwaliteit. EPC data zijn de enige officiële databron omtrent de woningkwaliteit die op grote schaal en volgens uniforme wijze gemeten zijn. Veelal zal een woning met goede energetische kwaliteiten ook andere goede kwaliteitskenmerken hebben zoals een recente keuken en badkamer. Door deze samenhang (multicollineariteit) en het feit dat de andere kwaliteitsvariabelen niet in het model opgenomen zijn (omitted variable bias) – want niet beschikbaar – zal de geschatte coëfficiënt van de EPC-score niet alleen de energetische, maar ook ten dele de waardering van de algemene kwaliteit van de woning schatten. Damen (2019) maakte het onderscheid tussen beiden en vond dat de geschatte woningprijs met 5% steeg indien de EPC-score met 100 verbetert indien het ook de algemene woningkwaliteit schat, en 3,5% indien de EPC-score met 100 verbetert en enkel het energetisch effect weergeeft (doordat de andere kwaliteitsvariabelen wel in het model opgenomen werden).

Een andere variabele die de woningkwaliteit weergeeft is het jaar van renovatie. Het model geeft weer dat de woningprijs 9% hoger geschat wordt indien de woning recent gerenoveerd is.

⁸ Voor Brussel zijn de resultaten iets minder stabiel, maar dit kan verklaard worden door het beperkter aantal huizen dat er verkocht wordt (en mogelijke vintage-effecten, waarbij oude huizen met een hoge architecturale waarde hoog gewaardeerd worden hoewel deze mogelijk een slechtere energiescore hebben).

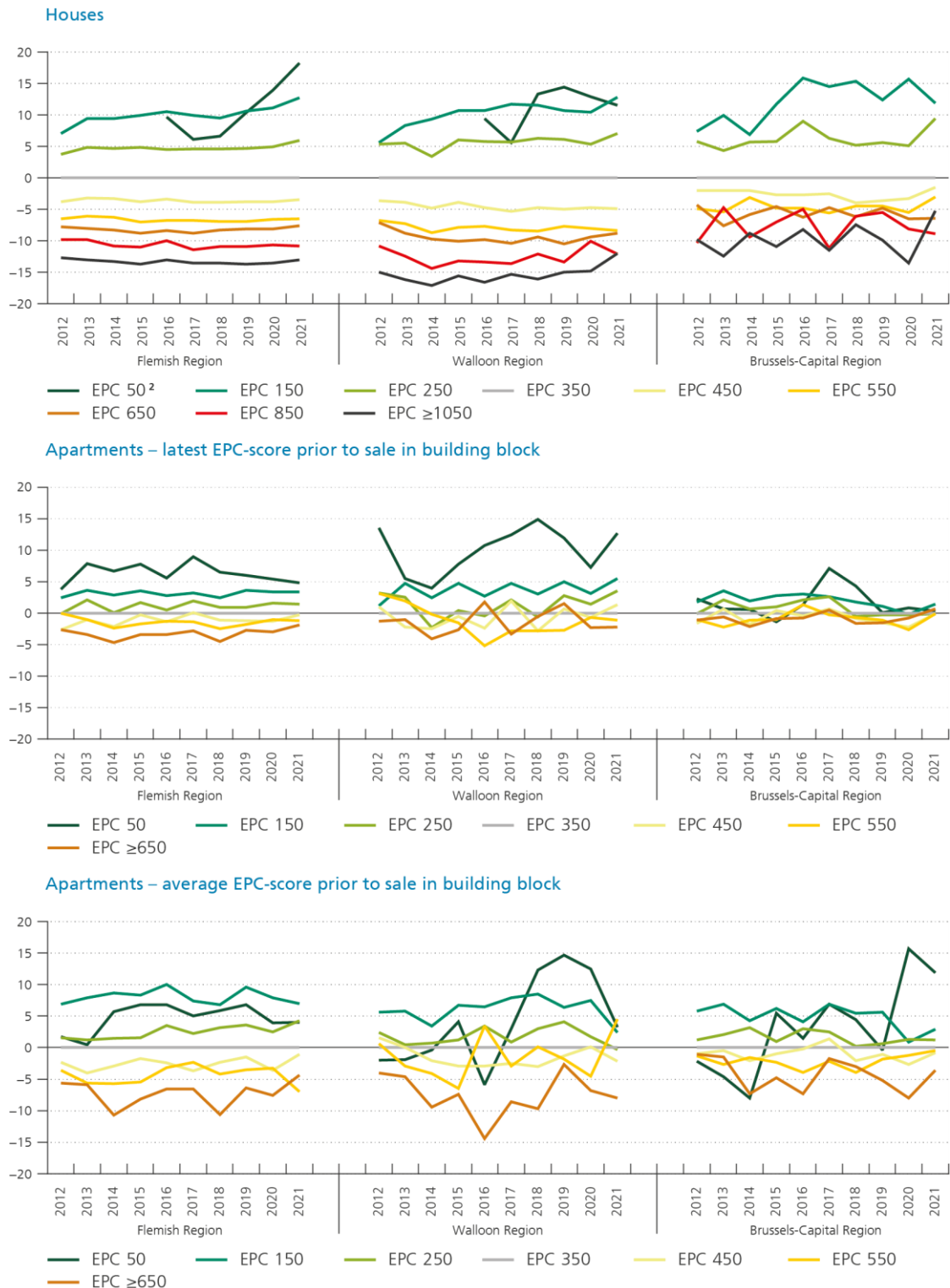
Figuur 2 Verdeling van de EPC-scores van verkochte huizen en appartementen volgens het jaar van verkoop¹ (in % (linker-as; gemiddeld EPC in kWh/m²y (rechteras))



* **Bemerk dat de jaren twee kwartalen voorwaarts geschoven zijn, bv. "2021" komt overeen met 2020Q3-2021Q2).**

Bron: Sources: FPS Finance, VEKA, Brussels Environment, SPW Wallonie, calculations Reusens, Vastmans en Damen (2022) chart 1

Figuur 3 Geschatte energie-efficiëntie prijspremie ¹ (in %, prijsverschil met een gelijkaardige woning met een EPC score van 350 kWh/m²)



*

Bron: Sources: FPS Finance, VEKA, Brussels Environment, SPW Wallonie, calculations Reussens e.a. (2022) chart 18

2.5 Energieprestatie-kenmerken van woningen in Vlaanderen

De EPC-databank is een zeer uitgebreide dataset van energetische woningkenmerken. Maar daarnaast bevat deze databank ook andere gegevens, zoals algemene woningwoningkenmerken omtrent de grootte van de woning, of gegevens omtrent de aanvraag van het EPC.

Indien we naar de predictiekracht kijken vergelijken we 3 modellen met elkaar. Het eerste basismodel is een log-lineair model met fixed effects voor gemeenten (FE). Vervolgens voegen we daar de EPC-scores aan toe, gelijkaardig aan de modellering in vorig punt. In een laatste stap gaan we na hoe de predictie verbetert bij het toevoegen van meer gedetailleerde data. Tabel 2 toont dat het basismodel zonder EPC een verklaarde variantie heeft van 73%. Dit is laag ten opzichte van het model waar het EPC kengetal is toegevoegd (77.6%). Voor het EPC-kengetal zijn er 10 variabelen toegevoegd die beter toelaten om de impact te meten van de verschillende energielabels (volgens de Vlaamse indeling). Het is niet verwonderlijk dat de verklaarde variantie sterk stijgt door het toevoegen van het EPC aangezien het EPC de enigste belangrijke variabele die de kwaliteit van de woning weergeeft. Indien we vervolgens nog extra variabelen toevoegen, 74 variabelen in totaal, dan zien we dat de meerwaarde beperkt is: de verklaarde variantie stijgt slechts met 1,7%.

Tabel 2 Verklaarde variantie (R^2) van EPC modellen

| Model | R^2 | Aantal variabelen |
|---------------------------|-------|-------------------|
| Basismodel FE | 73.0% | 348 |
| Model FE + EPC | 77.6% | 358 |
| Model FE + EPC uitgebreid | 79.3% | 432 |

Bron: AAPD verkoopprijzen van huizen, EPC-data VEKA, eigen verwerking

De meerwaarde van het toevoegen van extra variabelen neemt dus af naarmate er meer variabelen in het model zitten. Bovendien worden de coëfficiënten hierdoor minder goed interpreteerbaar en is het moeilijk om de invloed en de predictiekracht van deze variabelen afzonderlijk te meten vanwege de onderlinge samenhang tussen al deze variabelen (multicollineariteit). We bespreken kort de meeste significante resultaten:

2.5.1 Energetische variabelen

De impact van het EPC kengetal daalt ten voordele van meer gedetailleerde energetische variabelen. De resultaten zijn logisch. Zo heeft de aanwezigheid van een buitenvoeler, die de buitentemperatuur meet voor de weersafhankelijke regeling van de centrale verwarming, een positieve impact op de woningprijs, net zoals thermostatische radiatorcransen hoger gewaardeerd worden dan manuele. Het opzetten en interpreteren van al deze variabelen is echter een complexe oefening. Zo kan er in een woning zowel een houtkachel als een condensatieketel aanwezig zijn. Het is in dergelijk geval niet juist om aan te duiden dat er zowel een condensatieketel als een houtkachel aanwezig is via dummy-variabelen. Op basis van de databank werd daartoe een variabele gecreëerd die het aandeel van elke warmtebron weergeeft.

2.5.2 Algemene woningkenmerken vanuit de EPC-databank

In de EPC-databank zitten eveneens variabelen omtrent oppervlakte en bouwjaar. Gelijkaardige variabelen waren reeds via AAPD aanwezig en de kwaliteit van de AAPD variabelen bleek beter dan deze EPC-variabelen aangezien deze beter de variatie in de woningprijs kon verklaren. De nuttige woonoppervlakte in het kadaster is ook een gewogen gemiddelde, waarbij de woonoppervlakte van garages, kelders en zolders een kleiner gewicht krijgen. Deze benadering lijkt beter in lijn te liggen met

de marktwaarde van een woning dan de ruwe woonoppervlakte binnen het beschermd volume uit de EPC-databank. De nuttige woonoppervlakte biedt immers het voordeel dat ruimtes die veelal minder kwalitatief ingericht zijn (zolder, kelder, garage) ook slechts voor de helft meetellen.

Het beschermd volume uit de EPC-databank geeft echter wel een belangrijke meerwaarde. Deze capteert de combinatie van de oppervlakte en de hoogte van de kamers. De gemiddelde hoogte in een woning is dan ook een belangrijke variabele die niet in het kadaster beschikbaar is.

2.5.3 Transactie info

In de EPC-databank zitten ook gegevens die informatie bevatten over de verkoper en een indicatie dat het over een huurwoning gaat. Zo vonden we dat indien het EPC aangevraagd werd voor verhuur, de woningprijs 3% lager geschat wordt. Men kan verwachten dat een huurwoning met eenzelfde EPC als een eigenaarswoning voor de rest minder kwalitatieve kenmerken heeft. Vandaar de negatieve coëfficiënt bij de huur-dummy-variabele.

Daarnaast is het ook het type eigenaar belangrijk. Verkopen door overheden en sociale huisvestingsmaatschappijen worden 10% lager geschat in vergelijking met particulieren en private vennootschappen.

2.6 De aanwezigheid van een EPC

De EPC-score draagt bij tot een betere predictie van de woningwaarde, maar er is meer. Het is aangewezen om de steekproef van verkopen te beperken tot deze waarvoor een EPC beschikbaar is, en dus de andere verkopen niet in de steekproef op te nemen. Hiervoor kunnen twee redenen aangehaald worden.

- 1) Niet voor elke verkoop dient een EPC aangevraagd te worden. Dit is enkel het geval indien de verkoop **openbaar** is, met promotie, dus wanneer vraag en aanbod meer meespelen. Verkopen met EPC zijn waarschijnlijk deze waarbij de verkopen volgens de **marktprijs** verkocht worden. Als de woning zonder publieke reclame verkocht of verhuurd wordt, is er geen EPC nodig. Indien de verkoop niet publiek gebeurt, kan men veronderstellen dat de woning ook vaker verkocht wordt tegen een lagere prijs, niet noodzakelijk een “vriendenprijsje” waarbij mogelijk een deel in het zwart gebeurt (bv. om verkooprechten te ontlopen), maar ook omwille van het economisch gegeven dat een niet-publieke verkoop veelal samengaat met minder vraag. In dit geval is het dus niet gewenst dat deze observatie mee in de analyse genomen wordt aangezien we de woningwaarde wensen te schatten die tot stand komt in normale marktomstandigheden met voldoende vraag.
- 2) Daarnaast is een EPC niet verplicht indien de woning **ongeschikt of onbewoonbaar** is verklaard (art. 1.1.1., §2 Energiebesluit). De woning zal dan in een uitzonderlijk slechte staat zijn. Hoewel we geen indicatie hebben over het aandeel ongeschikt of onbewoonbare woningen in de verkoop, weten we wel dat er jaarlijks tussen de drie- à vierduizend woningen ongeschikt of onbewoonbaar verklaard worden in het Vlaams Gewest. In dit geval is het dus niet gewenst dat deze observatie mee in de analyse genomen wordt aangezien de woning niet bewoond kan worden. De woning zal dus ofwel gerenoveerd, ofwel gesloopt en heropgebouwd worden. De verkoopprijs kan wel tegen marktwaarde gebeuren, maar de huidige staat van de woning (en dus ook de marktwaarde) is niet deze van een bewoonbare woning.

Enkel het feit dat de verkochte woning een EPC-certificaat heeft, draagt bij tot een betere predictie door het feit dat de dataset hierdoor kwalitatiever is. Dit is nagegaan door 5 verschillende specificaties van het model te testen.

- Specificatie 1: het volledige model, waarbij ook de EPC-variabelen mee opgenomen zijn.

- Specificatie 2: zelfde steekproef als Specificatie 1, dus verkopen met EPC's, maar de EPC-variabelen zelf worden uit het model gelaten.
- Specificatie 3 : Niet alleen de EPC-variabelen worden uit het model gelaten, maar de steekproef wordt nu uitgebreid met verkopen waarvoor geen EPC's beschikbaar zijn.
- Specificatie 4: Aangezien specificatie 3 10% meer observaties heeft dan in specificatie 2 kan de verklaarde variantie (R^2) verschillen vanwege de grootte van de steekproef. In specificatie 4 wordt daartoe een willekeurige selectie gemaakt van +/-90% van totale steekproef van de verkopen uit specificatie 3, zodat de steekproef even groot is als deze van specificatie 2. Op die manier is de vergelijking van de resultaten niet beïnvloed door de grootte van de steekproef.
- Specificatie 5 : Aangezien we verwachten dat de verkopen zonder EPC een lagere verkoopprijs hebben, voegen we een variabele toe die aangeeft dat de observatie geen EPC heeft. Daarnaast kunnen we vermoeden dat het onbewoonbaar en ongeschikt zijn van een woning vooral bij oudere woningen mee zal spelen en een grotere prijsimpact heeft. We splitsen deze variabele daarom op in een variabele die aangeeft dat er geen EPC aanwezig is voor een woning jonger dan 70 jaar, en voor een woning met een leeftijd van 70 jaar of ouder.

De resultaten (Tabel 3) zijn zoals verwacht. Specificatie 1 waar het EPC is opgenomen in het model, heeft de grootste verklarende kracht (R^2). Specificatie 2 waarbij enkel verkopen met EPC zijn opgenomen verklaart een stuk minder dan specificatie 1 omdat de belangrijke EPC-variabele niet mee opgenomen is, maar verklaart wel beter dan specificatie 3 met de volledige set van verkopen. Dit is niet te wijten aan een verschil in steekproefgrootte, aangezien specificatie 4 hetzelfde resultaat geeft als specificatie 3, maar wel door het feit dat verkopen zonder EPC tegen een lagere prijs verkocht worden vergeleken met gelijkaardige woningen (zelfde grootte, locatie, bouwjaar). Specificatie 3 houdt hier geen rekening mee, specificatie 5 wel. Voor een woning met een leeftijd kleiner dan 70 jaar, is de geschatte verkoopprijs 66% lager voor een verkoop zonder EPC, in vergelijking met eenzelfde woning waarvoor wel een EPC is uitgereikt. Voor een woning met een leeftijd van minstens 70 jaar is de geschatte verkoopprijs -13,8% lager voor een verkoop zonder EPC. De resultaten suggereren dat de verkoopprijs van een woning zonder publiciteit gemiddeld 6% lager liggen dan deze van een gelijkaardige woning met publiciteit, in de veronderstelling dat ongeschikt en onbewoonbaar verklaarde woningen niet voorkomen bij recentere woningen. Over het effect van onbewoonbaar of ongeschikt kunnen we geen veronderstelling maken, al kunnen we verwachten dat dit groter is dan -13,8%, omdat de -13,8% lagere verkoopprijs voor woningen ouder dan 70 jaar ook ten dele verklaard worden door verkoop van woningen zonder publiciteit, waarvan het prijseffect lager geschat is, namelijk het resultaat dat we vonden voor de recente woningen. De opsplitsing op basis van leeftijd is gemaakt op basis een analyse in de verschillen in coëfficiënten tussen specificatie 2 en 3. De coëfficiënten verschillen nauwelijks tussen alle woningkenmerken, met uitzondering dus voor deze van oudere bouwjaar, wat in lijn ligt van de verwachtingen.

Tabel 3 Verklaarde variantie van modellen met diverse specificaties voor aanwezigheid EPC

| Specificatie | 1 | 2 | 3 | 4 | 5 |
|------------------------------|--------|--------|--------|--------|--------|
| R^2 (verklaarde variantie) | 78.21% | 73.23% | 72.76% | 72.75% | 73.24% |

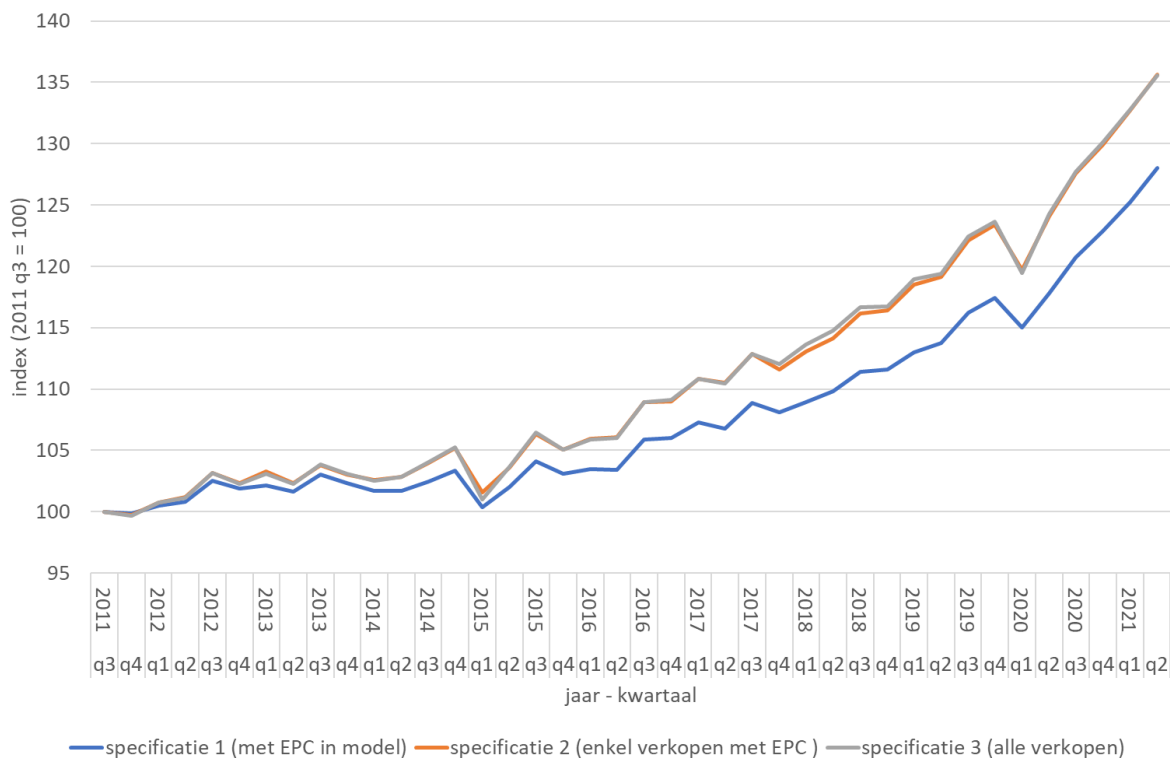
Bron: eigen berekeningen

Deze analyse illustreert dat het bij hedonische prijsanalyses aangewezen is om met een beperktere set van kwalitatieve data te werken, dan met de volledige set van verkopen. Het is evenwel niet zo dat huizen van minder dan 70 jaar oud die zonder EPC verkocht worden gemiddeld 6% onder de prijs verkocht worden. Het verschil kan ook verklaard worden door het feit dat ze van kwaliteit verschillen (omitted variable bias), waar in de hedonische prijsanalyse zonder EPC variabelen niet voor gecompenseerd wordt. Indien de kwaliteit van woningen zonder EPC gemiddeld lager is dan deze voor

gelijkaardige woningen met EPC verklaart dit ook de lagere prijs. Toch kunnen we besluiten dat deze observaties van mindere kwaliteit zijn met grotere schattingsfouten en vooral ook mogelijke vertekeningen.

Het al dan niet mee opnemen van EPC in de regressievergelijking heeft ook een methodologische kanttekening. Zo wordt er niet voor elke renovatie een nieuw EPC opgesteld. Maar tegelijkertijd worden woningen wel vaak gerenoveerd nadat ze gekocht worden (Vastmans 2020). Ook voor de hedonische prijsindex, die aangeeft hoe eenzelfde woning verandert doorheen de tijd, heeft dit een belangrijke impact. De hedonische prijsindex verschilt niet tussen specificatie 2 en 3 zoals uit onderstaande figuur duidelijk blijkt. We zien wel een groot verschil tussen de hedonische prijsindex tussen specificatie 1 en 2. Door niet te corrigeren voor de verbetering in het EPC doorheen de jaren, laat de hedonische prijsindex zonder EPC-scores een veel sterkere stijging zien dan deze waarbij de hedonische prijsindex wel rekening houdt met de toegenomen verbeteringen van de EPC-scores doorheen de jaren. Bovendien is het zo dat niet enkel de EPC-score verbetert doorheen de tijd, maar ook de waardering voor de EPC-score, alsook deze van andere woningkenmerken kan veranderen doorheen de tijd.

Figuur 4 Hedonische prijsindices volgens verschillende modelspecificaties



Bron: AAPD, VEKA, eigen berekeningen

3. EVOLUTIE VAN WOONVOORKEUREN DOORHEEN DE TIJD

Het schatten van woningwaarden gebeurt op basis van historische data. Hiervoor zijn voldoende data nodig. Om de coëfficiënten in een regressiemodel te schatten wordt soms als een vuistregel naar voren geschoven dat er minimaal 10 observaties nodig zijn per geschatte coëfficiënt. Voor het maken van predicties, dus niet enkel om aan te geven of een resultaat significant is, zijn meer observaties gewenst. Om de coëfficiënten van de woningkenmerken te schatten vormt dit veelal geen probleem aangezien het slechts een beperkte set van variabelen is en de steekproef zeer groot. Voor locatie-effecten is dit wel een aandachtspunt. Indien men het locatie-effect voor een bepaalde gemeente wil schatten dienen er reeds minimaal 10 observaties (en liefst meer) beschikbaar te zijn per gemeente. Om het locatie-effect binnen een gemeente verder te verfijnen zijn nog bijkomende observaties nodig. Daartoe maakt men niet enkel gebruik van de verkopen van het meest recente jaar, maar typisch de verkopen van de laatste 5 à 10 jaar. Hierdoor kan men de locatie-effecten, en ook de impliciete prijs van andere karakteristieken beter schatten en de verschillen tussen de locatie beter duiden. Dit is het voordeel van meer observaties.

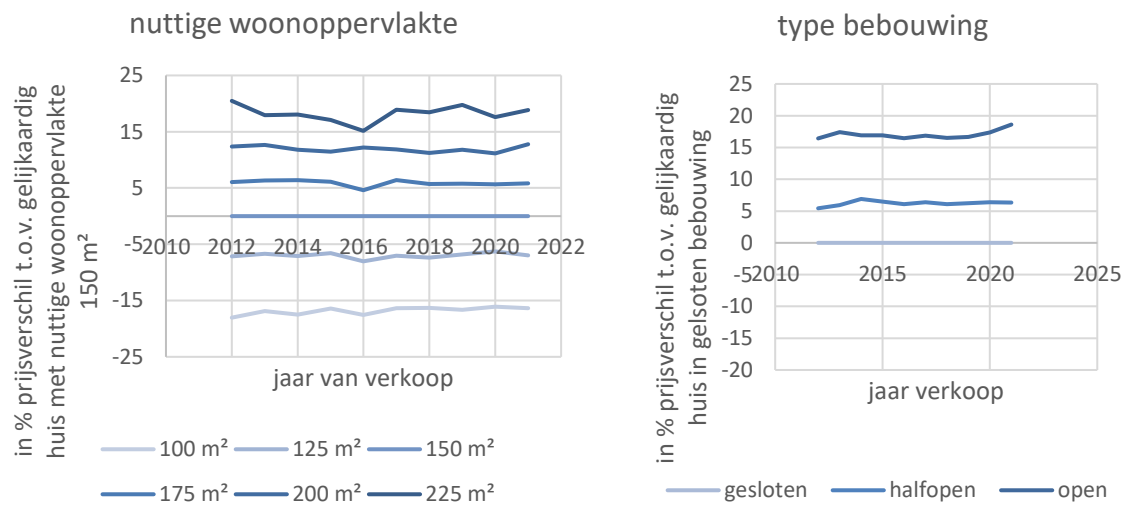
Het gebruik van 'oudere' observaties, dus verkopen van vroegere jaren, zorgt er wel voor dat het hedonisch model de gemiddelde voorkeur meet van de diverse woningkenmerken over een langere periode, en niet enkel de meest recente. Indien voorkeuren veranderen krijgt men zo schattingen die niet volledig up to date zijn. Dit is anno 2022 des te relevanter aangezien Covid-19 de vraag naar wonen versterkt heeft wat wereldwijd tot sterke prijsstijgingen heeft geleid. Een hedonisch prijsmodel houdt rekening met deze gemiddelde prijsstijging. Maar de meeste hedonische prijsmodellen houden geen rekening met veranderingen van woonvoorkeuren, die ertoe leiden dat bepaalde woningsegmenten, gebaseerd op type woning of locatie, een prijsstijging optekenen die meer of minder dan gemiddeld is. Typische log-lineaire modellen maken gebruik van een dummy variabele die enkel de algemene prijsevolutie meet. Het mogelijk nadeel van het gebruik van verkopen van minder recente datum is dat het de huidige woonvoorkeuren niet exact weergeeft.

Hedonische modellen kunnen echter rekening houden met veranderende woonvoorkeuren door gebruik te maken van coëfficiënten die variëren doorheen de tijd. Reusens, Vastmans en Damen (2022) hebben zo een theoretische methodologie van Eurostat praktisch verder uitgewerkt die de evolutie van de jaarlijkse woonvoorkeuren meet en pasten dit toe voor België, met extra aandacht voor de recente periode (tot en met het tweede kwartaal van 2021).

Zij vonden dat de woonvoorkeuren met betrekking tot de kenmerken van de woning veeleer stabiel waren. Op onderstaande figuur zien we dat de coëfficiënten voor nuttige woonoppervlakte en type bebouwing relatief stabiel zijn in het Vlaams Gewest⁹. Tijdens de COVID-periode (jaar 2021 geeft 1 juli 2020 tot 30 juni 2021 weer) zien we een iets hogere premie voor grotere woningen in open bebouwing. Het dient verder nagegaan te worden indien het werken van thuis uit na de COVID-periode tot een blijvend hogere premie leidt.

⁹ De cijfers voor Wallonië en Brussel laten een gelijkaardig maar gevarieerder beeld zien. De steekproef is er beperkter waardoor de jaarlijks geschatte coëfficiënten sterker variëren doorheen de tijd.

Figuur 5 Jaarlijkse * coëfficiënten voor nuttige woonoppervlakte en type bebouwing, Vlaams Gewest 2012-2021

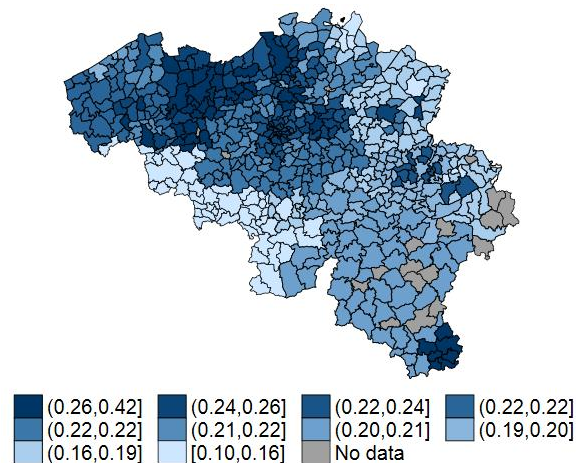


* jaar 2021 is van 1 juli 2020 tot 30 juni 2021

Bron: AAPD, VEKA, verwerking: Reusens, Vastmans en Damen (2022)

Daarnaast is er vooralsnog geen COVID-effect dat wijst op een uittocht uit de steden. De dynamische centrumsteden die sterkere prijsstijgingen optekenden voor COVID-19 stegen ook iets sterker tijdens COVID-19. Hiervoor zijn diverse mogelijke verklaringen. Huishoudens kunnen mogelijk meer belang hechten aan de huidige woonbuurt waarbij een verdere verhuis tijdens COVID minder aantrekkelijk is omdat COVID-19 de sociale interactie in een nieuwe woonbuurt beperkt. Daarnaast is het ook zo dat elk huishouden mogelijk opteert voor een grotere woning in verhouding tot de huidige grootte van de woning. Dat kan betekenen dat de vraagdruk op woningen van 100m² zeker even groot kan zijn als deze van 200m², omdat er meer huishoudens wensen te verhuizen van een woning van 80m² naar 100 m² dan van 180m² naar 200m². Het is de relatieve grootte die belangrijk is. Indien we meer in detail inzoomen in de regionale prijsontwikkelingen op langere termijn zien we dat het Waalse Gewest een iets minder sterke stijging optekende dan de andere gewesten (2021 t.o.v. 2012). Binnen Vlaanderen zijn het voornamelijk de stadsgewesten Gent, Antwerpen Mechelen, Leuven en Hasselt waar de prijs van een gelijkaardige woning sterker stegen. Voor verklaringen hiervoor verwijzen we naar Vastmans en Dreesen (2021).

Figuur 6 Regionale hedonische prijsindex (% stijging 2021 t.o.v. 2012) volgens Stadsgewesten*



* aangepaste indeling van Vanderstraeten et al (2019)

Bron: AAPD, VEKA, verwerking: Reusens, Vastmans en Damen (2022)

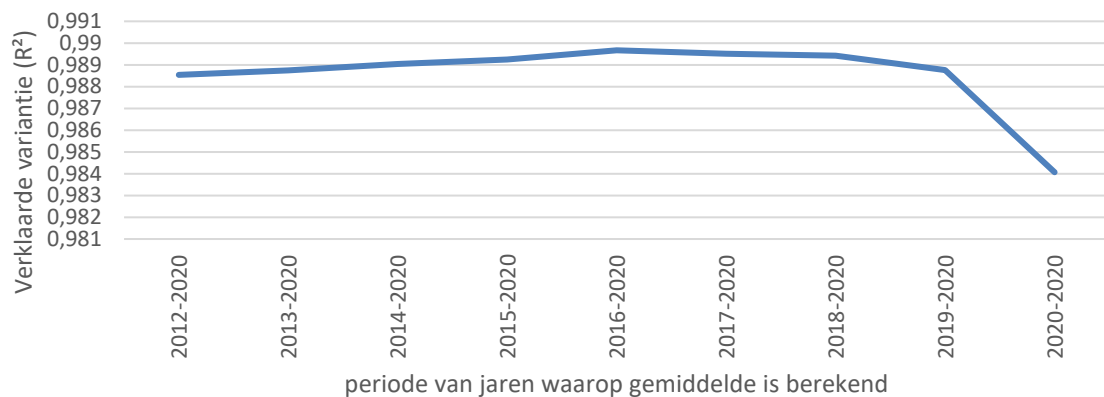
De specifieke vraag voor dit rapport is echter hoe vaak men best het model update en hoe lang de data periode moet zijn die men gebruikt om het model. Zoals eerder aangegeven is het een trade-off tussen een stabielere voorspelling (meer observaties en minder ruis) ten opzichte van een recente voorspelling (geen vertekening ten opzichte van de huidige trend door het gebruik van enkel recente data).

Voor het schatten van de coëfficiënten van woningkenmerken stelt dit probleem zich minder. Op basis van de laatste paar jaren (of zelfs laatste jaar) zijn er veelal voldoende observaties beschikbaar voor stabiele schattingen. Per woningkenmerk (of klasse van woningkenmerk) zijn er voldoende observaties. Dit komt omdat er veel meer observaties zijn dan woningkenmerken.

Om de regionale verscheidenheid in kaart te brengen heeft men zoals eerder gemeld voldoende observaties nodig. In de NBB studie werd de evolutie van de locatie-effecten daartoe niet op gemeenteniveau weergegeven maar op niveau van stadsgewesten. De jaarlijks geschatte gemeentelijke locatie-effecten zijn wel berekend¹⁰. Op basis van deze jaarlijkse gemeentelijke locatie-effecten kunnen we eenvoudig nagaan hoe stabiel deze gemeente-effecten zijn. In de mate dat er weinig ruis zit in de data zal het gemeente-effect weinig variantie vertonen en zal het huidige gemeente-effect best benaderd worden door de het laatste en meest recent beschikbare gemeente-effect. Indien er echter meer ruis in de data zit, zal het huidige gemeente-effect best geschat kunnen worden op basis van gegevens over meerdere jaren heen. Het is dus een afweging. Figuur 7 geeft weer hoe goed het model het meest recente gemeente-effect (2021) verklaart volgens het gemiddelde gemeentelijke liggingseffect berekend voor een periode van jaren. We zien dat het liggingseffect van 2020 (het dichtst bij 2021) niet het beste resultaat geeft. Een combinatie van twee jaren geeft veel betere resultaten en het beste resultaat vindt men indien het gemeentelijk liggingseffect gebaseerd is op dit van de laatste 5 jaar. Vervolgens, voor meer dan 5 jaren, neemt het af: het voordeel van de stabielere resultaten vanwege meer verkopen wordt dan kleiner ten opzichte van het nadeel dat de meeste recente trend te weinig gewicht krijgt.

¹⁰ Te downloaden op <https://www.nbb.be/nl/artikels/de-impact-van-veranderingen-woningkenmerken-en-voorkeuren-op-de-belgische-woningprijzen>

Figuur 7 verklaarde variantie van model dat gemeente-effect 2021 voorspelt op basis van gemiddelde gemeente-effect van voorgaande jaren (België 2012-2021)



Bron: gemeentelijke liggingseffecten Reusens, Vastmans en Damen (2022), eigen berekeningen

Indien we deze analyse differentiëren volgens grootte van de gemeente en het aantal verkopen, dan vinden we dat dit voornamelijk verklaard wordt door de kleinere gemeenten. Bij grotere gemeenten (met meer dan 100 verkopen per jaar), vinden we dat de ruis beperkter is, en men minder jaren nodig heeft voor stabielere resultaten.

Wat wil dit zeggen voor het gebruiken van geschatte woningwaarden als belastbare basis van het kadastraal inkomen en het gebruik van de KNN-methode?

Vooreerst lijkt het gepast dat de KNN-methode zijn observaties baseert op een tijdshorizon die groot genoeg is, aangezien op die manier het locatie-effect stabielers geschat kan worden. Ook voor gemeentedummies dient men voldoende observaties te hebben voor stabiele coëfficiënten. Schildermans en Damen (2021) gebruikten dit als uitgangspunt.

Op basis van het onderzoek van Reusens et al (2022) zou men best ook rekening houden met de evolutie op een iets bredere schaal, bvb. op het niveau van de stadsgewesten. Vanderstraeten et al. (2019) definieerde 19 stadsgewesten in België, met vier verschillende niveau's van stedelijkheid. Het grote voordeel van dergelijke benadering is dat elk stadsgewest (en zijn bijhorende urbane regio's) een eigen locatie-effect krijgen, en de eigen dynamiek beter gemodelleerd kan worden. Hoewel woningprijzen vaak gelijkaardig evolueren, zijn er toch grote regionale verschillen. Zo vond Reusens et al. (2022) als twee extremen dat een gelijkaardige woning in Gent tussen 2012-2021 met 40% is gestegen, terwijl dit in Bergen, Charleroi en Doornik eerder rond de 12% schommelde. De legende van Figuur 6 geeft hier een indicatie voor alle stadsgewesten. Gemiddeld gezien zijn de verschillen een stuk kleiner. Desalniettemin wijzen de resultaten erop dat de regionale verschillen tussen locaties sterker fluctueren dan deze van de woningkenmerken.

Beide resultaten kunnen eenvoudig gecombineerd worden door de KNN-methode naast de gemeentedummies uit te breiden met tijdsveranderlijke stadsgewest-effecten.

CONCLUSIE

De uitbreiding van de loglineaire hedonische prijsmethode met de K-Nearest-Neighbor methode (KNN), waarbij informatie van de dichtstbijgelegen observaties het locatie-effect bepalen blijkt een methode die verschillende voordelen biedt als methode om de verkoopwaarde mee te schatten als basis van het kadastraal inkomen. In dit onderzoeksrapport focusten we op verdere verfijningen binnen de KNN-methode, waarbij twee elementen centraal staan: enerzijds de beschikbaarheid van data, anderzijds de meerwaarde om deze toe te voegen aan de analyse verdere verfijningen aan het model zelf.

Data-beschikbaarheid

Er zijn nog diverse administratieve en technische uitdagingen met betrekking tot de beschikbaarheid van data die nodig zijn voor predictiemodellen. Hier wordt op verschillende domeinen aan gewerkt. Zo werkt men binnen Vlaanderen aan een gebouwenregister waarbij men zowel aandacht heeft voor het bepalen van appartementen binnen een appartementgebouw en het opstellen van unieke busnummers als het afstemmen en integreren van het grootschalig referentiebestand (GRB) met de kadastrale plannen. Op die manier zullen de EPC's voor appartementen beter gekoppeld kunnen worden. De energieprestaties van de woningen zullen waarschijnlijk ook aan belang winnen in de toekomst. Een meer algemene vraag wat betreft data-beschikbaarheid is of het wenselijk is om voor alle woningen in de woningvoorraad een EPC (of variant?) beschikbaar te hebben. Hoewel er binnen Vlaanderen reeds meer dan 1 miljoen EPC's beschikbaar zijn, en ook voor nieuwbouw EPB volledig beschikbaar zijn, beschikt het andere deel van de woningen niet over een EPC. Dit zorgt ervoor dat het EPC moeilijk kan opgenomen worden in de belastbare basis. In de mate dat de energieprestaties niet belast zouden worden volstaat het om voor alle woningen de waarde te schatten met een gelijkaardig EPC.

Naast de strikt gescheiden private delen die eenvoudiger te schatten zijn, spelen gemeenschappelijke delen ook een rol. Een analyse van het EPC voor de gemeenschappelijke delen is vooralsnog nog niet uitgevoerd. Het lijkt belangrijk om dit in de toekomst in meer detail op te volgen, en gemeenschappelijkheid in het algemeen, gezien het collectief wonen – hoewel nog kleinschalig in aandeel- aan belang toeneemt. Omgekeerd kan het ook interessant zijn om een woning verder op te splitsen in de waarde van de structuur en deze van grond. De gemiddelde stijging van de prijsindices van de bebouwde grond is groter dan deze van de structuur erop (Francke en Van de Minne, 2017).

In de mate dat de verschillende geografische informatie ook in hetzelfde coördinatenstelsel gemeten worden en op elkaar afgestemd zijn zouden ook meer verfijnde analyses mogelijk zijn met geografische informatie.

Meerwaarden van extra informatie

De analyse toont dat de meerwaarde van extra variabelen in het model beperkt is additioneel bovenop het gebruik van de KNN-methode. Dit is zowel het geval voor geografische informatie (overstromingsgevoelig, informatie over het soort straat waar de woning aan gelegen is, buurtkenmerken, knooppuntwaarde en voorzieningenniveaus). Maar dit wil niet zeggen dat deze niet nuttig zijn, aangezien het individueel wel een grote impact kan hebben. Het aandeel huizen dat in overstromingsgevoelig gebied ligt kan bijvoorbeeld beperkt zijn, maar het heeft individueel wel een impact. Daarnaast is de vraag in welke mate deze huizen anders belast dienen te worden. Beleidsmatig kan men ervoor opteren om woningen in overstromingsgevoelig gebied niet minder te belasten, maar mogelijk zelfs meer aangezien ze vaker een grotere kost betekenen voor de overheid in termen van interventie bij overstromingen. Hierbij rijst de vraag of er naast de woningwaarde ook rekening gehouden dient te worden met de positieve en negatieve

externaliteiten van woningen voor de belastbare basis. Zie bv. de studie van Vermeiren et al. (2019) over het monitaren van de urban sprawl.

De meerwaarde van extra variabelen over energetische kenmerken van de woning leek ook beperkt. Deze onderliggende kenmerken die de EPC-score bepalen zijn immers reeds mee opgenomen in de waardebeoordeling wanneer het EPC is toegevoegd in het model. Ook hier is de vraag indien deze coëfficiënt, dus de energieprestaties van een woning, mee opgenomen dienen te worden in de belastbare basis. Op die manier laat een hedonische prijsanalyse niet alleen toe om de waarde te bepalen van een woning, maar ook om sommige elementen al dan niet mee op te nemen in de berekening van de belastbare basis.

Opties van modelleren binnen de KNN-methode

Binnen de KNN-methode zijn er nog verschillende opties om verder te verfijnen. Zo zagen we dat voornamelijk voor locatie het interessant is om een trend mee op te nemen die weergeeft hoe de waarde van een locatie verandert doorheen de tijd. Voor stabiele trends te bepalen is het gebruik van stadsgewesten een optie, aangezien kleinere gemeenten veelal te weinig observaties hebben om een trend doorheen de tijd te bepalen.

Een ander verder te onderzoeken punt is eerder wiskundig en gaat na hoe een observatie beïnvloed wordt door de nearest neighbours. Dit kan modelmatig via de residuals (de schattingsfout als benadering van het niet-verklaarde deel van de locatie), waarbij ervan uitgegaan wordt dat de andere locatie-variabelen best apart gemeten worden. Maar dit kan ook door het opnemen van de locatie-effecten van de nearest neighbours (waar de residuals een deel van zijn). Andere elementen, zoals het bepalen van het aantal K nearest Neighbours, en het gewicht in functie van de afstand tot de observatie dienen dan herschat te worden. Meer onderzoek naar de ruimtelijke afhankelijkheid van de verschillende observaties en de link hiervan met de geografische informatie is in het algemeen interessant. Een groot deel wordt opgevangen door de KNN-methode, maar om deze ruimtelijke samenhang te verklaren (niet noodzakelijk om beter te schatten) zijn modellen waarbij deze ruimtelijke variabelen mee opgenomen worden in het model, zoals afstand tot het centrum, wel interessant.

BIJLAGEN

BIJLAGE 1 RESIDENTIEEL EN NIET-RESIDENTIËLE GEBOUWEN

Tabel B1 Totaal aantal residentiële gebouwen en het aandeel gebouwen waar eveneens een niet-residentiële gebouw op het grondperceel staat.

| | Classificatie-index | Niet-residentiële | | totaal |
|-----|--|-------------------|---------|-----------|
| 10 | Huis in een tuinderijk | 7% | 21,870 | 297,903 |
| 20 | Gekarakteriseerde hoeve | 0% | 205 | 65,233 |
| 30 | Villa | 2% | 16,734 | 727,673 |
| 31 | Bungalow | 3% | 9,293 | 345,653 |
| 32 | Fermette | 0% | 41 | 9,776 |
| 33 | Vakantieverblijf | 6% | 2,729 | 48,308 |
| 40 | Huis zonder bewoonbare kelder | 4% | 88,083 | 1,965,843 |
| 41 | Huis bel-étage | 6% | 3,423 | 59,553 |
| 50 | Huis met bewoonbare kelder | 14% | 3,875 | 27,904 |
| 60 | Huis met koetspoort als enige ingang | 13% | 461 | 3,637 |
| 70 | Huis met koetspoort en particuliere ingang | 10% | 563 | 5,601 |
| 80 | Huis zonder woonplaatsen op het gelijkvloers | 7% | 1,085 | 14,914 |
| 100 | Toebehorend aan één enkele eigenaar - Zonder lift | 7% | 3,749 | 53,544 |
| 101 | Wooneenheid - Zonder lift | 8% | 21,449 | 263,153 |
| 102 | Exploitatie-eenheid - Zonder lift | 9% | 2,866 | 32,364 |
| 103 | Garage, standplaats, parking, afzonderlijk gekadastreerd - Zonder lift | 6% | 6,399 | 115,045 |
| 104 | Diverse lokalen, afzonderlijk gekadastreerd - Zonder lift | 10% | 3,730 | 36,191 |
| 105 | Huis# | 9% | 382 | 4,272 |
| 110 | Toebehorend aan één enkele eigenaar - Met lift | 7% | 846 | 11,939 |
| 111 | Wooneenheid - Met lift | 8% | 48,045 | 599,267 |
| 112 | Exploitatie-eenheid - Met lift | 8% | 3,615 | 43,984 |
| 113 | Garage, standplaats, parking, afzonderlijk gekadastreerd - Met lift | 6% | 16,452 | 273,149 |
| 114 | Diverse lokalen, afzonderlijk gekadastreerd - Met lift | 7% | 6,714 | 98,375 |
| | totaal | 5% | 219,722 | 4,451,593 |
| | huizen | 4% | 145,633 | 3,523,690 |
| | appartementen | 8% | 74,089 | 927,903 |

Bron: AAPD, eigen verwerking

REFERENTIES

- Boogaerts, T., Damen, S. & Schildermans S. (2020). De staat van het kadastraal inkomen.
- Bracke, P. (2015). House Prices and Rents: Micro Evidence from a Matched Data Set in Central London. *Real Estate Econ.* 43 (2): 403–20.
- Campbell, S.D., Davis M.A., Gallin J., & Martin, R. (2009). What moves housing markets: A variance decomposition of the rent–price ratio, *Journal of Urban Economics*
- Damen, S., Vandesande, A., Bomans, K., Steenberghen, T., Van Balen, K., De Jaeger, S., ... Dugernier, M. (2017). Onderzoek naar de effecten van de erfgoedkarakteristieken van woningen en hun omgeving op de marktprijzen van woningen in Vlaanderen.
- Damen, S. (2019). Het effect van het EPC en energetische kenmerken op de verkoopprijs van woningen in Vlaanderen. Vlaams Energieagentschap, april 2019.
- Eisfeldt, A. and Demers, A.: 2018, Total returns to single family rentals. (NBER Working Paper No. 21804). Available at <https://www.nber.org/papers/w21804>.
- Eurostat (2017), Technical manual on Owner-Occupied Housing and House Price Indices, Eurostat
- Francke, M. and A. van de Minne. 2017. Land, Structure and Depreciation. *Real Estate Economics* 45(2): 415–451.
- Gallin, J. The Long-Run Relationship Between House Prices and Rents, <https://doi.org/10.1111/j.1540-6229.2008.00225.x>
- Goeyvaerts, G., & Buyst, E. (2019). Do market rents reflect user costs?. *Journal of Housing Economics*, 44, 112-130.
- Halket, J. and M.P.M. di Custozza. 2015. Homeownership and the Scarcity of Rentals. *Journal of Monetary Economics* 76: 107–123.
- Helgers R. (2016) Housing Prices and Housing Stock Composition: How Do We Value the Homes of our Neighbors? Essays on Local Housing and Real Estate Brokerage Markets, Dissertation presented to obtain the Degree of Doctor in Economics, KU Leuven
- Kryvobokov, M., Pradella, S. & Rosiers, F.D. Urban and Peri-Urban Residential Rental Markets in Wallonia: Similar or Different?. *Appl. Spatial Analysis* 13, 461–487 (2020). <https://doi.org/10.1007/s12061-019-09312-8>
- Mahieu, B., Heyndels, B., Bursens, J., Goeminne, S. & Smolders, C. (2012). Een analyse van de relatie tussen KI en woningprijzen in de Vlaamse centrumsteden. *Documentatieblad Federale Overheidsdienst Financiën*: 72(2).
- Mayer, M., Bourassa, S. C., Hoesli, M., & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European Real Estate Research*, 12(1), 134–150. <https://doi.org/10.1108/JERER-08-2018-0035>
- Reusens P, Vastmans F. & Damen S. (2022), The impact of changes in dwelling characteristics and housing preferences on house price indices before and during COVID-19, NBB working paper N° 406.
- Rifaad, Y., Sansen, J., & Ryckewaert, M. (2021). Studentenhuisvesting. Een verkenning van buitenlandse praktijkvoorbeelden. Leuven: Steunpunt Wonen, 38 p.
- Schildermans, S. & S. Damen (2021) Het schatten van de verkoopprijs van residentiële onroerende goederen, Retax (SBO, S005718N) gefinancierd door het Fonds voor Wetenschappelijk Onderzoek (FWO), 51 pag.

- Vanderstraeten L. and E. Van Hecke (2019), *Les régions urbaines en Belgique, Belgeo – Revue belge de géographie*, 1.
- Vastmans F., Buyst E., Helgers R. & Damen S. (2014), *Woningprijzen: woningprijs-mechanisme & marktevenwichten. De logica, nood en valkuilen van betaalbaarheid als woningprijs determinant*, Steunpunt Wonen, Leuven, p. 87.
- Vastmans F. (2016), *Een nieuwe rol voor de private huur. Deel 3: Een economische analyse van de huurmarkt*, Steunpunt Wonen, Leuven, 58 p.
- Vastmans, F. (2019), *De huurschatter, nieuwe resultaten versie 3, 2019*. Leuven: Steunpunt Wonen, p. 43.
- Vastmans, F. (2020). *De energieprestaties van de Vlaamse woningvoorraad. Cijfers en verklaringen*. Leuven: Steunpunt Wonen.
- Vastmans, F., & Dreesen, S. (2021). *Woningprijzen: algemene trends en regionale verschillen. Vaststellingen in Vlaanderen en verklaringen uit de literatuur van urban economics*. Leuven: Steunpunt Wonen, 146 p.
- Vastmans, F. (te verschijnen), *Het schatten van de huurprijs van woningen voor de basis van het kadastraal inkomen*, , Retax (SBO, S005718N) gefinancierd door het Fonds voor Wetenschappelijk Onderzoek (FWO)
- Verbeeck G. & Ceulemans W. (2015). *Analyse van de EPC-databank. Resultaten tot en met 2012*. Leuven: Steunpunt Wonen
- Verbrugge R., *The puzzling divergence of rents and user costs, 1980–2004*, *Rev. Income Wealth*, 54 (4) (2008), pp. 671-699
- Verbrugge, R., Poole R., (2010), *Explaining the rento inflation divergence, 1999–2007*, *Real Estate Econ.*, 38 (4), pp. 633-657
- Vermeiren et al. (2019), *Monetariseren van urban sprawl in Vlaanderen, uitgevoerd in opdracht van Departement Omgeving*.
- Winters, S. & K. Van den Broeck (2019). *Nadenken over een hervorming van de onroerende voorheffing*, Retax (SBO, S005718N) gefinancierd door het Fonds voor Wetenschappelijk Onderzoek (FWO)



**Research Foundation
Flanders**
Opening new horizons