

De invloed van NS-treinstations op huizenprijzen

Vak: Thesis Bachelor Economics
Major: Policy Economics
Begeleider: Dhr. Bas Jacobs
Tweede lezer: Dhr. Aart Gerritsen
Aantal pagina's: 44
Aantal woorden: 7733 (exclusief voorblad, referenties & appendix)
Datum: 22-06-2020
Universiteit: ERASMUS University Rotterdam
Faculteit: Erasmus School of Economics

Student Nummer	Achternaam	Voornaam
484198	Muntendam	Matthijs

Het eerste Nederlandse treinstation werd in 1839 geopend in Haarlem. Sindsdien zijn treinstations onderdeel van de openbare infrastructuur. Onderzoekers hebben reeds aangetoond dat voor de steden Amsterdam, Rotterdam & Enschede treinstations een positieve invloed hebben op de woningwaarde. In deze paper is met behulp van hedonische vraagtheorie onderzoek gedaan naar de invloed van NS-treinstations voor Nederlandse buurten. Een *Ordinary Least Squares regressie* is gebruikt om het causale effect van de gemiddelde afstand tot het treinstation op de gemiddelde woningprijzen te schatten. Er wordt een positieve coëfficiënt van 0,013 gevonden, die impliceert dat woningprijzen met 1,3% stijgen als de afstand tot het station met één kilometer toeneemt.

Inhoudsopgave

1. Inleiding	3
2. Theoretisch kader	5
2.1 Theorie.....	5
2.2 Literatuur.....	7
3. Data	10
3.1 Woningprijzen	10
3.2 Afstand tot treinstation.....	10
3.3 Kerncijfers van buurten	12
4. Veronderstellingen	14
4.1 Exogeniteit veronderstelling	14
4.2 Homoskedastisciteit van de error termen	17
4.3 Uitschieters van de observaties	17
5. Methode	18
6. Resultaten.....	21
7. Conclusie en Discussie.....	27
8. Referenties	29
9. Appendix.....	32
9.1 Figuren.....	32
9.2 Tabellen	34
9.3 STATA Codes.....	42

1. Inleiding

In Nederland is het mogelijk om te werken in Rotterdam en te wonen in Amsterdam, met de trein doe je er zo'n 30 minuten over om van Amsterdam naar Rotterdam Centraal Station te reizen. De gunstige Nederlandse infrastructuur geeft meer keuzevrijheid op de woningmarkt, echter wordt deze vrijheid beperkt door hoge woningprijzen. Ter illustratie, eind 2019 waren bestaande koopwoningen duurder dan ooit: een woning kostte in Nederland gemiddeld €308.000 (Van Bockxmeer, 2020). De stijging van de prijzen is positief voor huizenbezitters, maar een obstakel voor starters bij het kopen van een woning. Het is de omgekeerde wereld als je bedenkt dat in 2015 32% van de huishoudens in Nederland een fiscale hypotheekschuld hebben die hoger is dan de waarde van de eigen woning (Centraal Bureau voor de Statistiek, 2015). Toch is er verschil tussen regio's. In regio's waar de prijzen al hoog zijn, stijgen de prijzen minder hard (Van Bockxmeer, 2020).

Het is aannemelijk om te denken dat prijsverschillen tussen regio's worden bepaald door fundamentele verschillen in karaktereigenschappen van regio's. Bereikbaarheid per trein is zo'n mogelijk regionaal verschil. Utrecht heeft bijvoorbeeld de meeste kilometers spoor per vierkante kilometer en Friesland de minste kilometers spoor per vierkante kilometer (CBS, 2017). De Nederlandse Spoorwegen werkt samen met Nederlandse gemeenten aan het verbeteren van die bereikbaarheid. Medio 2019 werd op de grens van twee Zuid-Hollandse gemeenten een nieuw vervoersknooppunt opgeleverd boven de rijksweg A12 tussen Den Haag en Utrecht (Omroep West, 2019). We spreken hier over projecten met maatschappelijke opbrengsten, maar ook kosten.

Maatschappelijk gezien kunnen de uitkomsten van dit onderzoek een beter inzicht geven in de waarde van publieke voorzieningen. Essentiële vragenstukken zoals de terugverdientijd van deze investeringen kunnen hiermee gebaseerd worden op directe opbrengsten zoals treinkaartjes en op indirecte opbrengsten, zoals de waardeverhoging van woningen. Daardoor zouden op basis van dit onderzoek publieke investeringen, zoals in het openbaar vervoer lokaal toegeschreven kunnen worden. Oftewel wie winsten behaalt door de aanwezigheid van publieke goederen betaalt ook.

Ondanks dat het vermoeden bestaat dat er op wijk en buurt niveau voordelen zijn van de nabijheid van een treinstation, worden verbeteringen van de infrastructuur betaald met provinciaal belastinggeld. Nederland staat immers bekend om zijn goede infrastructuur, maar wat is de prijs hiervan? Om te onderzoeken wat de lokale voordelen zijn van een nabijgelegen station wordt gekeken naar de huizenmarkt. Wat betekent een treinstation voor de huizenprijzen en daarmee indirect voor de huizenbezitter. Dit leidt tot de volgende onderzoeksvraag:

Wat is de invloed van een NS-treinstation op de prijzen van koopwoningen in 2017 in Nederland?

Wetenschappelijk gezien kan deze paper bijdragen aan eerder onderzoek naar de relatie tussen huizenprijs en infrastructuur. Op niveau van 6-cijferige postcodes in de steden Amsterdam, Rotterdam en Enschede is gebleken dat een kortere afstand naar het station resulteert in een hogere woningwaarde (Debrezion, Pels & Rietveld, 2010). Waar het onderzoek van Debrezion et al. (2010) naar drie steden kijkt, heeft mijn onderzoek een onderzoeksgebied dat heel Nederland betreft, waardoor onderscheid gemaakt kan worden tussen provincies. Door te kijken naar de gemiddelde afstand tot het station wordt bijgedragen aan het blootleggen van specifieke determinanten die de waarde van een stuk grond of de prijs van een woning bepalen.

Deze paper bevat een theoretisch kader waarin op basis van voorgaande literatuur een hypothese opgesteld is. Daarna wordt in de sectie data besproken welke data gebruikt wordt. Vervolgens wordt in de sectie veronderstellingen inzichtelijk aan welke voorwaarden moet worden voldaan voor het uitvoeren van dit onderzoek. De methode bevat de uitvoering van een *Ordinary Least Squares* regressie. In het vervolg worden de uitkomsten van de uitgevoerde analyse in de sectie resultaten getoond. Tot slot wordt in de conclusie en discussie de onderzoeksvraag beantwoord en wordt het onderzoek in perspectief geplaatst.

2. Theoretisch kader

In deze sectie wordt eerst aan de hand van hedonische vraagtheorie uitgelegd hoe een treinstation de huizenprijzen kan beïnvloeden. Vervolgens wordt met behulp van de economische literatuur een hypothese opgesteld over het verband tussen treinstations en woningprijzen. Daarna wordt beargumenteerd welke andere variabelen belangrijk zijn om het effect van een treinstation op de huizenprijs te bepalen.

2.1 Theorie

Het onderzoeken van de vraag naar een product door de samengestelde kenmerken op te splitsen, ook wel hedonische vraagtheorie genoemd, is een geopenbaarde voorkeursmethode om de vraag naar een goed en daarmee de waarde ervan te schatten (Rosen, 1974). Dit maakt het mogelijk om waarde toe te schrijven aan karakteristieken van een locatie. Doordat consumenten een hoger nut ontleen aan de specifieke locatie eigenschappen stijgt de vraag op de ene locatie en daalt de vraag op de andere. Het is hierdoor mogelijk dat als je twee identieke objecten hebt op verschillende locaties er meer vraag is naar één van de objecten. Dit resulteert in verschillen tussen koop- en huurprijzen van huizen.

Al 190 jaar geleden is onderzoek gedaan naar de relatie tussen de locatie van vastgoed en de waarde van een object. Het blijkt dat er variatie bestaat in de waardes van boerenland, veroorzaakt door de toegankelijkheid van de marktplaats. Een betere toegankelijkheid van de marktplaats resulteert in lagere transportkosten, waardoor ook individuen met lagere opbrengsten een positief nut kunnen ontleen. Door positief nut voor meer individuen is er een hogere geaggregeerde vraag naar boerenland. Op deze manier kan verklaard worden dat land met dezelfde vruchtbaarheid, maar een andere locatie verschilt in waarde (Von Thünen, 1830).

De vraag naar huizen wordt daarnaast bepaald door de reële rentevoet. Individuen zijn bereid een huis te kopen totdat de marginale opbrengsten van een woning, gelijk zijn aan de marginale kosten in hun nutsfunctie. De reële rentevoet bevindt zich aan de kostenkant, dat betekent dat een stijging resulteert in een lager nut en minder vraag. Het aanbod van woningen, wordt bepaald door de voorraad van huizen volgens een productie relatie, deze is constant op de korte termijn (Poterba, 1984). Door de hogere vraag naar woningen vanwege een lagere reële rentevoet stijgen de prijzen op de markt waar vraag en aanbod samenkomen.

Een rente analyse van Alonso en Muth bevestigt de nutstheorie (Alonso, 1964 & Muth, 1969, zoals beschreven in Debrezion, Pels & Rietveld, 2010), er wordt gevonden dat de rentevoet afneemt als de afstand naar het centrale zakencentrum toeneemt. De langere afstand naar het centrale zakencentrum leidt tot meer reistijd naar het werk en daardoor minder vrije tijd. Doordat vrije tijd schaars is, wordt reistijd meegenomen in de nutsfunctie van ieder individu.

Dat de rentevoet verband houdt met de afstand naar het centrale zakencentrum kan uitgelegd worden als je bedenkt dat in de woningmarkt zelf onderscheid valt te maken tussen de eigenaar van een woning en de bewoner van een woning. De rentabiliteit van een woning wordt hierbij bepaald door de huurprijs voor huisvesting op de woningmarkt. Voor de bewoner reflecteert de huurprijs het marginale nut van huisvesting. Daarentegen meet de huurprijs de rentabiliteit voor de eigenaar (Van Ewijk, Jacobs & De Mooij, 2007). Doordat betere locatie eigenschappen de kosten van huisvesting beïnvloeden is de bewoner bereid de eigenaar meer huur te betalen. De eigenaar ziet hierdoor de waarde van zijn vastgoed stijgen. In het geval dat de bewoner ook de eigenaar van vastgoed is, verdwijnt simpelweg de huurprijs betaling, maar het uiteindelijke effect is vergelijkbaar wanneer een huis verkocht wordt. Doordat de nutsontlening aan het vastgoed groter geworden is, is een nieuwe eigenaar bereid maandelijks hogere kosten te aanvaarden en waardoor hij een hogere koopsom wil betalen.

2.2 Literatuur

Hedonische vraagtheorie is eerder gebruikt om onderzoek te doen naar onder andere de waarde van luchthavens en scholen. Het blijkt dat geluid van een luchthaven geen significant negatief effect heeft op huizenprijzen, echter ervaren bewoners positief nut van de aanwezigheid van de luchthaven, vanwege het reisgemak en de werkgelegenheid (Lipscomb, 2003). Verder blijkt dat verschillen in schoolresultaten leiden tot een hogere huizenprijs. De schoolresultaten worden geassocieerd met de kwaliteit van een school. Een 10 procentpunt verschil van standaarddeviatie in een gemiddeld schoolresultaat, leidt tot een 10 procentpunt stijging van de huizenprijs. Een kanttekening is dat een regressie discontinuïteit ontwerp de suggestie wekt dat de bevolking tussen districten in meerdere aspecten verschilt (Kane, Riegg & Staiger, 2006).

Onderzoekers hebben tot nu toe geen eenduidig antwoord op de vraag wat de afstand van een treinstation doet met de waarde van woningen. Onderzoek uitgevoerd door Dewees (1976) laat zien dat de aanwezigheid van een metrohalte binnen een radiaal gelijk aan één derde van een mijl de huren laat toenemen. Als toevoeging hierop is gebleken dat de afstand van een metrostation een statistisch significant effect heeft op de waarde van grond (Damm, Lerman, Lerner-Lam & Young, 1980). Grass (1992) gaat nog verder en vindt een direct effect van een nieuw metrostation op de prijzen van woningen. De bovengenoemde literatuur spreekt voornamelijk over een kortere afstand tot een station die leidt tot een hogere woningprijs, terwijl het niet ondenkbaar is dat een treinstation en vooral het bijbehorende spoor negatieve externe effecten met zich meebrengen, zoals geluidsoverlast. Nadat in Singapore bekend werd gemaakt dat de overheid iets ging doen aan de negatieve externe geluidseffecten van het spoor, stegen prijzen van huizen binnen een 400 meter grens relatief met 3,5 procent ten opzichte van huizen buiten diezelfde grens (Diao, Qin & Sing, 2015).

In tegenstelling tot het bovenstaande onderzoek blijkt uit Nederlands onderzoek op niveau van 6-cijferige postcodes in de steden Amsterdam, Rotterdam en Enschede, dat voor het dichtstbijzijnde station een kortere afstand tot het treinstation een positief effect heeft op de woningwaarde, echter blijkt dit effect nog sterker als gekeken wordt naar het meest gebruikte station in plaats van de afstand tot het station (Debrezion, Pels & Rietveld, 2010). Ondanks dat mijn onderzoek kijkt naar 4-cijferige postcodes, kan met het in acht nemen van al het bovenstaande de volgende hypothese worden opgesteld:

Hypothese: Een kleinere afstand van een NS-treinstation zorgt voor hogere woningprijzen in 2017.

Ondanks dat hedonische vraagtheorie kan bijdragen aan de waardebeoordeling van vastgoed, blijft het lastig om het daadwerkelijke effect van een bepaalde eigenschap te analyseren. Dit wordt veroorzaakt door twee problemen.

Ten eerste is lang niet alle data die daarvoor benodigd is beschikbaar. Ten tweede moet je zeker van zijn dat alle kenmerken of gedragingen zijn meegenomen die zowel de onafhankelijke als de afhankelijke variabele beïnvloeden. Dit tweede probleem kan verminderd worden door met behulp van bestaande literatuur mogelijke controle variabelen te vinden. In de volgende literatuur worden mogelijke controle variabelen genoemd.

Eerder onderzoek van Mathur (2008) heeft laten zien dat de kortere reistijd naar het zakencentrum een positief effect heeft op huizenprijzen. Een aanname hierbij is dat ieder belangrijk overstappunt gekoppeld kan worden aan een zakencentrum, maar dat niet ieder zakencentrum verbonden is aan een belangrijk overstappunt. De afstand tot een belangrijk overstappunt houdt op zijn beurt weer verband met de afstand tot het treinstation.

Behalve de toegankelijkheid, ofwel afstand tot het treinstation, heeft ook toegang tot de snelweg een aandeel in de bereikbaarheid van een buurt. Sterker nog, het blijkt dat de toegang tot de snelweg gezien kan worden als een substitutiegoed en concurrent voor toegang tot het spoor netwerk (Voith, 1993). Verder blijkt dat wanneer een snelweg binnen 100 meter ligt, er een significant negatief effect is op de huizenprijs (Debrezion, Pels & Rietveld, 2010).

Ondanks dat we bij publiek transport vaak te maken hebben met organisaties zonder winstoogmerk is het niet gezegd dat er in deze sector geen efficiëntie afwegingen worden gemaakt. Afwegingen maken die efficiënt zijn betekent dat openbaar vervoer aanbieders zoveel mogelijk mensen willen bereiken met zo min mogelijk kosten. Currie (2010) heeft onderzoek gedaan naar fricties tussen vraag en aanbod in de markt van publiek transport. Het onderzoek vindt significante verschillen tussen de aangeboden service en de sociale vraag. Er komt naar voren dat vraag en aanbod elkaar vinden in de dichtbevolkte binnen wijken, maar in mindere mate in de buitenwijken van Melbourne. Bevolkingsdichtheid blijkt verder een statistisch significant effect te hebben op de huisvestingskosten, die bestaan uit huren, rente en schuldbetalingen. De bevolkingsdichtheid en huisvestingskosten hebben een zeer lage, maar positieve elasticiteit (Palm, Gregor, Wang & McMullen, 2014).

Opvallend is dat investeringen in het publieke transport de vraagfrictie rondom zakencentrum kunnen verminderen doordat huishoudens zich nu vestigen rondom het station in plaats van in het centrum (Fajarang, 1994), er is hier nu immers ook een goede bereikbaarheid gecreëerd.

Uit onderzoek van Bowes & Ihlanfeldt (2001) blijkt dat er sprake is van een relatie tussen het treinstation en de woningprijzen, echter verandert de relatieve waarde van de effecten als er tussen buurten verschillen zijn in de demografische segmentatie. In buurten waar een lager gemiddeld inkomen is, is de waarde die bewoners aan het station koppelen groter, hierdoor is de coëfficiënt die de relatie tussen de afstand tot het station en woningprijzen aangeeft verschillend. Dit is te verklaren doordat lagere-inkomensgroepen per definitie meer gebruik maken van het openbaar vervoer, omdat het bezitten van een auto wellicht te duur is.

Overheden spelen een rol in het aanbod van publiek transport (Banister, 2005 & Ingram & Liu, 1999, zoals beschreven in Buehler, 2011). Je kan hier denken aan de macht van lokale overheden om te lobbyen voor een nieuw station. Sommige lokale bestuurders hebben betere relaties met de provinciale bestuurders om financiële steun te krijgen voor projecten.

Daarbij komt dat prijzen van woningen verschillen tussen gemeenten, veroorzaakt door de kwaliteit van voorzieningen, werkgelegenheid of simpelweg reputatie. In een meervoudige regressie kunnen zulke tijds-irrelevante verschillen door middel van gemeente vaste effecten meegenomen worden.

3. Data

In deze paper wordt gebruik gemaakt van cross-sectie data. Deze data bevatten onder andere de gemiddelde woningwaarde voor 9866 buurten in 2017 (CBS, 2020). Daarnaast is er informatie voor iedere woonwijk over de nabijheid van voorzieningen (CBS, 2019). De gemiddelde afstand tot het treinstation is de onafhankelijke variabele waarvan ik het effect op de gemiddelde woningprijzen wil weten. De gemiddelde afstand tot het treinstation is beschikbaar voor 12 583 verschillende buurten in Nederland.

Tabel 1: Beschrijvende statistieken van variabelen in 2017.

Variable	Observaties	Gemiddelde	Std. Dev.	Min	Max
woningprijs	9,866	249.18	114.70	53	1633
afstand tot treinstation	12,583	6.34	6.68	.2	60.3
afstand tot overstappunt	12,583	12.90	9.85	.2	71.9
afstand tot hoofdweg	12,583	1.88	1.80	.1	46.4
bevolkingsdichtheid	12,610	3037.29	3966.28	0	35921
inkomen	10,857	26.92	7.37	9.3	142

3.1 Woningprijzen

De variabele $WPR_{i,j,2017}$ geeft voor iedere buurt i in gemeente j de gemiddelde woningprijs in duizenden euro's in 2017. De bovenstaande tabel laat zien dat deze beschikbaar is voor 9866 buurten. De gemiddelde woningwaarde is hierbij gelijk aan €249.000,-. Waarbij de buurt met de laagste waarde een woningwaarde heeft van €53.000,-. De buurt met de hoogste waarde in deze dataset een woningwaarde heeft van 1,6 miljoen euro.

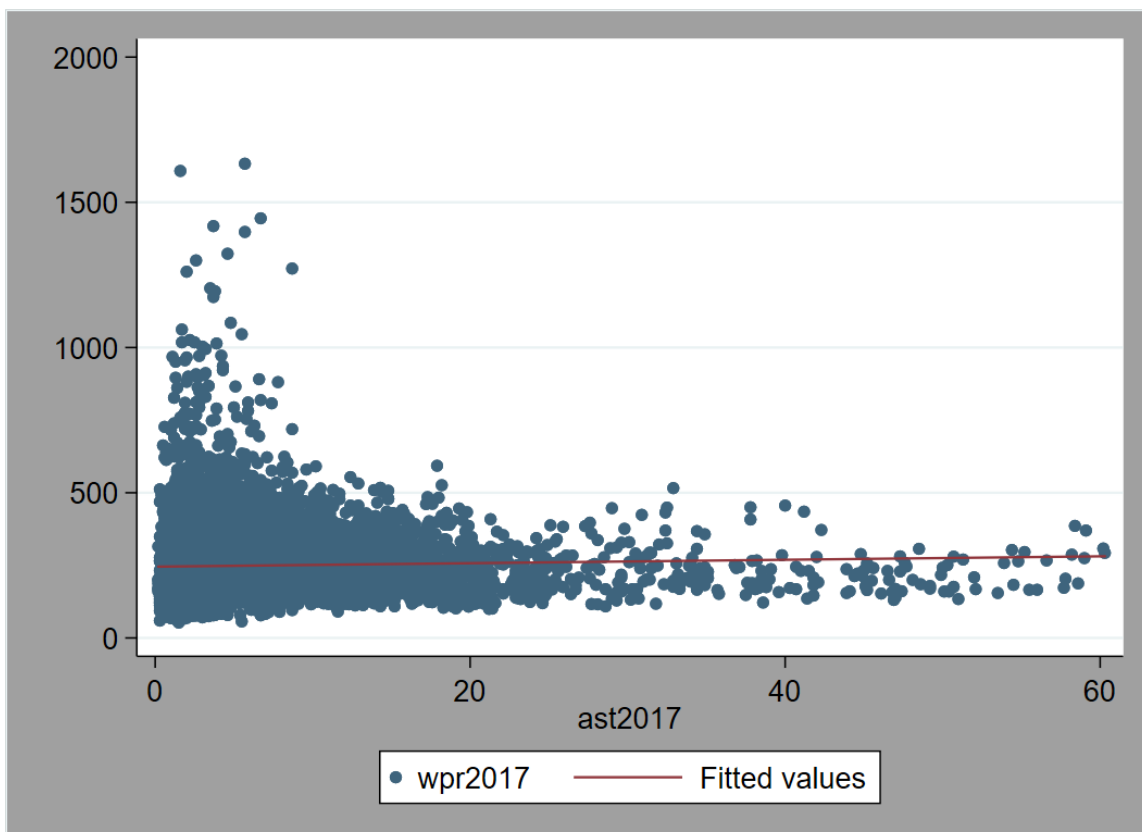
In figuur 1 in de appendix is te zien dat de gemiddelde woningwaarde neigt naar een normale distributie, hoewel er ook een lichtelijk naar rechts afwijkende distributie zichtbaar is.

3.2 Afstand tot treinstation

De variabele $AST_{i,j,2017}$ geeft voor 12 583 Nederlandse buurten de gemiddelde afstand in kilometers tot het dichtstbijzijnde NS-treinstation. Deze afstand is nauwkeurig tot op 1 decimaal, vandaar dat ook af te lezen valt dat de buurt met de kleinste afstand tot het station gelijk is aan 0,2 km. De grootste afstand ter vergelijking bedraagt 60.2 km. De gemiddelde afstand tot een treinstation in Nederland bedraagt 6,3 km.

Uit figuur 2 in de appendix blijkt dat de gemiddelde afstand tot het treinstation een verdeling heeft met een afwijking naar rechts. Dit wil zeggen dat er in Nederland relatief veel buurten zijn met een kleine afstand tot het station en minder buurten met een grote afstand. Een verklaring hiervoor is dat stations in het centrum worden neergezet, zodat de afstand om er te komen voor iedereen klein is. We praten hier over een range van 0,1 tot 20 km.

Het onderstaande scatterplot geeft een indruk van de relatie tussen afstand tot het treinstation en gemiddelde woningwaarde in 2017.



Figuur 3: Scatterplot van woningprijzen uitgezet tegen afstand tot het treinstation.

Figuur 3 lijkt een positieve relatie weer te geven tussen afstand tot het treinstation en de gemiddelde woningprijs beiden in 2017. Het gaat hier echter om een partiële correlatie zolang er niet gecontroleerd wordt voor eigenschappen van buurten die zowel de afstand tot het treinstation als ook de woningprijs beïnvloeden. Verder moet uit het scatterplot opgemerkt worden dat de werkelijke woningprijzen niet even ver van de geschatte waardes liggen voor elke afstand tot het treinstation. De gevolgen hiervan worden verder uitgewerkt in de sectie veronderstellingen.

3.3 Kerncijfers van buurten

Tot slot zijn voor het uitsluiten van *Omitted Variable Bias* verschillende kerncijfers van buurten nodig. Respectievelijk zijn de variabelen $ADO_{i,j,2017}$, $ASW_{i,j,2017}$ & $BVD_{i,j,2017}$, ofwel afstand tot dichtstbijzijnde overstappunt, afstand tot oprit hoofdverkeersweg & bevolkingsdichtheid beschikbaar voor ten minste 12 583 buurten in Nederland over 2017. Voor $ADO_{i,j,2017}$ & $ASW_{i,j,2017}$ geldt dat ze beiden gemeten zijn in kilometers tot op 1 decimaal nauwkeurig. De kleinste afstand tot het dichtstbijzijnde overstappunt bedraagt 0,2 kilometer. Daartegenover staat een maximale afstand van 71,9 kilometer. De gemiddelde afstand tot een overstappunt is 9,8 kilometer.

De gemiddelde afstand tot een hoofdverkeersweg is met 1,9 km aanzienlijk lager, dan de aan het treinstation gerelateerde variabelen. De buurt met de kleinste afstand tot een hoofdverkeersweg vertegenwoordigt een afstand van 0,1 kilometer. Daartegenover is een buurt met een afstand van 46,4 km om bij een hoofdverkeersweg te komen, de buurt met de grootste waarde voor deze variabele.

De variabele $BVD_{i,j,2017}$ geeft het aantal inwoners per km² weer. Respectievelijk zijn de buurten met het minimale en maximale bevolkingsdichtheid 0 inwoners per km² en 35 921 per vierkante kilometer. Het minimum van 0 is te verklaren met buurten waar geen woningen staan, zoals een sportpark. In mijn meervoudig regressie model zorgt dit niet voor problemen aangezien de variabele woningprijs in dit geval als missend is gerapporteerd. De gemiddelde bevolkingsdichtheid in een Nederlandse buurt ligt op 3037 inwoners per vierkante kilometer.

Een andere variabele is $INK_{i,j,2017}$, de variabele geeft het gemiddelde inkomen in € (x1000) per inwoner aan. Dat betekent dat het gemiddelde inkomen per inwoner lager is als er minder mensen werken in een buurt. Deze variabele is voor 10 857 buurten beschikbaar. Het gemiddelde inkomen in een Nederlandse buurt is €26 000,-. Het laagste gemiddelde buurtinkomen is gelijk aan €9 300,-, terwijl het maximum €142 000,- is. Tabel 2 in de appendix geeft voor alle variabelen de volledige naam en eenheid weer.

Tot slot wordt gebruik gemaakt van gemeente vaste effecten om zo te controleren voor alle tijdsirrelevante factoren die op gemeenteniveau zowel de afstand tot het treinstation als de woningprijs beïnvloeden. Voor het uitvoeren deze controle variabelen bevindt zich in de dataset voor iedere buurt de naam van de bijbehorende gemeente. Er zijn in totaal 388 verschillende gemeenten in Nederland, hiervoor worden 387 dummy's toegevoegd die het verschil meten ten opzichte van één gemeente.

4. Veronderstellingen

In deze sectie worden veronderstellingen besproken die nodig zijn om het effect van afstand tot het treinstation op woningprijzen beter te bepalen met behulp van een *Ordinary Least Squares* regressie.

4.1 Exogeniteit veronderstelling

Het schenden van deze aanname kan komen door drie verschillende bronnen:

- *Zero Conditional Mean*
- Incorrecte vorm van variabelen
- Simultane causaliteit

Zero Conditional Mean

De *Zero Conditional Mean* veronderstelling impliceert dat de error-term ε onafhankelijk is van de onafhankelijke variabele (afstand tot het treinstation). De verwachte waarde van de error term, moet het hetzelfde zijn ongeacht de afstand tot het treinstation:

$$E(\varepsilon_{i,j,2017} | AST_{i,j,2017}) = 0$$

Als de *Zero Conditional Mean* veronderstelling wordt gehaald, geldt het volgende:

$$\text{corr}(\varepsilon_{i,j,2017}, AST_{i,j,2017}) = 0$$

Deze expressie geeft aan dat er geen correlatie bestaat tussen de afstand tot het station en de errorterm. Er mag geen relatie zijn tussen de onafhankelijke variabele en andere variabelen, die zich uiten via de errorterm. Dit kan opgelost worden door de variabelen vanuit de errorterm op te nemen in de regressievergelijking, indien ze de afhankelijke variabele (gemiddelde woningprijs in 2017) beïnvloeden.

Aangezien uit de literatuur blijkt dat onafhankelijke variabele niet volledig exogeen is, wordt een aantal correlaties onderzocht. Op buurtniveau zijn er variabelen die zowel de afstand tot het treinstation als de woningprijzen beïnvloeden, de gemeente waarin een buurt zich bevindt is daarvan een voorbeeld. Een gemeente in de randstad heeft hogere woningprijzen en een beter ontwikkeld openbaar vervoer netwerk vergeleken met een gemeente in het oosten van Nederland.

Uit tabel 3 in de appendix blijkt dat de variabelen afstand tot het station en afstand tot het dichtstbijzijnde overstappunt een positieve relatie hebben met een correlatiecoëfficiënt van 0,64. Hoewel de relatie met de afstand tot een hoofdverkeersweg ook positief is, geeft de coëfficiënt van 0,17 een zwakkere relatie weer. De bevolkingsdichtheid daarentegen heeft een negatieve relatie met de afstand tot het treinstation, de coëfficiënt is gelijk aan -0,30. Dit lijkt betrouwbaar als je er van uit gaat dat een station alleen aanwezig is als het ook daadwerkelijk rendabel is voor de Nederlandse Spoorwegen. Een hogere bevolkingsdichtheid betekent automatisch meer vraag en daarom de incentive om een station te bouwen dat dichterbij is. Uit dezelfde tabel komt naar voren dat de afstand tot het treinstation en het gemiddelde inkomen in een buurt nagenoeg niet gecorreleerd zijn met elkaar. De coëfficiënt is gelijk aan -0,04, er is in dit geval geen relatie tussen deze variabelen. Samenvattend blijkt uit bovenstaande resultaten dat de variabelen afstand tot dichtstbijzijnde overstappunt, afstand tot hoofdverkeersweg en bevolkingsdichtheid gecorreleerd zijn met de onafhankelijke variabele. We spreken hier over potentiële controle variabelen. Het gemiddelde inkomen lijkt geen invloed te hebben op de afstand tot het treinstation.

Ondanks dat bovenstaande controle variabelen *Omitted Variable Bias* voorkomen, blijft het lastig om volledig aan de exogeniteit aanname te voldoen. De *Conditional Independence Assumption* stelt namelijk dat conditioneel op alle controle variabelen de onafhankelijke variabele random moet zijn. Ik kan dit niet volledig garanderen, doordat niet bekend is of er controle variabelen zijn die niet in de dataset zijn opgenomen. Hierdoor blijft het lastig te zeggen of er daadwerkelijk sprake is van een causaal verband. Echter moet aangetekend worden dat er relatief weinig variabelen zijn die de afstand naar het treinstation beïnvloeden.

Wanneer de variabelen worden toegevoegd moet economisch te beargumenteren zijn dat het ook daadwerkelijk controle variabelen zijn. Het moet zo zijn dat de controle variabelen bepaald zijn voordat de onafhankelijke variabele bepaald wordt. Als je variabelen opneemt in de regressie die beïnvloed worden door de afstand tot het station heb je te maken met mogelijke *mechanisms* en *colliders*. Een *mechanism* is een variabele die door de afstand tot het treinstation beïnvloed wordt, maar ook effect heeft op de woningprijs. Een *collider* wordt door zowel de afstand tot het treinstation als de woningprijs beïnvloed. *Mechanisms* en *colliders* zorgen respectievelijk voor selectiebias en het observeren van een onechte relatie.

Incorrecte vorm van variabelen

Er is sprake van meetfouten als variabelen incorrect zijn opgenomen in de regressie. Door het gebruik van voorgaande literatuur is het onwaarschijnlijk dat variabelen incorrect worden opgenomen in de regressie. Er lijkt geen sprake te zijn van kwadratische variabelen en interactie-effecten tussen variabelen.

Simultane causaliteit

Er is sprake van simultane causaliteit als de afhankelijke variabele wordt beïnvloed door de onafhankelijke variabele en omgekeerd. Een *Granger causality test* kan eventuele simultane causaliteit uitsluiten of bevestigen. De test onderzoekt of de vertraagde variabelen de woningprijzen beïnvloeden en aan de andere kant of de vertraagde variabelen de afstand tot het treinstation beïnvloeden. Als een vertraagde variabele een significant effect heeft op de afhankelijke variabele suggereert dit dat deze variabele eerder is bepaald dan afhankelijke variabele, hierdoor weten we de richting van het effect. Dat de vertraging van de afstand tot het treinstation een significant effect heeft, is onwaarschijnlijk aangezien deze nauwelijks veranderd over tijd. Het uitvoeren van twee regressies ziet er als volgt uit:

$$WPR_{i,j,2017} = a + \rho WPR_{i,j,2016} + \tau AST_{i,j,2016} + \varepsilon$$

&

$$AST_{i,j,2017} = \mu + \phi AST_{i,j,2016} + \lambda WPR_{i,j,2016} + \eta$$

Wanneer blijkt dat τ significant is, maar λ niet, betekent dit dat de afstand tot het treinstation de woningprijs beïnvloedt, maar niet omgekeerd. Voor het uitvoeren van de test is een additionele dataset gebruikt die achtereenvolgens de variabelen, $WPR_{i,j}, AST_{i,j}, ADO_{i,j}, ASW_{i,j}, BVD_{i,j}$ en $INK_{i,j}$ bevat voor de jaren 2016 en 2017. De variabelen hebben exact dezelfde kenmerken als degene benoemd in de sectie data met het verschil dat er nu sprake is van panel-data in plaats van cross-sectie.

In tabel 4 in de appendix blijkt dat τ significant is en λ niet. Hierdoor is de afstand tot het treinstation Granger-causaal voor de woningprijzen, maar woningprijzen niet Granger-causaal voor de afstand tot het treinstation. Uit het tweede model in tabel 4 in de appendix blijkt tevens dat de afstand tot het treinstation over de jaren nagenoeg constant is doordat de afstand tot het treinstation in 2016 een perfecte voorspeller is voor de afstand tot het treinstation in 2017.

4.2 Homoskedasticiteit van de error termen

De error termen in de regressie moeten allen dezelfde variantie hebben.

$$Var(\varepsilon|AST_{2017}) = \sigma^2$$

Uit figuur 3 in de sectie data moet opgemerkt worden dat de werkelijke woningprijzen niet even ver van de geschatte waardes liggen voor elke afstand tot het treinstation. Dit impliceert dat er sprake is van heteroskedasticiteit. Het feit dat er geen constante variatie van de errorterm is zorgt ervoor dat bij het regresseren gebruik gemaakt moet worden van heteroskedastisch-robuste standaardfouten.

4.3 Uitschieters van de observaties

Wanneer er in een regressie model grote uitschieters worden meegenomen, is het model onnauwkeurig. *OLS* is immers gevoelig voor uitschieters, waardoor de coëfficiënten beïnvloed worden. Ondanks dat uit figuur 3 in de sectie data blijkt dat er niet veel grote uitschieters zijn, neem ik het logaritme van de gemiddelde woningprijzen. Grote uitschieters hebben bij het nemen van de logaritme minder effect op de coëfficiënten. Extreme waardes worden door deze handeling minder extreem.

5. Methode

Om de hypothese te onderzoeken wordt gebruik gemaakt van een kwantitatieve onderzoeksmethode. Er wordt gebruik gemaakt van *Ordinary Least squares* regressies om te achterhalen of een kleinere afstand tot het treinstation resulteert in hogere woningprijzen. Voor het uitvoeren van de regressie wordt van de woningprijzen de logaritme genomen. Hiermee worden buurten met een hoge gemiddelde woningprijs minder extreem. Het meervoudig log-lineair regressie model kan hierbij als volgt worden weergegeven:

$$\ln WPR_{i,j,2017} = a + B_1 AST_{i,j,2017} + B_2 ADO_{i,j,2017} + B_3 ASW_{i,j,2017} + B_4 BVD_{i,j,2017} + B_5 INK_{i,j,2017} + \sum_{j=2}^{J=388} \gamma GEM_j + \varepsilon_{i,j,2017}$$

In deze regressie is de onafhankelijke variabele $AST_{i,j,2017}$ de gemiddelde afstand tot het NS-treinstation in buurt i gemeente j in 2017. De afhankelijke variabele $\ln WPR_{i,j,2017}$ is het logaritme van de gemiddelde prijs van een woning in buurt i in gemeente j in 2017. Dit betekent dat de coëfficiënt B_1 de procentuele verandering van de woningprijs weergeeft als de gemiddelde afstand tot het station verandert.

Aangezien ik in dit onderzoek geïnteresseerd ben in het causale effect van $AST_{i,j,2017}$ op $\ln WPR_{i,j,2017}$ is het belangrijk om in gedachten te houden dat het uitbreiden van de regressie verband houdt met *The Zero Conditional Independence Assumption*. Deze is al eerder aan bod gekomen in de sectie veronderstellingen. Het toevoegen van variabelen is in dit geval niet vanwege hun voorspellende functie voor de afhankelijke variabele, maar om als controle-variabelen het effect zo betrouwbaar mogelijk te maken.

De controle-variabelen worden stuk voor stuk ingevoegd en als blijkt dat ze helpen het causale verband beter weer te geven worden ze in het definitieve model opgenomen. Voor alle controle variabelen geldt dat ze nauwkeurig zijn voor iedere buurt i in gemeente j in 2017. De dummy-variabelen voor gemeente vaste effecten worden weergegeven door de operator over GEM_j . Er zijn in Nederland in totaal 388 verschillende gemeenten.

De eerste controle variabele die toegevoegd wordt is gemiddelde afstand tot het dichtstbijzijnde overstappunt, ofwel $ADO_{i,j,2017}$. De aanname over het dichtstbijzijnde overstappunt en het centrale zakencentrum in de sectie theoretisch kader maakt het mogelijk om het teken van de coëfficiënt te bepalen. Aangezien uit eerder onderzoek blijkt dat een kortere afstand tot het zakencentrum, leidt tot een hogere woningwaarde, verwacht ik dat de coëfficiënt B_2 negatief is. Tegelijkertijd wordt verondersteld dat als de gemiddelde afstand tot het dichtstbijzijnde overstappunt afneemt, ook de gemiddelde afstand naar het treinstation afneemt. Vanwege deze positieve correlatie en de verwachte negatieve coëfficiënt is eerder geschatte coëfficiënt B_1 hoogstwaarschijnlijk lager dan de werkelijke coëfficiënt.

De tweede controle variabele die toegevoegd wordt, is de gemiddelde afstand tot een hoofdverkeersweg, ofwel $ASW_{i,j,2017}$. Uit eerder onderzoek komt naar voren dat een goede bereikbaarheid per auto gezien kan worden als een substituuut voor openbaar vervoer. Echter leiden negatieve externe effecten van een snelweg ook tot een lagere woningwaarde. Dit leidt tot de verwachting dat de coëfficiënt B_3 positief is. Het is verder aannemelijk dat een kortere afstand naar het openbaar vervoer verband houdt met een kortere afstand naar de hoofdverkeersweg om substituten van elkaar te kunnen zijn. Vanwege de positieve correlatie en verwachte positieve coëfficiënt is de eerder geschatte coëfficiënt B_1 hoogstwaarschijnlijk hoger dan de werkelijke coëfficiënt.

De derde controle variabele die toegevoegd wordt, is de bevolkingsdichtheid per vierkante kilometer, ofwel $BVD_{i,j,2017}$. Uit eerder onderzoek komt naar voren dat bevolkingsdichtheid een klein positief significant effect heeft op de huisvestingskosten. Ik verwacht daarom dat de coëfficiënt van bevolkingsdichtheid B_4 positief is. De relatie tussen bevolkingsdichtheid en afstand tot het treinstation lijkt negatief te zijn, zeker als je bedenkt dat het aanbod van transport in binnen wijken verschilt van buiten wijken. Het is geen onlogische gedachte dat op dichtbevolkte plekken voor verschillende redenen meer publiek transport aanwezig is. Echter bestaat de mogelijkheid dat afstand tot het treinstation een voorspeller is voor bevolkingsdichtheid (Fajarang, 1994). Dit betekent dat wanneer bevolkingsdichtheid wordt toegevoegd, we te maken hebben met een mechanisme in plaats van een controle variabele. Een mechanisme introduceert bias, daarom moeten mechanismes niet in de regressie opgenomen worden. Indien ik bevolkingsdichtheid toevoeg, zorgen de negatieve correlatie en

verwachte positieve coëfficiënt dat de eerder geschatte coëfficiënt B_1 hoogstwaarschijnlijk hoger is dan de werkelijke coëfficiënt.

De vierde controle variabele die toegevoegd wordt is het gemiddelde inkomen per inwoner, ofwel $INK_{i,j,2017}$. Uit eerder onderzoek komt naar voren dat in buurten met een lager inkomen het openbaar vervoer meer waarde heeft. Hier is echter niet mee gezegd dat inkomen ook daadwerkelijk de afstand naar het treinstation beïnvloedt. Uit tabel 3 in de sectie veronderstellingen blijkt er nauwelijks correlatie te zijn met afstand tot het treinstation. Daarnaast moet worden opgepast, want het is aannemelijker dat woningprijzen het gemiddelde inkomen, dat nodig is om een hypotheek te krijgen, beïnvloeden in plaats van dat het inkomen de gemiddelde woningprijs beïnvloedt. Doordat er nauwelijks correlatie is en de afhankelijke variabele mogelijk de controle variabele beïnvloedt, is er niets te zeggen over de eerder geschatte coëfficiënt B_1 ten opzichte van de werkelijke coëfficiënt.

Het eerdere onderzoek naar inkomen, openbaar vervoer en woningprijzen maakt het echter wel interessant om naar de coëfficiënt B_1 te kijken voor verschillende inkomensgroepen. Dit kan gedaan worden de regressies uit te voeren conditioneel op $INK_{i,j,2017}$.

De laatste controle variabelen die toegevoegd worden, zijn de dummy variabelen voor gemeente vaste effecten, ofwel GEM_j . Er zijn in Nederland 388 gemeenten en iedere buurt behoort tot een bepaalde gemeente. Deze 387 dummy variabelen nemen de waarde één aan als wijken in de desbetreffende gemeente liggen, anders geven ze 0 weer. De dummy's controleren voor alle verschillen tussen gemeenten die zowel afstand tot het treinstation als ook de gemiddelde huizenprijzen beïnvloeden en constant zijn over de tijd.

6. Resultaten

Deze sectie bevat de resultaten van verschillende log-lineaire regressie modellen. Deze modellen zijn zowel enkelvoudig als meervoudig, daarbij is aangegeven of er gemeente vaste effecten opgenomen zijn. De controle variabelen die besproken zijn in voorgaande secties worden stuk voor stuk toegevoegd. Gebaseerd op voorgaande literatuur, assumpties van *Ordinary Least Squares* en de gevonden resultaten wordt bepaald of de controle variabele in de uiteindelijk regressie wordt gebruikt om het effect van afstand tot het treinstation op woningprijzen te bepalen.

Tabel 5: Lineaire regressieschattingen van de determinanten van woningwaarde in 2017.

	(1) logaritme van woningprijs	(2) logaritme van woningprijs	(3) logaritme van woningprijs	(4) logaritme van woningprijs	(5) logaritme van woningprijs
afstand tot treinstation	0.00452*** (0.000558)	0.00604*** (0.000705)	0.00578*** (0.000707)	0.00434*** (0.000630)	0.00223*** (0.000516)
afstand tot overstappunt		-0.00158*** (0.000520)	-0.00159*** (0.000518)	-0.00594*** (0.000519)	0.000480 (0.000387)
afstand tot hoofdweg			0.00767*** (0.00220)	0.0131*** (0.00207)	0.00676*** (0.00157)
bevolkings- dichtheid				-0.0000299*** (0.00000155)	-0.0000185*** (0.00000100)
inkomen					0.0391*** (0.00103)
constante	5.408*** (0.00556)	5.418*** (0.00709)	5.406*** (0.00758)	5.571*** (0.0107)	4.426*** (0.0296)
Vaste effecten	NEE	NEE	NEE	NEE	NEE
N	9860	9860	9860	9860	9820

Standaardfouten tussen haakjes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Het regresseren van de onafhankelijke variabele gemiddelde afstand tot het treinstation op de afhankelijke variabele de logaritme van de gemiddelde woningprijs in 2017 levert een positieve coëfficiënt op in het eerste model. De coëfficiënt van 0,005 geeft aan dat als de gemiddelde afstand naar het treinstation met 1 kilometer toeneemt de gemiddelde woningwaarde met 0,5% toeneemt zoals weergegeven in tabel 5. De coëfficiënt is significant tot op één procent, waarbij het interval van afstand tot het treinstation $[0,003; 0,006]$ tot stand is gekomen met een methode die in 95% van de gevallen een interval op zal leveren waar het onbekende populatiegemiddelde in ligt zoals weergegeven in tabel 7 in de appendix. Dit interval laat daarnaast zien dat de coëfficiënt in het eerste model niet gelijk is aan 0, dit wordt ook weergegeven door de F-test. De null-hypothese dat alle coëfficiënten gelijk zijn aan 0 is hiermee verworpen.

Het tweede regressiemodel is meervoudig nu afstand tot het dichtstbijzijnde overstappunt is toegevoegd. De coëfficiënt van de afstand tot het treinstation is nog steeds significant tot op een 1% significantie niveau, maar geeft met een grote van 0,006 nu een 0,6% stijging van de gemiddelde huizenprijzen als de afstand met 1 km toeneemt. Deze waarde laat zien dat er in het eerste model een *downward bias* van de coëfficiënt B_1 was, ofwel de coëfficiënt van afstand tot het treinstation. Dit is in overeenstemming met de verwachting in de methode. De F-test in tabel 8 in de appendix laat zien dat minstens één van de coëfficiënten niet gelijk is aan 0.

In het derde meervoudig regressiemodel wordt ook de afstand tot de hoofdverkeersweg toegevoegd. Deze controle variabele verandert de coëfficiënt B_1 , maar het verschil met model 2 is dermate klein dat de op 3-decimalen afgeronde coëfficiënt 0,006 blijft. Deze is significant tot op een 1% significantie niveau. De verwachting in de methode dat er sprake is van een *upward bias*, zonder het toevoegen van afstand tot hoofdverkeersweg blijkt te kloppen. De coëfficiënt B_3 is positief zoals verwacht en zorgt ervoor dat de coëfficiënt B_1 in het derde model lager is dan in model 2, wat een *upward bias* impliceert. De hypothese dat één van de coëfficiënten gelijk is aan 0 wordt verworpen door de F-test in tabel 9 (appendix).

Het vierde model heeft met de bevolkingsdichtheid nog steeds een tot op 1% significante coëfficiënt voor afstand tot het treinstation. Echter neemt de gemiddelde woningwaarde nu met 0,4% toe wanneer het station 1 kilometer verder ligt, aangezien de parameter een waarde van 0,004 heeft. De hogere waarde van B_1 in model 3 ligt in de lijn met de verwachte met de *upward bias* in de sectie methode. Ondanks dat ook in dit model de F-test in tabel 10 een p-waarde van 0,000 aangeeft, blijft het twijfelachtig of we bevolkingsdichtheid definitief moeten toevoegen. Uit voorgaande literatuur van Fajarang, (1994) blijkt dat huishoudens zich vestigen rondom het treinstation. Hierdoor is bevolkingsdichtheid een mechanisme in plaats van een controle variabele.

Gemiddeld inkomen wordt in het vijfde model als controle variabele toegevoegd. De variabele waarin ik geïnteresseerd ben, heeft nu een coëfficiënt van 0,002 op een één procent significantieniveau, ofwel een wijk waar de gemiddelde afstand tot het station 1 kilometer meer is, heeft huizen die 0,2% duurder zijn. De F-test in tabel 11 laat zien dat tenminste één van de coëfficiënten niet gelijk is aan 0 met een F-waarde van 497. De positieve coëfficiënt van inkomen suggereert dat mensen met een hoger inkomen in duurdere huizen wonen. Doordat uit de correlatie tabel in de sectie veronderstellingen blijkt dat er geen correlatie bestaat tussen inkomen en afstand tot het treinstation lijkt het erop dat inkomen wellicht een beeld geeft van niet waargenomen variabelen.

Desalniettemin is het vanuit economisch perspectief niet logisch dat inkomen de woningprijzen in een buurt beïnvloedt. Verschil in prijzen wordt bepaald door de specifieke karakteristieken van een woning. De waarde van een woning neemt niet toe als de buurman een nieuwe functie krijgt met een beter inkomen. Echter is het wel mogelijk dat als in een buurt de prijzen hoog zijn er mensen met een hoog salaris komen wonen, aangezien de maximale hoogte van een hypotheek verband houdt met inkomen. Daarnaast is de afstand tot het treinstation voor buurten meestal gelijk over tijd en fluctueert het gemiddelde inkomen. Dit impliceert dat het gemiddelde inkomen bepaald wordt nadat de afstand tot het treinstation tot stand gekomen is. Dit zorgt ervoor dat afstand tot het treinstation een voorspeller is van het inkomen. Wanneer het bovenstaande in acht genomen wordt, lijkt inkomen een *collider* te zijn in plaats van een controle variabele.

In model 6 en 7 zijn de gemeente vaste effecten toegevoegd. Model 6 bevat enkel de vaste effecten als controle variabelen. Model 7 daarentegen bevat ook de afstand naar het dichtstbijzijnde overstappunt en de hoofdverkeersweg. Doordat bevolkingsdichtheid een mechanisme lijkt te zijn en inkomen een mogelijke collider is, is er sprake van onjuiste controle variabelen. Een mechanisme en collider zorgen respectievelijk voor bias en een onechte relatie. Daarom heb ik ervoor gekozen om de bevolkingsdichtheid en het inkomen uit de meervoudige regressie te laten. Echter blijkt uit de voorgaande literatuur dat bevolkingsdichtheid en inkomen waarde bepalend zijn voor het openbaar vervoer, daarom worden later in de resultaten regressies conditioneel op provincie en inkomen besproken.

Tabel 6: Lineaire regressieschattingen van de determinanten van woningwaarde met vaste effecten in 2017.

	(1) logaritme van woningprijs	(6) logaritme van woningprijs	(7) logaritme van woningprijs
afstand tot treinstation	0.00452*** (0.000558)	0.0147*** (0.00130)	0.0129*** (0.00140)
afstand tot overstappunt			0.00143 (0.00111)
afstand tot hoofdweg			0.0247*** (0.00328)
constante	5.408*** (0.00556)	5.333*** (0.0559)	5.285*** (0.0555)
Vaste Effecten	NEE	JA	JA
N	9860	9860	9860

Standaardfouten tussen haakjes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In model 6 zijn 387 dummy's om te controleren voor gemeente vaste effecten toegevoegd. De coëfficiënt voor afstand tot het treinstation is gelijk aan 0.0147 en significant tot op een 1 procent significantieniveau. Dit wil zeggen dat als de afstand naar het station met één kilometer toeneemt, de gemiddelde woningprijs met 1,5% stijgt. Het betrouwbaarheidsinterval $[0,012; 0,017]$ van coëfficiënt B_1 is tot stand gekomen met een methode die in 95% van de gevallen een interval op zal leveren waar het onbekende populatiegemiddelde in ligt zoals weergegeven in tabel 12 in de appendix.

Model 7 heeft door het toevoegen van twee andere controle variabelen een B_1 met een waarde van 0,013, ofwel als de gemiddelde afstand tot het station in een buurt met 1 kilometer toeneemt, stijgt de gemiddelde woningwaarde met 1,3%. De coëfficiënt heeft een significantieniveau van 1% zoals ook blijkt uit tabel 6. Het betrouwbaarheidsinterval [0,010; 0,016] is tot stand is gekomen met een methode die in 95% van de gevallen een interval op zal leveren waar het onbekende populatiegemiddelde in ligt zoals weergegeven in tabel 13 in de appendix. De F-test voor deze regressie is niet beschikbaar, maar het betrouwbaarheidsinterval sluit tot op een 95% significantieniveau uit dat de coëfficiënt B_1 gelijk is aan 0.

Aangezien bevolkingsdichtheid niet in de regressie van model 7 is opgenomen, maak ik onderscheid tussen provincies. De regressies conditioneel op de 12 Nederlandse provincies laten de verschillen tussen provincies zien. Tabel 14 en 15 in de appendix laten zien dat voor de provincies Friesland, Flevoland, Utrecht en Zuid-Holland er geen significante coëfficiënt voor het effect van de afstand tot het treinstation op woningprijzen is. Voor Friesland en Flevoland kan een klein aantal observaties de oorzaak zijn, terwijl dit bij Utrecht en Zuid-Holland niet van toepassing is. Een mogelijke verklaring is dat er veel spoorlijnen per vierkante kilometer liggen in Utrecht en Zuid-Holland, waardoor er weinig onderscheid is tussen woningen en de afstand tot het treinstation geen effect meer heeft op woningprijzen.

Voor alle overige provincies geldt dat de coëfficiënt B_1 positief is, dit impliceert dat als de gemiddelde afstand tot het treinstation groter is, de woningprijzen hoger zijn. De grote van de B_1 verschilt echter wel tussen de verschillende provincies. Overijssel en Gelderland hebben respectievelijk een coëfficiënt van 0,027 en 0,023 welke een 2,7% en 2,3% stijging laten zien wanneer de afstand tot het treinstation met één kilometer toeneemt. Noord-Holland en Limburg hebben kleinere coëfficiënten. In Noord-Holland stijgt de prijs van een huis met 1,6% als de afstand tot het treinstation met 1 kilometer toeneemt. In Limburg is de stijging iets groter namelijk 1,9% door een coëfficiënt van 0,019. In Drenthe is de coëfficiënt B_1 gelijk aan 0,010, ofwel huizenprijzen stijgen met 1 procent als de gemiddelde afstand tot het treinstation met 1 kilometer toeneemt. In Zeeland is de stijging gelijk aan 1,4%, echter geldt voor zowel Drenthe als Zeeland dat de coëfficiënt significant is tot op een 5% niveau. Het effect van de afstand tot het treinstation op de woningprijzen is nagenoeg gelijk voor Groningen en Noord-Brabant. Afgerond geeft de coëfficiënt van 0,009 een 0,9% stijging weer. De bovenstaande

resultaten laten zien dat ondanks dat de coëfficiënt B_1 voor alle provincies positief is, de waarde van de stijging kleiner is voor bepaalde provincies. Hierdoor stijgt de waarde van een huis minder in Groningen en Drenthe vergeleken met Noord-Holland en Gelderland wanneer de afstand tot het station toeneemt. Dit zou kunnen komen doordat in de noordelijke provincies dicht bij het station wonen positiever ervaren wordt.

De keuze om gemiddeld inkomen per buurt niet in de regressie van model 7 op te nemen zorgt ervoor dat ik regressies conditioneel op inkomen kan uitvoeren. Inkomen is hiervoor opgedeeld in 4 kwartielen. Tabel 16 in de appendix laat echter zien dat niet iedere regressie hetzelfde aantal observaties heeft. Dit komt omdat er wijken zijn met een bekend inkomen, maar een onbekende woningwaarde of afstand tot het station, waardoor de regressie niet kan worden uitgevoerd. Voor alle kwartielen geldt dat de coëfficiënt B_1 significant is tot op een 1 procent significantieniveau. In het eerste kwartiel, ofwel 25% van de buurten met het laagste gemiddelde inkomen, leidt een toename van 1 kilometer tot het station tot een woningprijs toename van 1%. Hoewel de coëfficiënt voor het tweede kwartiel iets verschilt, gaat het hier ook om een waarde van 0,010. Hiermee lijkt er tussen het eerste en tweede kwartiel geen verschil te zitten in de waarde die de afstand tot het station heeft voor verschillende inkomensgroepen. Tabel 16 in de appendix laat daarentegen wel zien dat in het derde kwartiel huizenprijzen meer stijgen als de afstand naar het treinstation toeneemt. Een toename van één kilometer houdt verband met een 1,3% stijging van de huizenprijs door een coëfficiënt van 0,013. Het vierde en laatste kwartiel heeft ten opzichte van het derde kwartiel een kleinere parameter van 0,012. Daarmee neemt de huizenprijs dus 1,2% toe als de afstand tot het station met 1 kilometer wordt vermeerderd. Deze regressies laten hiermee weliswaar de trend zien dat wanneer een buurt een lager inkomen heeft het openbaar vervoer belangrijker is, maar de coëfficiënt B_1 is voor alle inkomensgroepen positief. Dit betekent dat een huis in een buurt in het eerste en tweede kwartiel ten opzichte van het derde en vierde kwartiel, minder in waarde stijgt wanneer de afstand tot het treinstation toeneemt.

7. Conclusie en Discussie

In deze paper is onderzocht of er sprake is van hogere gemiddelde woningprijzen in buurten waar het treinstation gemiddeld dichterbij ligt. In eerste instantie is gekeken of de afstand tot het treinstation resulteert in significante verschillen in woningprijzen. Door gebruik te maken van hedonische vraagtheorie en Nederlandse data van buurten is met behulp van een *Ordinary Least Squares* regressie gekeken wat het effect is van de afstand tot een treinstation op woningprijzen. Aan de hand hiervan is de volgende onderzoeksvraag onderzocht:

Wat is de invloed van een NS-treinstation op de prijzen van koopwoningen in 2017 in Nederland?

Voor het beantwoorden van de onderzoeksvraag is één hypothese opgesteld. De hypothese, die stelt dat een kleinere afstand naar het NS-treinstation leidt tot hogere gemiddelde woningprijzen in een buurt, is verworpen. De coëfficiënt van afstand tot het treinstation is namelijk positief. Wat betekent dat als de gemiddelde afstand naar een treinstation toeneemt, de huizenprijs ook toeneemt. Een mogelijke verklaring hiervoor is dat het station, maar vooral het bijbehorende spoor negatieve effecten met zich meebrengt (Diao, Qin & Sing, 2015).

Verder blijkt uit regressies conditioneel op de kwartielen van het gemiddelde inkomen in een buurt dat de coëfficiënt van afstand tot het treinstation verschilt. Buurten met een lager gemiddeld inkomen hebben een kleinere coëfficiënt vergeleken met buurten met een hoger gemiddeld inkomen. Een mogelijke verklaring hiervoor is dat in buurten met een lager gemiddeld inkomen, de waarde die bewoners aan het station koppelen groter is. Hierdoor is de coëfficiënt die de relatie tussen de afstand tot het station en woningprijzen aangeeft verschillend (Bowes & Ihlanfeldt, 2001).

Concluderend kan als antwoord op de onderzoeksvraag gesteld worden dat een gemiddeld grotere afstand naar het treinstation een significant positief effect heeft op woningprijzen, waarbij er sprake is van een verschil tussen verschillende inkomens per buurt. Hiermee is het onderzoek in tegenspraak met het onderzoek van Debrezion, Pels & Rietveld (2010), maar in overeenstemming met Bowes & Ihlanfeldt (2001).

Ondanks dat gewerkt is met controle variabelen kunnen er problemen zijn met de interne validiteit. Wanneer er bij een *Ordinary Least Squares* regressie sprake is van *Omitted Variable Bias* kan de coëfficiënt een foutieve relatie weergeven tussen gemiddelde afstand tot het treinstation en de gemiddelde woningprijs. Een mogelijke oplossing hiervoor is om meer controle variabelen toe te voegen of een andere onderzoeksmethode te gebruiken, zoals een met een Instrumentele Variabele. Een ander mogelijk probleem is dat voor sommige buurten de dataset missende waarden bevat. Vooral voor woningprijzen is dit het geval. Een mogelijke reden is dat het dan bijvoorbeeld gaat om een bedrijventerrein. Het kan waardevol zijn om de waarde van kantoren en bedrijfshallen ook mee te nemen in de analyse.

De externe validiteit van dit onderzoek is hoog voor Nederland. Dit komt omdat het onderzoek is uitgevoerd met data van Nederlandse buurten. Wanneer de uitkomsten van dit onderzoek in een ander land worden gebruikt is de externe validiteit laag. De fiscale behandeling van koopwoningen verschilt tussen landen en dit heeft mogelijk ook gevolgen voor de prijzen. Het spreekt in het voordeel van het onderzoek dat de data afkomstig is uit 2017, waardoor deze representatief is.

Ruimte voor vervolgonderzoek is reeds te vinden. Dit onderzoek vindt in tegenstelling tot voorgaand onderzoek een positieve coëfficiënt voor afstand. Het onderzoek verschilt weliswaar ten opzichte van eerder onderzoek, doordat er gebruik wordt gemaakt van gemiddelde afstanden en woningprijzen, maar toch is lang gedacht dat het tegenovergestelde waar zou zijn. Er is wellicht een verband met de bouwjaren van huizen. Nieuwbouwwijken staan vaak aan de buitenrand van de stad, ver van het station, maar zijn de laatste jaren duur. Verder is het interessant om dit onderzoek voor grondwaarde uit te voeren en de verschillende coëfficiënten tussen provincies uit te leggen.

Aangezien dat de nabijheid van een station een negatief effect heeft op woningprijzen lijkt het lastig om publieke investeringen lokaal toe te schrijven. Echter kunnen deze nieuwe inzichten wel bijdragen aan de overwegingen voor overheidsinvesteringen in de toekomst. Dat een treinstation een negatief effect heeft op woningprijzen biedt de kans om te kijken of andere voorzieningen zwaarder wegen, zoals P+R bij of een openbaar vervoer verbinding naar het station toe. Het is wellicht mogelijk om met hetzelfde budget het openbaar vervoer efficiënter in te richten om zo het fysiek kapitaal in Nederland verder te laten toenemen.

8. Referenties

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press.

Bowes, D.R. & Ihlanfeldt, K.R. (2001). Identifying the Impacts of Rail Transit Stations on Residential Property Values. *Journal of Urban Economics*, 50(1), 1-25. Doi: <https://doi.org/10.1006/juec.2001.2214>Get rights and content

Buehler, R. (2011). Determinants of transport mode choice: a comparison of Germany and the USA. *Journal of Transport Geography*, 19(4), 644-657. Doi: <https://doi.org/10.1016/j.jtrangeo.2010.07.005>

Centraal Bureau voor de Statistiek. (2015). *Minder huizen onder water*. Geraadpleegd via <https://www.cbs.nl/nl-nl/nieuws/2015/44/minder-huizen-onder-water>

Centraal Bureau voor de Statistiek. (2017). *Lengte van spoorwegen; spoorwegkenmerken, provincie*. Geraadpleegd via <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/71024ned/table?ts=1591886533811>

Centraal Bureau voor de Statistiek. (2019). *Nabijheid voorzieningen; afstand locatie, wijk- en buurtcijfers 2017*. Geraadpleegd via <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84334NED/table?ts=1588609588199>

Centraal Bureau voor de Statistiek. (2020). *Kerncijfers wijken en buurten 2017*. Geraadpleegd via <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83765NED/table?ts=1588610598789>

Currie, G. (2010). Quantifying spatial gaps in public transport supply based on social needs. *Journal of Transport Geography*, 18(1), 31-41. Doi: <https://doi.org/10.1016/j.jtrangeo.2008.12.002>

Damm, D., Lerman, S. R., Lerner-Lam, E. & Young, J. (1980). Response of urban real estate values in anticipation of the Washington Metro. *Journal of Transport Economics and Policy*, 14(3), 315–336. Geraadpleegd via https://www.jstor-org.eur.idm.oclc.org/stable/20052588?seq=1#metadata_info_tab_contents

Debrezion, G., Pels, E. & Rietveld, P. (2010). The Impact of Rail Transport on Real Estate Prices: An Empirical Analysis of the Dutch Housing Market. *Urban Studies*, 48(5), 997-1015. Doi: <https://doi-org.eur.idm.oclc.org/10.1177/0042098010371395>

Deweese, D. N. (1976). The effect of a subway on residential property values in Toronto. *Journal of Urban Economics*, 3(4), 357–369. Doi: [https://doi.org/10.1016/0094-1190\(76\)90035-8](https://doi.org/10.1016/0094-1190(76)90035-8)

Diao, M., Qin, Y. & Sing, T.F. (2015). Negative Externalities of Rail Noise and Housing Values: Evidence from the Cessation of Railway Operations in Singapore. *Real Estate Economics*, 44(4), 878-917. Doi: <https://doi-org.eur.idm.oclc.org/10.1111/1540-6229.12123>

Fejarang, R. A. (1994) *Impact on property values: a study of the Los Angeles Metro Rail*. London: PTRC Education and Research Services on behalf of the Planning and Transport Research and Computation International Association.

Grass, R.G. (1992). The Estimation of Residential Property Values Around Transit Station Sites in Washington, D.C. *Journal of Economics and Finance*, 16(2), 139-147. Doi: 10.1007/BF02920114

Kane, T. J., Riegg, S. K. & Staiger, D. O. (2006). School Quality, Neighbourhoods, and Housing Prices. *American Law and Economic Review*, 8(2), 183-212. Doi: <https://doi-org.eur.idm.oclc.org/10.1093/aler/ahl007>

Lipscomb, C. (2003). Small Cities Matter, Too: The Impacts of an Airport and Local Infrastructure on Housing Prices in a Small Urban City. *Review of urban and regional development studies*, 15(3), 255-273. Doi: <https://doi-org.eur.idm.oclc.org/10.1111/j.1467940X.2003.00076.x>

Mathur, S. (2008). Impact of Transportation and Other Jurisdictional-Level Infrastructure and Services on Housing Prices. *Journal of Urban Planning and Development*, 134(1). Doi: [https://doi-org.eur.idm.oclc.org/10.1061/\(ASCE\)0733-9488\(2008\)134:1\(32\)](https://doi-org.eur.idm.oclc.org/10.1061/(ASCE)0733-9488(2008)134:1(32))

Muth, R. F. (1969). *Cities and Housing*. Chicago, IL: University of Chicago Press.

Omroep West. (2019). Station Lansingerland-Zoetermeer officieel open: 'een laan boven de A12'. Geraadpleegd via: <https://www.omroepwest.nl/nieuws/3834036/Station-Lansingerland-Zoetermeer-officieel-open-een-laan-boven-de-A12>

Palm, M., Gregor, B., Wang, H. & McMullen, S. (2014). The trade-offs between population density and households' transportation-housing costs. *Transport Policy*, 36, 160-172. Doi: <https://doi.org/10.1016/j.tranpol.2014.07.004>

Poterba, J.M. (1984). Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach. *The Quarterly Journal of Economics*, 99(4), 729-752. Doi: <https://doi-org.eur.idm.oclc.org/10.2307/1883123>

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34-56. Geraadpleegd via <https://www-journals-uchicago-edu.eur.idm.oclc.org/doi/abs/10.1086/260169?journalCode=jpe>

Van bockxmeer, J., (2020). *Huizenprijzen weer sterker gestegen*. Geraadpleegd via <https://fd.nl/economie-politiek/1331570/huizenprijzen-7-gestegen-in-2019>

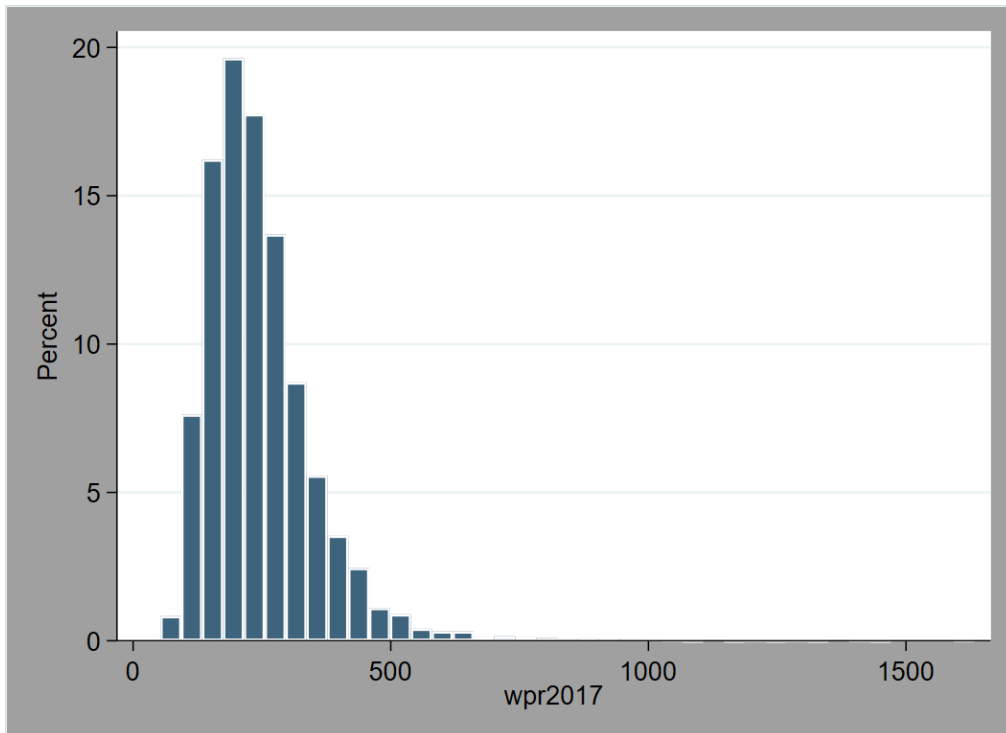
Van Ewijk, C., Jacobs, B. & De Mooij, R. (2007). Welfare effects of fiscal subsidies on home ownership in the Netherlands. *The Economist*, 155, 323-336. Doi: 10.1007/s10645-007-9064-z

Voith, R. (1993). Changing Capitalization of CBD-Oriented Transportation Systems: Evidence from Philadelphia, 1970–1988. *Journal of urban economics*, 33(3), 361-376. Doi: <https://doi.org/10.1006/juec.1993.1021>

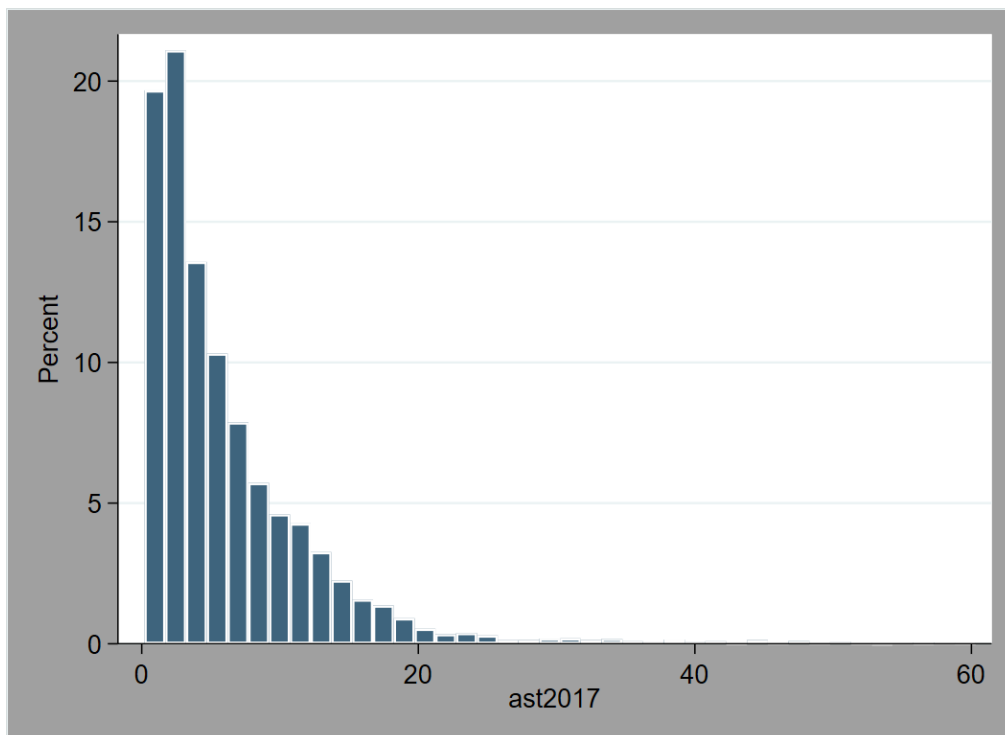
Von Thünen, J.H. (1830). *Der Isolierte staat in Beziehung auf Landwirtschaft und nationalökonomie*. München: Pflaum.

9. Appendix

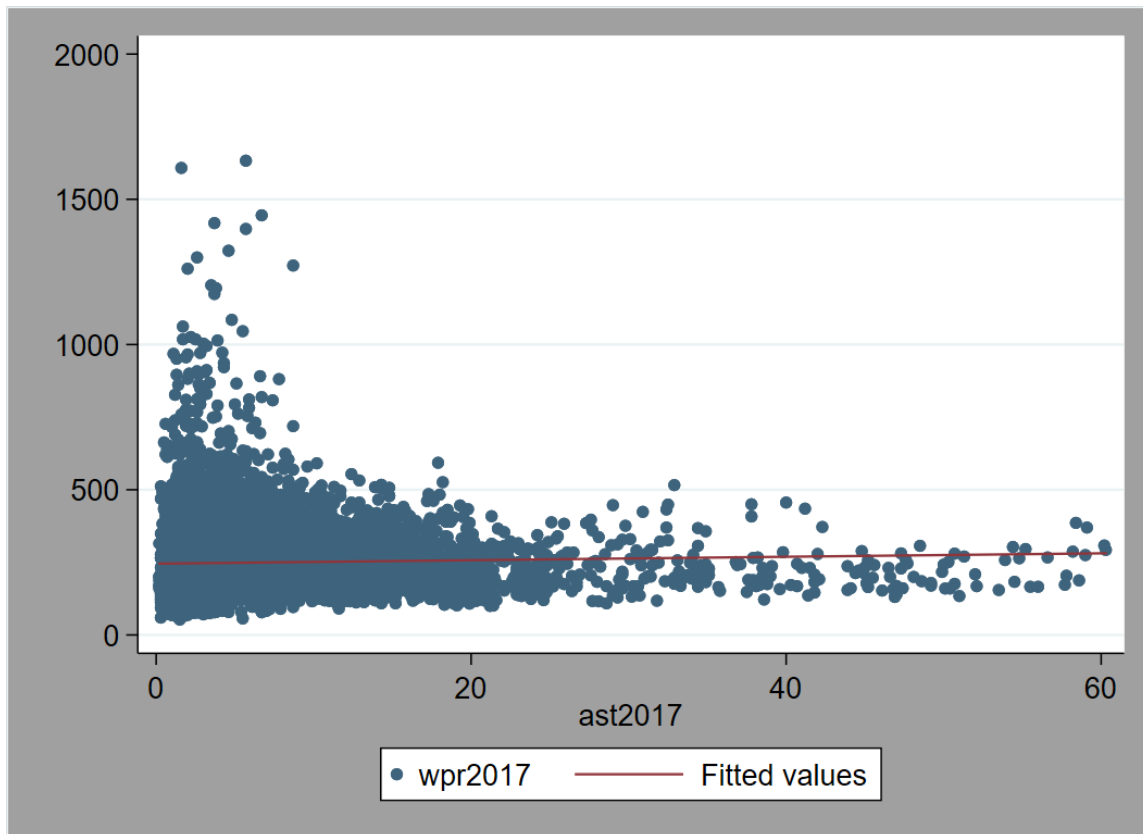
9.1 Figuren



Figuur 1: Histogram van woningprijzen.



Figuur 2: Histogram van afstand tot het treinstation.



Figuur 3: Scatterplot van woningprijzen uitgezet tegen afstand tot het treinstation.

9.2 Tabellen

Tabel 1: Beschrijvende statistieken van variabelen in de dataset.

Variable	Observaties	Gemiddelde	Std. Dev.	Min	Max
woningprijs	9,866	249.18	114.70	53	1633
afstand tot treinstation	12,583	6.34	6.68	.2	60.3
afstand tot overstappunt	12,583	12.90	9.85	.2	71.9
afstand tot hoofdweg	12,583	1.88	1.80	.1	46.4
bevolkingsdichtheid	12,610	3037.29	3966.28	0	35921
inkomen	10,857	26.92	7.37	9.3	142

Tabel 2: Naam en eenheid van variabelen in de dataset.

Variable	Naam	Eenheid	Bijzonderheden
wpr2017	gemiddelde woningprijs in 2017	€	x1000
ast2017	afstand tot station in 2017	km	tot op 100m
ado2017	afstand tot dichtbijzijnde overstappunt in 2017	km	tot op 100m
asw2017	afstand tot hoofdverkeersweg in 2017	km	tot op 100m
bvd2017	bevolkingsdichtheid in 2017	per km ²	
ink2017	gemiddeld inkomen per inwoner	€	x1000

Tabel 3: Correlatiecoëfficiënten voor ZCM-aanname.

	afstand tot treinstation	afstand tot overstap-punt	afstand tot hoofdweg	bevolkings-dichtheid	inkomen
afstand tot treinstation	1.0000				
afstand tot overstappunt	0.6410	1.0000			
afstand tot hoofdweg	0.1674	0.1131	1.0000		
bevolkings-dichtheid	-0.2930	-0.3823	0.0083	1.0000	
inkomen	-0.0359	-0.1196	0.0155	-0.1016	1.0000

Tabel 4: Granger Causality test voor woningprijs en afstand tot treinstation.

	(1) woningprijs in 2017	(2) afstand tot treinstation in 2017
woningprijs in 2016	1.036*** (0.00504)	0.0000318 (0.0000406)
afstand tot trein- station in 2017	-0.225*** (0.0204)	1.000*** (0.000210)
constante	-0.152 (1.040)	-0.00466 (0.00923)
N	9312	9555

Standaardfouten tussen haakjes

* p<0.10, ** p<0.05, *** p<0.01

Tabel 5: Lineaire regressieschattingen van de determinanten van woningwaarde in 2017.

	(1) logaritme van woningprijs	(2) logaritme van woningprijs	(3) logaritme van woningprijs	(4) logaritme van woningprijs	(5) logaritme van woningprijs
afstand tot treinstation	0.00452*** (0.000558)	0.00604*** (0.000705)	0.00578*** (0.000707)	0.00434*** (0.000630)	0.00223*** (0.000516)
afstand tot overstappunt		-0.00158*** (0.000520)	-0.00159*** (0.000518)	-0.00594*** (0.000519)	0.000480 (0.000387)
afstand tot hoofdweg			0.00767*** (0.00220)	0.0131*** (0.00207)	0.00676*** (0.00157)
bevolkings- dichtheid				-0.0000299*** (0.00000155)	-0.0000185*** (0.00000100)
inkomen					0.0391*** (0.00103)
constante	5.408*** (0.00556)	5.418*** (0.00709)	5.406*** (0.00758)	5.571*** (0.0107)	4.426*** (0.0296)
Vaste effecten	NEE	NEE	NEE	NEE	NEE
N	9860	9860	9860	9860	9820

Standaardfouten tussen haakjes

* p<0.10, ** p<0.05, *** p<0.01

Tabel 6: Lineaire regressieschattingen van de determinanten van woningwaarde met vaste effecten in 2017.

	(1) logaritme van woningprijs	(6) logaritme van woningprijs	(7) logaritme van woningprijs
afstand tot treinstation	0.00452*** (0.000558)	0.0147*** (0.00130)	0.0129*** (0.00140)
afstand tot overstappunt			0.00143 (0.00111)
afstand tot hoofdweg			0.0247*** (0.00328)
constante	5.408*** (0.00556)	5.333*** (0.0559)	5.285*** (0.0555)
Vaste Effecten	NEE	JA	JA
N	9860	9860	9860

Standaardfouten tussen haakjes

* p<0.10, ** p<0.05, *** p<0.01

Tabel 7: Enkelvoudig lineair regressiemodel 1.

Aantal Observaties	=	9860
F	=	65,63
Kans > F	=	0,0000
R-kwadraat	=	0,0058
Aangep. R-kwadraat	=	0,0057
Wortelgem. Kwadr. Fout	=	0,3955

logaritme van woningprijs	Coef.	Robuuste Std. Fout.	t	P> t	95% Onder grens	95% Boven grens
afstand tot treinstation	0,0045	0,0005582	8,10	0,000	0,0034279	0,0056163
constante	5,4078	0,0055647	971,79	0,000	5,3968475	5,4186635

Tabel 8: Meervoudig lineair regressiemodel 2 met controle variabele.

Aantal Observaties	=	9860
F	=	40,67
Kans > F	=	0,0000
R-kwadraat	=	0,0067
Aangep. R-kwadraat	=	0,0065
Wortelgem. Kwadr. Fout	=	0,3954

logaritme van woningprijs	Coef.	Robuuste Std. Fout.	t	P> t	95% Onder grens	95% Boven grens
afstand tot treinstation	0,0060	0,0007052	8,56	0,000	0,0046573	0,0074222
afstand tot overstappunt	-0,0016	0,0005196	-3,05	0,002	-0,0026028	-0,0005658
constante	5,4181	0,0070921	763,96	0,000	5,4042341	5,4320381

Tabel 9: Meervoudig lineair regressiemodel 3 met controle variabelen.

Aantal Observaties	=	9860
F	=	32,76
Kans > F	=	0,0000
R-kwadraat	=	0,0076
Aangep. R-kwadraat	=	0,0073
Wortelgem. Kwadr. Fout	=	0,3952

logaritme van woningprijs	Coef.	Robuuste Std. Fout.	t	P> t	95% Onder grens	95% Boven grens
afstand tot treinstation	0,0058	0,0007072	8,17	0,000	0,0043929	0,0071654
afstand tot overstappunt	-0,0016	0,0005185	-3,06	0,002	-0,0026047	-0,0005721
afstand tot hoofdweg	0,0077	0,0022030	3,48	0,001	0,00334716	0,0119839
constante	5,4057	0,0075770	713,44	0,000	5,39084834	5,4205532

Tabel 10: Meervoudig lineair regressiemodel 4 met controle variabelen.

Aantal Observaties	=	9860
F	=	128,53
Kans > F	=	0,0000
R-kwadraat	=	0,0905
Aangep. R-kwadraat	=	0,0901
Wortelgem. Kwadr. Fout	=	0,3784

logaritme van woningprijs	Coef.	Robuuste Std. Fout.	t	P> t	95% Onder grens	95% Boven grens
afstand tot treinstation	0,0043	0,0006301	6,89	0,000	0,0031070	0,0055771
afstand tot overstappunt	-0,0059	0,0005192	-11,45	0,000	-0,0069626	-0,0049273
afstand tot hoofdweg	0,0131	0,0020692	6,33	0,000	0,0090330	0,0171453
Bevolkingsdichtheid	0,0000	0,0000015	-19,34	0,000	-0,0000330	-0,0000269
constante	5,5710	0,0106663	522,30	0,000	5,5501048	5,5919210

Tabel 11: Meervoudig lineair regressiemodel 5 met controle variabelen.

Aantal Observaties	=	9820
F	=	497,08
Kans > F	=	0,0000
R-kwadraat	=	0,6057
Aangep. R-kwadraat	=	0,6055
Wortelgem. Kwadr. Fout	=	0,2488

logaritme van woningprijs	Coef.	Robuuste Std. Fout.	t	P> t	95% Onder grens	95% Boven grens
afstand tot treinstation	0,0022	0,0005164	4,32	0,000	0,0012202	0,0032448
afstand tot overstappunt	0,0005	0,0003867	1,24	0,215	-0,0002782	0,0012379
afstand tot hoofdweg	0,0068	0,0015672	4,32	0,000	0,0036922	0,0098362
Bevolkingsdichtheid	0,0000	0,0000010	-18,46	0,000	-0,0000205	-0,0000166
inkomen	0,0391	0,0010296	37,98	0,000	0,0370861	0,0411224
constante	4,4259	0,0295773	149,64	0,000	4,3678728	4,4838282

Tabel 12: Meervoudig lineair regressiemodel 6 met vaste effecten.

Aantal Observaties	=	9860
F	=	.
Kans > F	=	.
R-kwadraat	=	0,3701
Aangep. R-kwadraat	=	0,3448
Wortelgem. Kwadr. Fout	=	0,3211
Vaste Effecten	=	JA

logaritme van woningprijs	Coef.	Robuuste Std. Fout.	t	P> t	95% Onder grens	95% Boven grens
afstand tot station	0,0147	0,0013004	11,32	0,000	0,0121740	0,0172721
constante	5,3334	0,0558981	95,41	0,000	5,2238377	5,4429821

Tabel 13: Meervoudig lineair regressiemodel 7 met controle variabelen & vaste effecten.

Aantal Observaties	=	9860
F	=	.
Kans > F	=	.
R-kwadraat	=	0,3738
Aangep. R-kwadraat	=	0,3486
Wortelgem. Kwadr. Fout	=	0,3202
Vaste Effecten	=	JA

logaritme van woningprijs	Coef.	Robuuste Std. Fout.	t	P> t	95% Onder grens	95% Boven grens
afstand tot treinstation	0,0129	0,0014016	9,24	0,000	0,0101971	0,0156920
afstand tot overstappunt	0,0014	0,0011144	1,29	0,198	-0,0007507	0,0036182
afstand tot hoofdweg	0,0247	0,0032773	7,53	0,000	0,0182528	0,0311011
constante	5,2852	0,0554882	95,25	0,000	5,1764555	5,3939931

Tabel 14: Meervoudig lineair regressiemodel conditioneel op provincies (1/2), controle variabelen & vaste effecten.

	(1)	(2)	(3)	(4)	(5)	(6)
	logaritme van woningprijs	logaritme van woningprijs	logaritme van woningprijs	logaritme van woningprijs	logaritme van woningprijs	logaritme van woningprijs
Provincie	Groningen	Friesland	Drenthe	Overijssel	Flevoland	Gelderland
afstand tot treinstation	0.00915** (0.00433)	-0.00183 (0.00357)	0.0101** (0.00464)	0.0272*** (0.00357)	0.0201 (0.0129)	0.0229*** (0.00400)
afstand tot overstappunt	-0.00254 (0.00290)	0.00165 (0.00279)	-0.00506 (0.00342)	0.00728** (0.00300)	0.0279* (0.0147)	0.00281 (0.00287)
afstand tot hoofdweg	0.0290** (0.0116)	0.0190** (0.00741)	0.0457*** (0.0117)	0.0179 (0.0111)	-0.0113 (0.0332)	0.0185* (0.00957)
constante	5.060*** (0.159)	5.180*** (0.0854)	5.378*** (0.0756)	4.906*** (0.0587)	5.106*** (0.0853)	5.333*** (0.0486)
Vaste Effecten	JA	JA	JA	JA	JA	JA
N	431	521	406	851	189	1266

Standaardfouten tussen haakjes

* p<0.10, ** p<0.05, *** p<0.01

Tabel 15: Meervoudig lineair regressiemodel conditioneel op provincies (2/2), controle variabelen & vaste effecten.

	(7)	(8)	(9)	(10)	(11)	(12)
	logaritme van woningprijs	logaritme van woningprijs	logaritme van woningprijs	logaritme van woningprijs	logaritme van woningprijs	logaritme van woningprijs
Provincie	Utrecht	Zuid-Holland	Noord- Holland	Zeeland	Noord- Brabant	Limburg
afstand tot treinstation	-0.00374 (0.0115)	0.00606 (0.00512)	0.0163*** (0.00494)	0.0135** (0.00579)	0.00887* (0.00460)	0.0189*** (0.00434)
afstand tot overstappunt	0.0298*** (0.00845)	-0.0238*** (0.00442)	0.00821** (0.00368)	-0.00166 (0.00536)	0.0163*** (0.00447)	0.00624 (0.00398)
afstand tot hoofdweg	-0.0114 (0.0161)	0.0498*** (0.0106)	0.0256*** (0.00958)	0.0187 (0.0137)	0.0146* (0.00826)	0.0211* (0.0114)
constante	5.366*** (0.0593)	5.899*** (0.0849)	5.054*** (0.0909)	5.282*** (0.0950)	5.056*** (0.105)	5.277*** (0.0757)
Vaste Effecten	JA	JA	JA	JA	JA	JA
N	590	1507	1709	305	1386	699

Standaardfouten tussen haakjes

* p<0.10, ** p<0.05, *** p<0.01

Tabel 16: Meervoudig lineair regressiemodel met kwartielen voor inkomen, controle variabelen & vaste effecten.

	(1)	(2)	(3)	(4)
	logaritme van woningprijs	logaritme van woningprijs	logaritme van woningprijs	logaritme van woningprijs
Kwartiel	1	2	3	4
afstand tot treinstation	0.00971*** (0.00192)	0.0102*** (0.00185)	0.0133*** (0.00238)	0.0117*** (0.00303)
afstand tot overstappunt	0.00221 (0.00154)	0.00667*** (0.00151)	0.00514*** (0.00178)	0.00391 (0.00242)
afstand tot hoofdweg	0.0154*** (0.00427)	0.0174*** (0.00417)	0.0120** (0.00478)	0.0138** (0.00640)
constante	4.959*** (0.0815)	5.072*** (0.104)	5.325*** (0.0633)	5.553*** (0.0780)
Vaste Effecten	JA	JA	JA	JA
N	2707	2272	2298	2543

Standaardfouten tussen haakjes

* p<0.10, ** p<0.05, *** p<0.01

9.3 STATA Codes

```
summarize wpr2017 ast2017 ado2017 asw2017 bvd2017 ink2017 if
type=="Buurt      "
histogram wpr2017 if type=="Buurt      ", percent fcolor(edkblue)
lcolor(white) lpattern(solid) ylabel(, angle(horizontal))
graphregion(fcolor(gs10))
histogram ast2017 if type=="Buurt      ", percent fcolor(edkblue)
lcolor(white) lpattern(solid) ylabel(, angle(horizontal))
graphregion(fcolor(gs10))
twoway (scatter wpr2017 ast2017 if type=="Buurt      ", sort
mcolor(edkblue) msize(small)) (lfit wpr2017 ast2017 if type=="Buurt
"), ylabel(, angle(horizontal)) graphregion(fcolor(gs10))

correlate ast2017 ado2017 asw2017 bvd2017 ink2017 if type=="Buurt      "
eststo: regress wpr2017 wpr2016 ast2016 if type=="Buurt      ", robust
eststo: regress ast2017 ast2016 wpr2016 if type=="Buurt      ",
robustesttab using Output.csv, se star(* 0.10 ** 0.05 *** 0.01)
eststo clear

eststo: regress wpr2017_ln ast2017 if type=="Buurt      ", robust
eststo: regress wpr2017_ln ast2017 ado2017 if type=="Buurt      ",
robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 if type=="Buurt
", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 bvd2017 if
type=="Buurt      ", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 bvd2017 ink2017 if
type=="Buurt      ", robust
esttab using Output2.csv, se star(* 0.10 ** 0.05 *** 0.01)
eststo clear
eststo: regress wpr2017_ln ast2017 if type=="Buurt      ", robust
eststo: regress wpr2017_ln ast2017 i.gemeente_id if type=="Buurt      ",
robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt      ", robust
esttab using Output3.csv, se star(* 0.10 ** 0.05 *** 0.01)
eststo clear

regress wpr2017_ln ast2017 if type=="Buurt      ", robust
putexcel set "THS B5 Output Regression", sheet("Blad1")
putexcel E1=("Number of Observations") G1=(e(N))
putexcel E2=("F") G2=(e(F))
putexcel E3=("prob > F") G3=(Ftail(e(df_m), e(df_r),
e(F)))
putexcel E4=("R-squared") G4=(e(r2))
putexcel E5=("adj R-squared") G5=(e(r2_a))
putexcel E6=("ROOT MSE") G6=(e(rsme))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A8=matrix(a), names
regress wpr2017_ln ast2017 ado2017 if type=="Buurt      ", robust
putexcel set "THS B5 Output Regression 2", sheet("Blad1")
putexcel E1=("Number of Observations") G1=(e(N))
putexcel E2=("F") G2=(e(F))
putexcel E3=("prob > F") G3=(Ftail(e(df_m), e(df_r),
e(F)))
putexcel E4=("R-squared") G4=(e(r2))
```

```

putexcel E5=("adj R-squared") G5=(e(r2_a))
putexcel E6=("ROOT MSE") G6=(e(rmse))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A8=matrix(a), names
regress wpr2017_ln ast2017 ado2017 asw2017 if type=="Buurt ",
robust
putexcel set "THS B5 Output Regression 3", sheet("Blad1")
putexcel E1=("Number of Observations") G1=(e(N))
putexcel E2=("F") G2=(e(F))
putexcel E3=("prob > F") G3=(Ftail(e(df_m), e(df_r),
e(F)))
putexcel E4=("R-squared") G4=(e(r2))
putexcel E5=("adj R-squared") G5=(e(r2_a))
putexcel E6=("ROOT MSE") G6=(e(rmse))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A8=matrix(a), names
regress wpr2017_ln ast2017 ado2017 asw2017 bvd2017 if type=="Buurt
", robust
putexcel set "THS B5 Output Regression 4", sheet("Blad1")
putexcel E1=("Number of Observations") G1=(e(N))
putexcel E2=("F") G2=(e(F))
putexcel E3=("prob > F") G3=(Ftail(e(df_m), e(df_r),
e(F)))
putexcel E4=("R-squared") G4=(e(r2))
putexcel E5=("adj R-squared") G5=(e(r2_a))
putexcel E6=("ROOT MSE") G6=(e(rmse))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A8=matrix(a), names
regress wpr2017_ln ast2017 ado2017 asw2017 bvd2017 ink2017 if
type=="Buurt ", robust
putexcel set "THS B5 Output Regression 5", sheet("Blad1")
putexcel E1=("Number of Observations") G1=(e(N))
putexcel E2=("F") G2=(e(F))
putexcel E3=("prob > F") G3=(Ftail(e(df_m), e(df_r),
e(F)))
putexcel E4=("R-squared") G4=(e(r2))
putexcel E5=("adj R-squared") G5=(e(r2_a))
putexcel E6=("ROOT MSE") G6=(e(rmse))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A8=matrix(a), names
regress wpr2017_ln ast2017 i.gemeente_id if type=="Buurt ", robust
putexcel set "THS B5 Output Regression 6", sheet("Blad1")
putexcel E1=("Number of Observations") G1=(e(N))
putexcel E2=("F") G2=(e(F))
putexcel E3=("prob > F") G3=(Ftail(e(df_m), e(df_r),
e(F)))
putexcel E4=("R-squared") G4=(e(r2))
putexcel E5=("adj R-squared") G5=(e(r2_a))
putexcel E6=("ROOT MSE") G6=(e(rmse))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A8=matrix(a), names
regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt ", robust

```

```

putexcel set "THS B5 Output Regression 7", sheet("Blad1")
putexcel E1=("Number of Observations") G1=(e(N))
putexcel E2=("F") G2=(e(F))
putexcel E3=("prob > F") G3=(Ftail(e(df_m), e(df_r),
e(F)))
putexcel E4=("R-squared") G4=(e(r2))
putexcel E5=("adj R-squared") G5=(e(r2_a))
putexcel E6=("ROOT MSE") G6=(e(rmse))
matrix a = r(table)'
matrix a = a[.,1..6]
putexcel A8=matrix(a), names

xtile ink2017_4 = ink2017, nq(4)
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & ink2017_4==1, robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & ink2017_4==2, robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & ink2017_4==3, robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & ink2017_4==4, robust
esttab using Output4.csv, se star(* 0.10 ** 0.05 *** 0.01)
eststo clear

eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Groningen", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Friesland", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Drenthe", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Overijssel", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Flevoland", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Gelderland", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Utrecht", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Noord-Holland", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Zuid-Holland", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Zeeland", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Noord-Brabant", robust
eststo: regress wpr2017_ln ast2017 ado2017 asw2017 i.gemeente_id if
type=="Buurt " & provincie=="Limburg", robust
esttab using Output5.csv, se star(* 0.10 ** 0.05 *** 0.01)
eststo clear

```