

Modelmatige Waardebepaling Vastgoed

Taxatie woningwaarde met sociale indicatoren

Scriptie

Samir Marghich



Modelmatige Waardebepaling Vastgoed

Taxatie woningwaarde met sociale indicatoren

Samir Marghich

Scriptie



Vrije Universiteit

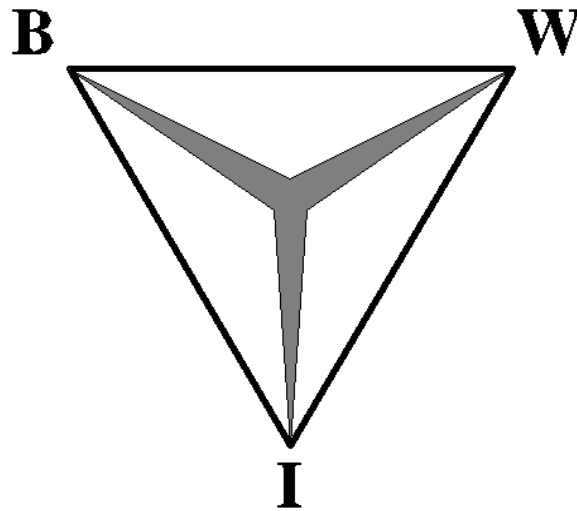
Faculteit der Exacte Wetenschappen

Divisie Wiskunde en Informatica

Studierichting *Bedrijfs*Wiskunde en *Informatica*

De Boelelaan 1081a

1081 HV Amsterdam



nbwo
Nederlands Bureau Waardebepaling Onroerende zaken

Klaprozenweg 75g
1033 NN Amsterdam

Juli 2005

Voorwoord

Als onderdeel van de opleiding Bedrijfswiskunde & Informatica aan de Vrije Universiteit te Amsterdam is dit stageverslag tot stand gekomen. Dit stageverslag samen met een presentatie in het bedrijf dient ter afronding van mijn afstudeerstage.

Mijn stage van vier maanden heeft plaats gevonden bij het bedrijf NBWO, Nederlands Bureau Waardebepaling Onroerende zaken, in Amsterdam. NBWO is een organisatie die zich bezighoudt met modelmatig waardebeoordeling van woningen in Nederland. Deze organisatie is een dochteronderneming van Davinci Advies BV, een adviesbedrijf met jarenlange ervaring in het adviseren aan o.a. financiële instellingen en overheidsinstanties. Ook is NBWO een dochteronderneming van Grontmij Taxaties vastgoed. Grontmij is als projectontwikkelaar al jaren actief op de vastgoedmarkt en heeft daardoor veel kennis van de woningmarkt.

Het doel van mijn stage was het ontwikkelen van statistisch onderbouwde analyses over de mate van invloed van sociale indicatoren op de verkoopwaarden van eengezinswoningen. Hierbij was het ook de bedoeling alternatieve taxatiemodellen te creëren die met behulp van sociale indicatoren in combinatie met andere indicatoren de verkoopwaarden van woningen zo nauwkeurig mogelijk kunnen schatten.

Mijn ouders en vrienden wil ik bedanken voor hun steun tijdens mijn studie. Binnen NBWO wil ik mijn stagebegeleider Jan Wigle Visser (Senior Modelontwikkelaar) bedanken voor zijn begeleiding. Ook gaat mijn dank uit naar alle andere medewerkers van NBWO die mij hebben bijgestaan. Verder wil ik mijn stagebegeleider Bert Kersten (hoogleraar van de afdeling Wiskunde & Informatica) en de tweede lezer van mijn afstudeerverslag Jan Rouwendal (hoogleraar van de afdeling Ruimtelijke Economie) bedanken voor hun begeleiding gedurende de stageperiode vanuit de Vrije Universiteit.

Amsterdam, juli 2005

Samenvatting

Probleemstellingen

In het kader van modelmatige waardebeoordelingen van woningen is statistisch onderzoek gedaan naar de invloed van zogenaamde sociale indicatoren op de verkoopwaarde van de woning. Hierbij kan men denken aan sociale indicatoren als sociale klasse, inkomensniveau, gezinssamenstelling, bevolkingsamenstelling etc. In het onderzoek moet gekeken worden naar de invloed van sociale indicatoren in combinatie met de kenmerken van de woning en lokaliteit van de woning bij het schatten van de verkoopwaarde van een woning. Hierbij zijn de onderzoeksvragen:

- Welke soort (regressie)modellen zijn in het onderzoek gebruikt?
- Welke sociale indicatoren leveren een significante bijdrage bij het schatten van de verkoopwaarde van een woning?
- Hoe groot is deze bijdrage van de sociale variabelen?
- Hoe nauwkeurig is de verkoopwaarde van de woning te schatten met sociale indicatoren in combinatie met andere indicatoren?

Doelstellingen

Een doelstelling van de opdracht is inzicht krijgen in de sociale indicatoren die de verkoopwaarde van woningen beïnvloeden. Ook is een doelstelling het ontwikkelen van regressiemodellen die de verkoopwaarde van de woning met behulp van sociale indicatoren, kenmerken van de woning en lokaliteitindicatoren (postcodegebieden) zo nauwkeurig mogelijk kunnen schatten. Hierbij wordt verwacht en gewenst dat de invloed van sociale indicatoren deels de invloed van lokaliteitindicatoren zullen vervangen.

Aanpak

Literatuuronderzoek heeft geholpen bij het verzamelen van de sociale indicatoren die relevant waren geweest voor het onderzoek. Ook heeft literatuuronderzoek een toegevoegde waarde gehad bij het creëren van ideeën omtrent het bepalen van de soort en de bouw van regressiemodellen die met behulp van sociale en andere indicatoren de verkoopwaarde van de woning zo nauwkeurig mogelijk moeten kunnen schatten. Na het literatuuronderzoek is begonnen met het verzamelen en verwerken van data. Vervolgens hebben de statistische analyses plaatsgevonden, waarbij verschillende logaritmisch getransformeerde lineaire meervoudige regressiemodellen zijn opgesteld en kort toegelicht. Deze regressiemodellen zijn tot stand gekomen doordat begonnen is met het opstellen eenvoudige regressiemodellen. Deze modellen zijn verder uitgebreid tot steeds grotere en complexere modellen. Het uitbreiden van de modellen bestond onder andere uit toevoegen van nieuwe soorten variabelen. Van die modellen zijn de beste twee geselecteerd en uitgebreider behandeld. Hierbij is vooral gelet op de bijdrage van sociale indicatoren. Tenslotte zijn deze modellen vergeleken met de huidige (kenmerken)modellen van modelmatig waardebeoordeling bij NBWO.

Voornaamste onderzoeksresultaten

“Vertrouwelijk”

Conclusies

“Vertrouwelijk”

Aanbevelingen

“Vertrouwelijk”

INHOUDOPGAVE

1. INLEIDING	8
1.1 PROBLEEMSTELLINGEN	8
1.2 DOELSTELLINGEN	8
1.3 STAGEPLAATS	9
1.4 GESCHIEDENIS MODELMATIGE WAARDEBEPALING	9
1.5 OPBOUW VAN HET VERSLAG	9
DEEL I	10
2. NBWO	11
2.1 ORGANISATIE	11
2.2 MODELMATIG WAARDEREN	12
2.3 INFORMATIE PRODUCTEN	12
2.3.1 Voor hypotheekmaatschappij	12
2.3.2 Voor intermediair	13
2.3.3 Voor gemeentes	13
2.4 NBTO	15
2.4.1 De taxatie is maatwerk	15
2.4.2 Waartoe dient de modelwaarde?	15
2.4.3 Wat doet de taxateur?	16
DEEL II	17
3. DE MODELLEN VAN NBWO	18
3.1 INDEXATIE WONINGWAARDEN UIT HET VERLEDEN MBV HET INDEXATIE-MODEL	18
3.2 MODELMATIGE WAARDEBEPALING (ONLINE WAARDE CHECK)	18
3.3 TAXATIEMODEL	21
3.4 GEBRUIKTE DATA BINNEN DE MODELLEN	21
4. OPZET EN UITVOERING VAN HET ONDERZOEK	22
4.1 DOEL ONDERZOEK	22
4.2 ONDERZOEKSPOPULATIE	22
4.3 MEETNIVEAUS VARIABELEN	22
4.4 VERWERKING EN PREPARATIE VAN DE GEGEVENS	23
4.4.1 Programma's	24
4.4.2 Omschrijving en codering variabelen	24
4.5 REGRESSIE-ANALYSE	29
4.5.1 Het regressiemodel 1	30
4.5.2 Het regressiemodel 2	33
4.5.3 Schatten van de parameters	35
4.5.4 Evaluatiematen van het model	36
4.5.5 Analyse residuen	37
4.5.6 Directe en stapsgewijze regressie	38
5. RESULTATEN ONDERZOEK	40
5.1 BESCHRIJVENDE STATISTIEK	40
5.2 REGRESSIEANALYSE	40
5.3 BESTE REGRESSIEMODELLEN	40
5.3.1 Regressiemodel 6A	40
5.3.2 Regressiemodel 6B	40
5.4 ANALYSE VAN RESIDUEN	40
5.4.1 Analyse van residuen regressiemodel 6A	40
5.4.2 Analyse van residuen regressiemodel 6B	40

6 VERGELIJKING ALTERNATIEVE EN HUIDIGE MODELLEN	41
7. CONCLUSIES & AANBEVELINGEN	42
8. BIJLAGEN.....	43
8.1 BIJLAGE: CLASSIFICATIE SOCIALE KLASSEN	43
8.2 BIJLAGE: OUTPUT SPSS REGRESSIEMODEL 6A	43
8.3 BIJLAGE: OUTPUT SPSS REGRESSIEMODEL 6B	43
8.4 BIJLAGE: GRAFIEKEN RESIDUENANALYSE REGRESSIEMODEL 6A.....	43
8.5 BIJLAGE: GRAFIEKEN RESIDUENANALYSE REGRESSIEMODEL 6B.....	43
9. LITERATUUROPGAVE	44

1. Inleiding

1.1 probleemstellingen

De vraag naar snel beschikbare, actuele en betrouwbare informatie over de waarde van onroerende zaken groeit. Overheden die hun eigen belastingcapaciteit willen vaststellen, bedrijven die de boekwaarde van hun objecten willen kennen en particulieren die een woning willen kopen of verkopen, zijn gebaat bij de juiste informatie. NBWO voorziet zijn klanten zoveel mogelijk in hun behoeften naar de waarde van woningen door middel van modelmatige waardebeoordelingen. Modelmatige waardebeoordeling is het taxeren van de waarde van een onroerende zaak door^[7]:

- het systematisch uit de marktanalyse afleiden welke objectkenmerken relevant zijn voor de waarde;
- het uit de geregistreerde objectkenmerken van de te taxeren panden halen van de kenmerken die relevant zijn voor de waarde;
- het berekenen van een waarde door een directe vergelijking met een verkocht pand of met meerdere verkochte panden;
- het beoordelen van de waarschijnlijkheid dat de berekende waarde een correcte weergave is van de waarde;
- het afzonderlijk verifiëren van de waarde indien er onvoldoende zekerheid bestaat dat de berekende waarde correct is .

NBWO is continue bezig met het verbeteren van haar taxatiemodellen. In het kader van de modelmatige waardebeoordelingen van woningen is statistisch onderzoek gedaan naar de invloed van zogenaamde sociale factoren op de verkoopwaarde van de woning. Hierbij kan men denken aan sociale factoren als sociale klasse, inkomensniveau, gezinssamenstelling, bevolkingsamenstelling etc. In het onderzoek is gekeken naar de invloed van sociale factoren in combinatie met de kenmerken van de woning en lokaliteit van de woning (postcodegebied).

Literatuuronderzoek heeft geholpen bij het selecteren van de sociale indicatoren die relevant zijn geweest voor het onderzoek. Ook heeft literatuuronderzoek een toegevoegde waarde gehad bij het creëren van ideeën omtrent het bepalen van de soort en de bouw van regressiemodellen die zo goed mogelijk moesten kunnen worden toegepast bij het schatten van verkoopwaarden van woningen. Dit zijn regressiemodellen die met behulp van sociale indicatoren en andere indicatoren de verkoopwaarde van de woning zo nauwkeurig mogelijk kunnen schatten.

Welke soort (regressie)modellen zijn in het onderzoek gebruikt? Welke sociale indicatoren leveren een significante bijdrage bij het schatten van de verkoopwaarde van een woning? Hoe groot is de bijdrage van de sociale variabelen? Hoe nauwkeurig is de verkoopwaarde van de woning te schatten met sociale indicatoren in combinatie met andere indicatoren? Dit zijn onderzoeksvragen die zo goed mogelijk worden beantwoord in dit verslag.

Sinds 1993 heeft NBWO data verzameld van verkooptransacties van woningen. Deze data samen met andere data afkomstig van instellingen als het Kadaster en Dataland zijn benut voor het onderzoek.

1.2 Doelstellingen

Een doelstelling van de opdracht is inzicht te krijgen in de sociale indicatoren die de verkoopwaarde van woningen beïnvloeden. Ook is een doelstelling het ontwikkelen van regressiemodellen om de verkoopwaarde van de woning met behulp van sociale factoren,

kenmerken van de woning en lokaliteit (zoals postcodegebieden) te schatten. Hierbij wordt verwacht en gewenst dat de invloed van sociale indicatoren deels de invloed van lokaliteitindicatoren (postcodegebieden en straten) zullen vervangen.

De modelmatige waardebeoordeling is gebaseerd op de combinatie van de volgende doelstellingen^[7]:

- het objectiveren van de waardebeoordeling;
- het reduceren van de kosten;
- het inzichtelijk maken van de waardebeoordeling voor belanghebbenden;
- het vastleggen en vergroten van de kennis van de marktverwerking;
- het bewaken van de kwaliteit van de waardebeoordeling (conform marktniveau, vergelijkbaarheid object).

Door verbeteringen proberen in te brengen op de huidige taxatiemodellen kan NBWO deze doelstellingen nog beter verwezenlijken. Hiermee kan het bedrijf een betere concurrentiepositie verkrijgen, waarbij op lange termijn de kans groter wordt dat de huidige klanten behouden kunnen worden en potentiële klanten aangetrokken kunnen worden.

1.3 Stageplaats

De stage van vier maanden heeft plaats gevonden bij de organisatie NBWO, Nederlands Bureau Waardebeoordeling Onroerende zaken, in Amsterdam. Hier houdt men zich bezig met modelontwikkeling om woningwaarden te kunnen taxeren. Een ander kantoorgebouw van NBWO, ligt in Amersfoort. In dit gebouw werkt men aan de implementatie van de ontwikkelde taxatiemodellen en men werkt daar ook aan andere automatiseringtaken.

1.4 Geschiedenis modelmatige waardebeoordeling

Modelmatige waardebeoordeling van woningen is geen nieuw concept. Dit wordt al jaren in de Verenigde Staten gedaan. Ook in Nederland is dit concept niet al te nieuw gezien instellingen, zoals taxatiebureaus, gemeenten etc. al een tijdje bezig zijn met modelmatig taxeren. Net zoals deze instellingen is NBWO ook zo een instelling die actief bezig is met het modelmatig bepalen van de waarde van woningen. NBWO is wel een van de eerste die met haar model actief om de markt is. Dit model wordt op grote schaal aangeboden en met grote regelmaat gebruikt. Dit model kan gebruikt worden als bijlage in een taxatierapport, of om fysieke taxatierapporten modelmatig te scannen op mogelijke fraude, of als instrument voor een betere portfeuilleanalyses etc.

1.5 Opbouw van het verslag

Deze scriptie bestaat uit twee delen. In deel 1 die hoofdstuk 2 bevat, wordt algemene informatie gegeven over het bedrijf NBWO. Deel 2 begint bij hoofdstuk 3, hierin wordt informatie gegeven over de huidige taxatiemodellen in het bedrijf. Vervolgens wordt in hoofdstuk 4 uitgelegd hoe het onderzoek opgezet en uitgevoerd is. In hoofdstuk 5 worden de resultaten besproken. In hoofdstuk 6 worden de alternatieve regressiemodellen vergeleken met de huidige (kenmerken)modellen van modelwaardige waardebeoordeling bij NBWO. Conclusies en aanbevelingen staan in hoofdstuk 7. Voor de bijlagen wordt verwezen naar hoofdstuk 8. Tenslotte, zijn in hoofdstuk 9 de gebruikte literatuurbronnen te vinden.

-NBWO-

DEEL I

-NBWO-

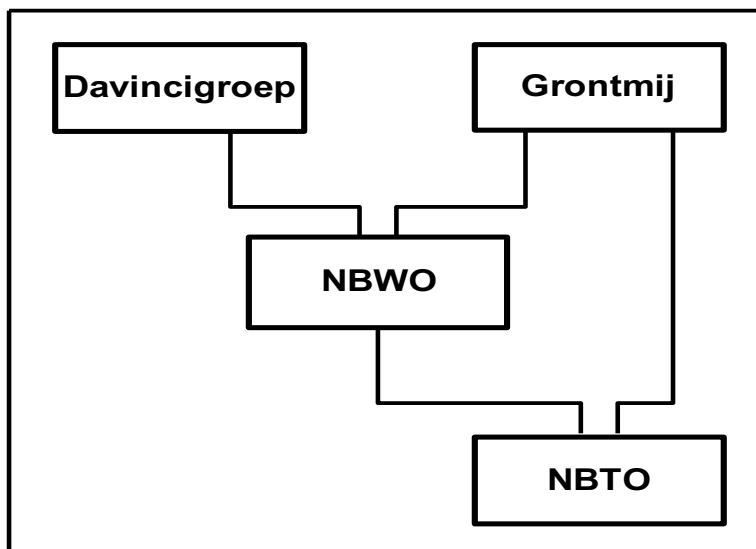
2. NBWO

2.1 Organisatie

NBWO staat voor Nederlands Bureau Waardebepaling Onroerende Zaken en heeft de vaardigheid ontwikkeld om Nederlandse woningen volledig modelmatig te taxeren. NBWO levert producten aan doelgroepen als hypotheekmaatschappijen, verzekeraars, intermediairs, makelaars en overheidsinstanties. Met haar innovatieve producten heeft NBWO in de loop der jaren een grote klantenkring opgebouwd.

Met het NBWO-model kunnen particuliere woningen in enkele seconden worden getaxeerd. Het model doet dit aan de hand van uitgebreide statistische analyses van verkoopgegevens en daarbij geregistreerde woningkenmerken. Voor deze analyses put het model uit een eigen database waarin onder andere alle sinds 1993 verrichte particuliere verkooptransacties geregistreerd staan. Momenteel beslaat deze database ruim 2.200.000 transacties, op maandbasis worden hier gemiddeld tussen de 15.000 en 20.000 nieuwe transacties aan toegevoegd.

NBWO is een samenwerkingsverband van Davinci Advies BV en Grontmij Taxaties vastgoed. Grontmij is als projectontwikkelaar al jaren actief op de vastgoedmarkt en heeft daardoor veel kennis van de woningmarkt. Davinci is een adviesbureau met jarenlange ervaring in het adviseren aan o.a. financiële instellingen en overheidsinstanties. Door deze combinatie van kennis en ervaring is NBWO in staat toegevoegde waarde te creëren voor haar klanten. Het Nederlands Bureau Taxaties Onroerende zaken (kortweg NBTO) is in 2001 ontstaan als samenwerkingsverband van enerzijds het landelijk opererend ingenieursbureau Grontmij en anderzijds het Nederlands Bureau Waardering Onroerende zaken (NBWO). In figuur 2.1 is de bedrijfsstructuur weergegeven.



Figuur 2.1 Bedrijfsstructuur NBWO.

2.2 Modelmatig waarden

Modelmatige woningwaardering staat voor het taxeren van een woning zonder dat deze daarvoor hoeft te worden bezichtigd. Op basis van verzamelde gegevens en statistische analyses kan desondanks een betrouwbare taxatie worden gegeven. Voordelen van modelmatige woningwaardering zijn onder andere:

- Veel goedkoper dan fysieke taxatie doordat geen taxateur nodig is;
- Woningen kunnen in enkele seconden worden getaxeerd;
- Grotere objectiviteit dan fysieke taxatie, doordat bij de waardering gebruik wordt gemaakt van objectieve gegevens.

Het modelmatig waarden van woningen is begin jaren negentig ontwikkeld in de Verenigde Staten. Het gebruik van woningwaarde modellen is hier al volledig geaccepteerd. Een groot aantal bedrijven en overheidsinstellingen bieden inmiddels diensten aan op basis van woningwaarde modellen.

De betrouwbaarheid van woningwaarde modellen is in de afgelopen jaren sterk toegenomen. Mede hierdoor is de afzet van woningwaarde modellen in de VS sterk gestegen. Hypotheekmaatschappijen en overheidsinstellingen zijn in de VS de grootste afnemers van modelmatige waardebeoordelingen.

De trend van modelmatig waarden van woningen, is halverwege de jaren negentig door NBWO naar Nederland gehaald. In eerste instantie stonden bedrijven huiverig tegenover deze innovatie. Veel bedrijven twijfelden over de betrouwbaarheid van de modellen. Inmiddels heeft deze modelmatige woningwaardering zich voor veel bedrijven bewezen. Bedrijven uit verschillende branches hebben met woningwaarde-modellen structurele besparingen weten te realiseren.

Voor uitleg van de taxatiemodellen die NBWO gebruikt, wordt verwezen naar hoofdstuk 3: *De modellen van NBWO*.

2.3 Informatie producten

De toepassingen, zoals NBWO deze heeft ontwikkeld op basis van het model, richten zich vooral op de hypotheekmaatschappij, het intermediair en gemeentes. NBWO kent een onderscheid naar toepassingen gericht op commercie en toepassingen gericht op risicobeheersing, kapitaalmanagement en de Wet WOZ. Deze worden besproken bij:

- Voor hypotheekmaatschappij (2.3.1);
- Voor intermediair (2.3.2);
- Voor gemeentes (2.3.3).

2.3.1 Voor hypotheekmaatschappij

De waarde van het onderpand speelt op verschillende momenten in het hypotheekproces een rol. Op basis van het NBWO model zijn verschillende producten ontwikkeld voor hypotheekmaatschappijen. Met deze producten kunnen in elke fase van het hypotheekproces structurele verbeteringen worden aangebracht.

2.3.2 Voor intermediair

Voor de intermediair, zijn er diverse diensten van NBWO die van toegevoegde waarde kunnen zijn voor haar organisatie.

Met het kenmerkenmodel kunnen bezoekers van de website van de intermediair op een snelle en eenvoudige wijze de actuele waarde van hun woning bepalen.

De klanten van intermediair wijzen op een eventuele overwaarde kan met de Overwaarde Calculator van NBWO.

Voor het intermediair die haar taxaties wil uitbesteden, NBWO werkt samen met 2 taxatiebureaus, welke het modelwaarde rapport van NBWO leveren bij hun taxatie. Dit zijn NBTO en Spunt20.

2.3.3 Voor gemeentes

WOZ Waarderingscentrum

De herwaardering in het kader van de Wet WOZ legt telkens weer een behoorlijk beslag op de ambtelijke organisatie. De gemeente heeft er belang bij dit proces zo goedkoop en snel mogelijk uit te voeren. Bovendien bepaalt de nauwkeurigheid van de (modelmatig) vastgestelde waarden in hoge mate de omvang van de uit te voeren vervolgwerkzaamheden.

Het WOZ Waarderingscentrum bundelt specifieke expertise

Het WOZ Waarderingscentrum is opgericht door het landelijk opererende ingenieursbureau Grontmij en het Nederlands Bureau Waardering Onroerende zaken (NBWO). Grontmij is sinds jaar en dag actief op het gebied van waardering, ontwikkeling, beheer en exploitatie van vastgoed. NBWO ontwikkelt sinds 1996 geautomatiseerde modellen waarmee taxaties van onroerende zaken kunnen worden uitgevoerd. Die kennis en kunde is nu gebundeld om de gemeente te ondersteunen bij de uitvoering van de Wet WOZ.

WOZ-waarderingen: snel, correct én voordelig

Het WOZ Waarderingscentrum treedt op als het logische en praktische verlengstuk van elke taxatieafdeling. NBWO voert de WOZ-processen zeer efficiënt en snel uit. Hiervoor gebruiken zij gestandaardiseerde en geautomatiseerde waarderingsmodellen en hun uitgebreide kennis van WOZ-processen. Met minimale inspanning van de gemeente kan men beschikken over complete en correcte waarderungen inclusief de taxatierapporten. Uiteraard levert dit de gemeente een flinke kostenbesparing op.

WOZ-waarderingen op een innovatieve manier

Innovatieve oplossingen zijn de kracht van deze dienstverlening. Het WOZ Waarderingscentrum werkt met een modelmatige, volledig geautomatiseerde waardebeoordeling, waarmee woningen in enkele seconden kunnen worden getaxeerd. Dat doen ze op basis van historische en actuele marktgegevens en statistische analyses. Het levert een betrouwbare WOZ-waardering op, zonder dat daarvoor een tijdrovende fysieke bezichtiging noodzakelijk is.

Modelmatige waardebeoordeling is objectief en nauwkeurig

Het model dat zij gebruiken is door NBWO op basis van woninggegevens van het Kadaster en Dataland ontwikkeld. Daarbij is gebruik gemaakt van alle verkooptransacties sinds 1993. Momenteel zijn dat ruim 2,5 miljoen transacties; op maandbasis worden hier gemiddeld tussen de 15.000 en 20.000 nieuwe transacties aan toegevoegd. Uiteraard worden alle gegevens uit uw WOZ-administratie door hen gebruikt om de waardebeoordeling op correcte

gegevens te baseren. Met diverse statistische en taxatietechnische modellen, waarin de kennis in het kader van de Wet WOZ van Grontmij is verwerkt, wordt vervolgens een aantal mogelijke uitkomsten bepaald. Aan de hand van deze uitkomsten wordt ten slotte één gewogen WOZ-waarde bepaald, waarbij rekening is gehouden met de behaalde nauwkeurigheid van de onderliggende modellen. Deze betrouwbaarheidstoets is voor de gemeente van belang om inzicht te krijgen in de omvang van de uit te voeren vervolgwerkzaamheden.

Waarderingen conform de Wet WOZ

De gemeente behoudt de regie over de uitvoering zodat gemeente niet alleen marktconform maar zeker ook WOZ-conform waardeert. Nadat de gemeente opdracht heeft gegeven voor het uitvoeren van de waardebepaling, levert NBWO de gemeente binnen één week de eerste resultaten. Deze resultaten worden bovendien op hun betrouwbaarheid getoetst. De gemeente heeft daardoor volledig inzicht in de afwijkingen of twijfelgevallen. Vervolgens kan de gemeente gericht bepalen welke werkzaamheden er uitgevoerd moeten worden om tot een waardebepaling te komen die geheel voldoet aan de wettelijke eisen en, indien gewenst, aan de eisen van de gemeente.

Gebruik regionale marktgegevens

Bij de modelmatige waardebepaling maakt NBWO gebruik van regionale marktgegevens. De WOZ-waarderingen van de gemeente worden niet alleen met marktgegevens van hun gemeente uitgevoerd, maar tevens met die van aangrenzende gemeentes.

Beperkte doorlooptijd

De voorlopige resultaten van de modelmatige waardering zijn binnen een week beschikbaar. Een volledige waardering van de objecten in een gemeente kan binnen een maand worden afgerond.

Kostenbesparend

Het gebruik van geavanceerde datamodelen beperkt de doorlooptijd, waardoor de kosten beter beheersbaar en lager worden.

Extra zekerheid door meervoudige waardering

NBWO levert een waarde op die gewogen is aan de hand van twee waarden, waarvan de ene is gebaseerd op taxatietechnische en de andere op statistische uitgangspunten. Dit wordt de betrouwbaarheidstoets genoemd. De betrouwbaarheidstoets zorgt niet alleen voor een goed beargumenteerde waarde, maar bovendien voor extra zekerheid. Zij leveren de gemeente waardering en controle in één.

Volledig transparant voor burger en gemeente

De gemeente heeft inzicht in het gehele WOZ-waarderingsproces. Telkens kan de gemeente de beslissing nemen over de volgende stap. Dit kan de gemeente bovendien aan de burger laten zien. De burger is daardoor verzekerd van een objectieve waarderingmethode die marktconform is.

De partners van het WOZ Waarderingscentrum

Het Nederlands Bureau Waardebepaling Onroerende Zaken, kortweg NBWO heeft de vaardigheid ontwikkeld om Nederlandse woningen volledig modelmatig te taxeren. NBWO is een samenwerkingsverband van Davinci Advies BV en Grontmij Nederland BV. NBWO adviseert o.a. financiële instellingen en overheidsinstanties en brengt kennis in van modelmatige waardebepaling en van financiële processen en systemen. Met innovatieve producten heeft NBWO een grote klantenkring opgebouwd onder hypotheekmaatschappijen, verzekeraars, intermediairs, makelaars en overheidsinstanties.

Grontmij beschikt over een compleet en modulair dienstenpakket voor het proces van uitvoering van WOZ-dienstverlening. Uitgangspunt voor de dienstverlening zijn de bestaande of gewenste werkprocessen in een gemeente. De combinatie met de dienstverlening van NBWO maakt die dienstverlening compleet. Grontmij gaat uit van de bestaande systemen die een gemeente gebruikt. Indien gewenst zet Grontmij eigen systemen in, zo nodig op locatie. De voordelen van deze aanpak zijn evident: de communicatielijnen zijn kort en er vindt voortdurend uitwisseling van kennis en informatie plaats. De specialisten van Grontmij zijn inzetbaar voor alle soorten werkzaamheden en objecten. Door het landelijk vestigingennetwerk zijn zij bovendien goed bekend met de regionale en lokale marktomstandigheden.

2.4 NBTO

Het Nederlands Bureau Taxaties Onroerende zaken (kortweg NBTO) is in 2001 ontstaan als samenwerkingsverband van enerzijds het landelijk opererend ingenieursbureau Grontmij en anderzijds het Nederlands Bureau Waardering Onroerende zaken (NBWO). NBWO ontwikkelt sinds 1996 geautomatiseerde modellen waarmee waardebepalingen van onroerende zaken kunnen worden verricht. Grontmij is sinds jaar en dag actief op het gebied van ontwikkeling, beheer en exploitatie van vastgoed, en heeft uit hoofde daarvan expertise op het gebied van het taxeren van onroerend goed.

2.4.1 De taxatie is maatwerk

Na de aanvraag voor een taxatie, worden verschillende activiteiten uitgevoerd:

- het verzamelen van kadastrale gegevens van het te taxeren object;
- het toevoegen van de modelmatige waardebeoordeling aan het rapport;
- het opdragen van de taxatie aan een taxateur in de regio;
- het maken van een afspraak voor bezichtiging door de taxateur;
- de fysieke inspectie van het object en het uitwerken van de gegevens;
- het vervaardigen van het rapport en het uitvoeren van kwaliteitscontrole;
- het goedkeuren en verzenden van het taxatierapport.

Door de keuze van een bekwame taxateur in de regio en door diens bezoek aan het te taxeren object, wordt de kwaliteit van de taxatie gewaarborgd. Hoewel het rapport een gestandaardiseerde lay-out heeft, wordt door een of meer foto's en beschrijvingen van het object in het rapport het individuele karakter van het object optimaal weergegeven.

2.4.2 Waartoe dient de modelwaarde?

De door de taxateur-fysiek-vastgestelde waarden van het object zijn de uitkomsten van een door de taxateur ingesteld onderzoek naar vele facetten van het object. Om deze uitkomsten te ondersteunen wordt gebruik gemaakt van de modelmatige waardebeoordeling; deze treedt niet in de plaats van de vaststelling door de taxateur, maar kan zeer goed ter ondersteuning en/of aanvulling behulpzaam zijn bij het werk van de taxateur.

De modelwaarde kan ook worden gebruikt door o.a. banken en financiële instellingen bijvoorbeeld voor portefeuille onderzoek, fraudedetectie, en vele andere toepassingen.

2.4.3 Wat doet de taxateur?

Het werk van de taxateur is gebonden aan normen die daarvoor landelijk door de brancheorganisaties zijn ontwikkeld. Dit houdt in dat naast de lay-out van het taxatierapport, ook de werkzaamheden van de taxateur zijn gestructureerd.

Na ontvangst van de taxatieopdracht, via het kantoor van NBTO, maakt de taxateur als eerste een afspraak voor opname van het te taxeren object. Hij bereidt op kantoor de opname voor door het raadplegen van beschikbare documentatie, het eventueel opvragen van specifieke informatie over de omgeving bij instanties en registers, en het verrichten van enkele andere administratieve werkzaamheden.

Tijdens de opname van het object, zowel extern als intern, noteert de taxateur alle van belang zijnde gegevens van het object. Deze worden daarna in het taxatiesysteem van NBTO ingevoerd. Tevens onderzoekt de taxateur andere factoren die voor het taxatierapport van belang zijn; daarbij dient gedacht te worden aan de eigendomsakte van het object, eventuele gevestigde zakelijke rechten (erfpacht, recht van overpad), milieu omstandigheden, anti speculatie beding etc. Daarbij kan hij zich laten bijstaan door lokale instanties en of andere deskundigen.

-NBWO-

DEEL II

-NBWO-

3. De modellen van NBWO

NBWO beschikt voor het (her)waarderen van objecten over een drietal methodieken:

- A) Indexatie van de woningwaarde uit het verleden middels het indexatiemodel;
- B) Volledige 'herberekening' van de actuele vrije verkoopwaarde middels het woningwaarde-model (online waarde check);
- C) Het gebruik van taxatiegegevens voor het waarderen van de woning (taxatiemodel).

Onderstaand wordt in paragraaf 3.1 t/m 3.3 ingegaan op de werking van de modellen. In paragraaf 3.4 wordt wat verteld over de gebruikte data binnen de modellen.

3.1 Indexatie woningwaarden uit het verleden mbv het indexatie-model

Elke maand genereert NBWO voor alle woningen in Nederland waarde-indexen. Deze indexen zijn een weergave van de waardeontwikkeling van ieder afzonderlijke woning ten opzichte van een waarderingsmoment gelegen in het verleden.

NBWO put hiervoor o.a. uit de Kadaster-transactiedatabase waarin alle woningverkopen in Nederland sinds 1993 worden bijgehouden.

Veel bekende woningwaarde-indexen zijn gevoelig voor afwijkingen door a-typische verkopen in een meetperiode. Het kan bijvoorbeeld voorkomen dat bovengemiddeld veel goedkope woningen zijn verkocht, waardoor de indexen een prijsdaling aangeven. Maar dat hoeft niet te gelden voor de gehele markt. NBWO maakt alleen gebruik van transacties van woningen die twee of meer keer zijn verkocht. Met Repeat Sales methode worden de transactiepreizen van de herhaalde verkopen van dezelfde woning vergeleken om tot schattingen van prijsindexcijfers te komen. Deze methode is aanmerkelijk ongevoeliger voor afwijkingen en geeft een nauwkeuriger beeld van de werkelijke markttrends. NBWO berekent indexen per woningmarktgebied (verzameling van postcodegebieden) en per woningtype.

Zoals genoemd kan NBWO haar eigen indexen bepalen sinds 1993. Voor de periode 1985 tot 1993 wordt gebruik gemaakt van NVM indexen. Indexen van voor 1985 heeft NBWO niet beschikbaar.

3.2 Modelmatige waardebepaling (online waarde check)

De waarde van de woning wordt bepaald op basis van 4 submodellen, welke gecombineerd de uiteindelijke waarde van de woning bepalen. Het gaat om de volgende modellen:

1. Waarde op basis van vorige verkoop van de woning
2. Waarde op basis van specifieke kenmerken van de woning
3. Waarde op basis van verkopen in het postcodegebied
4. Waarde op basis van verkopen in de straat

1. Waarde op basis van vorige verkoop van de woning

Het vorige verkoop model gaat uit van de meest recente verkooptransactie zoals bij het Kadaster bekend is. De koopsom die hoort bij deze verkooptransactie wordt geïndexeerd naar de huidige datum.

Er wordt geen rekening gehouden met:

- een waardevermeerdering van de woning als gevolg van bijvoorbeeld een verbouwing.
- een waardevermindering van de woning als gevolg van bijvoorbeeld slecht onderhoud.
- een niet representatieve koopsom (wanneer de koopsom lager / hoger is dan de algemene woningwaarde op moment van verkoop), bijvoorbeeld bij familieverkoop of onderlinge verkoop van een woning.

2. Waarde op basis van specifieke kenmerken van de woning

Dit model bepaalt de waarde van de woning op basis van specifieke kenmerken van de woning in combinatie met verkooptransacties in de directe omgeving. Bij het selecteren van deze verkooptransacties wordt rekening gehouden met verschillen in woningtype, perceelgrootte, inhoud, perceelgrootte, bouwjaar etc. van de woning. In tabel 3.2 worden deze gegevens beschreven.

Hier wordt gebruik gemaakt van meervoudige regressie met logaritmische transformatie en dummyvariabelen. Voor een overzicht van alle gebruikte variabelen, zie tabel 3.2.

3. Waarde op basis van verkopen in het postcodegebied

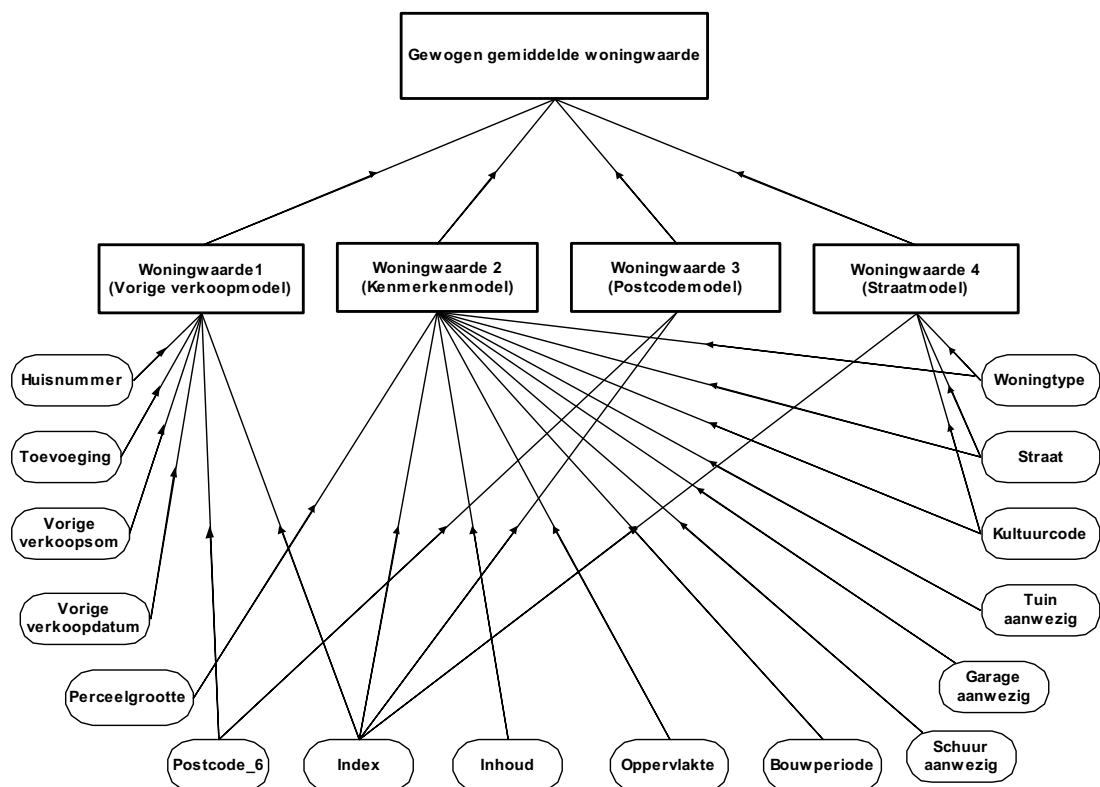
De waarde bepaald door het postcodemodel is een gemiddelde van alle verkooptransacties die recent in de betreffende postcode hebben plaatsgevonden. Het vorige kwartaal ten opzichte van het kwartaal waarin de waardebepaling plaats vindt, is voldoende recent. Hierbij worden niet representatieve koopsommen zoveel mogelijk uitgesloten, zoals de verkoop van een winkelwoning.

4. Waarde op basis van verkopen in de straat

De waarde bepaald door het straatmodel is een gemiddelde van alle koopsommen van hetzelfde woningtype (hoekwoning, tussenwoning etc.) die recent in de betreffende straat hebben plaatsgevonden. Hierbij worden niet representatieve koopsommen zoveel mogelijk uitgesloten, zoals de verkoop van een woning met een praktijkruimte.

Aan de hand van de uitkomsten van deze vier modellen wordt tenslotte één gewogen waarde gegenereerd die rekening houdt met de behaalde nauwkeurigheid van de vier modellen.

In figuur 3.2 is een schematische weergave te zien van huidige eindmodel van modelmatig waardebepaling. Per submodel wordt aangeven welke gegevens gebruikt worden en de omschrijving van deze gegevens staan in tabel 3.2.



Figuur 3.2 Schematische weergave NBWO-modellen van modelmatige waardebeoordeling.

In de tabel 3.2 staan de beschrijvingen van de gegevens (indicatoren) van het huidige model van modelmatig waardebeoordeling.

Nr	Indicatoren	Beschrijving
1	Postcode_6	De volledige postcode (2 letters en vier cijfers)
2	Huisnummer	Huisnummer van de woning.
3	Toevoeging	Toevoeging aan het huisnummer.
4	Kultuurcode	11= eengezinswoningen en 12= Appartementen.
5	Woningtype	V = Vrijstaand woning : woning dat niet aangebouwd is aan een ander gebouw. T = Tussenwoning: rijtjeswoning grenst aan beide kanten aan ander gebouw. H = Hoekwoning: woning dat grenst aan een kant aan een ander gebouw. K = Twee-onder-een-kap: twee woningen onder één dak. O = Overige woning: woning die geen Vrijstaand woning, Tussenwoning, Hoekwoning of Twee-onder-een-kap is.
6	Straat	De straat waar de woning staat.
7	Perceelgrootte	Aantal vierkante meters grond waar de woning op staat
8	Inhoud	Aantal kubieke meters van de woning.
9	Oppervlakte	Aantal vierkante meters vloeroppervlakte binnen de woning
10	Bouwperiode	De tijdperiode waarin de woningen in een postcodegebied gebouwd zijn.
11	Garage aanwezig	Ja/nee, garage aanwezig
12	Tuin aanwezig	Ja/nee, tuin aanwezig
13	Schoor aanwezig	Ja/nee, schoor aanwezig
14	Index	Index van de woningprijzen
15	Vorige verkoopsom	Vorige verkoopprijs
16	Vorige verkoopdatum	Vorige verkoopdatum

Tabel 3.2 Beschrijving van de gegevens van het huidige model bij modelmatige waardebeoordeling.

3.3 Taxatiemodel

Het taxatiemodel is een model dat verder gaat dan de online waarde check. Extra invulling van gegevens of gegevens die verkregen zijn via een taxatie zijn extra kenmerken waarbij NBWO in staat is om een waardebeoordeling uit te voeren met weging van alle relevante onderdelen.

Aspecten als bijvoorbeeld een verbouwing of ligging kunnen hierdoor meegenomen worden wat bij geen van de andere modellen (i.v.m. het ontbreken van de benodigde gegevens) mogelijk is.

3.4 Gebruikte data binnen de modellen

Binnen de verschillende modellen wordt gebruik gemaakt van de transactiedatabase van het Kadaster waar vanaf 1993 verkopen van objecten met doel woonbestemming staan geregistreerd. Deze database beslaat per september 2004 zo'n 2.319.000 gegevens waarvan 1.723.000 unieke objecten.

De gegevens van het Kadaster wordt sinds 2002 aangevuld met Dataland-gegevens. Dataland is een initiatief van en door de Nederlandse gemeenten en bevat gemeentelijke gegevens zoals postcode, x- en y-coördinaat tot bouwjaar, bestemming en bruto inhoud. Per september 2004 beslaat de database 2.750.000 gegevens waarvan 2.453.000 unieke objecten.

4. Opzet en uitvoering van het onderzoek

In de paragrafen 4.1 en 4.2 zullen in deze volgorde het doel en de doelgroep van het onderzoek worden besproken. Welke meetniveaus van variabelen gebruikt zijn, is te vinden in paragraaf 4.3. In paragraaf 4.4 wordt verteld hoe de gegevens verwerkt en voorbereid zijn om onderzocht te worden. De laatste paragraaf, 4.5 legt uit hoe regressieanalyse hier in zijn werk gaat. In deze paragraaf krijg je antwoord op onder andere de volgende vragen. Welke modellen zijn er gebruikt? Hoe wordt de kwaliteit van het model gemeten? Hoe zijn de residuen verdeeld?

4.1 Doel onderzoek

Een doelstelling van de opdracht is inzicht te krijgen in de sociale indicatoren die de verkoopwaarde van woningen beïnvloeden. Ook is een doelstelling het ontwikkelen van regressiemodellen om de verkoopwaarde van de woning met behulp van sociale indicatoren, kenmerken van de woning en lokaliteitindicatoren (postcodegebieden) zo nauwkeurig mogelijk te kunnen schatten. Hierbij wordt verwacht en gewenst dat de invloed van sociale indicatoren deels de invloed van lokaliteitindicatoren zullen vervangen.

4.2 Onderzoekspopulatie

De doelgroep van het onderzoek zijn eengezinswoningen uit de provincie Noord-Holland van het jaar 2000. Deze woningen bestaan uit vijf verschillende woningtypen, namelijk: hoekwoningen, tussenwoningen, twee-onder-een-kap, vrijstaande woningen en overige woningen. De verkoopwaarden van de woningen liggen tussen 50.000 en 2.000.000 euro. Totaal zijn ongeveer 30.000 woningen onderzocht.

4.3 Meetniveaus variabelen

De variabelen die in een statistisch onderzoek gebruikt worden, kunnen onderverdeeld worden in twee verschillende meetniveaus^[13]:

- 1) Nominaal meetniveau;
- 2) Ratio meetniveau;

1) Nominaal meetniveau

Nominaal meetniveau is een meetniveau waarbij de waarden (getallen) van een variabele alleen maar zijn te interpreteren als etiket of naamgeving.

Voorbeeld

De variabele *Woningtype* met de waarden:

V = Vrijstaande woning;

H = Hoekwoning;

T = Tussenwoning;

K = Twee-onder-een-kap;

O = Overige woning.

Om nominale variabelen te kunnen gebruiken in regressieanalyse zal elke nominale variabele omgezet moeten worden in een of meerdere *dummyvariabelen*. Een dummyvariabele is een variabele met twee mogelijke uitkomsten: 0 of 1. Een nominale variabele met twee groepen heeft één dummyvariabele nodig bij regressieanalyse. Een nominale variabele met drie

groepen heeft twee dummyvariabelen nodig. Een nominale variabele met K groepen, heeft K-1 dummyvariabelen nodig.

Voorbeeld

Als voorbeeld wordt nog eens de nominale variabele *Woningtype* gebruikt. Deze variabele maakt onderscheid tussen vijf verschillende groepen, namelijk de groepen *Vrijstaande woning*, *Hoekwoning*, *Tussenwoning*, *Twee-onder-een-kap* en *Overige woning*.

Dit betekent dat hier vier dummyvariabelen nodig zijn. Laat de dummyvariabelen d_1 , d_2 , d_3 en d_4 genoemd worden. Dan kunnen nu de verschillende groepen aangeven worden met behulp van deze dummyvariabelen. Een van de groepen zal als referentiegroep moeten dienen en deze groep zal moeten worden aangegeven door alle vier dummyvariabelen de waarde 0 te geven. In tabel 4.3 is te zien dat de groep *Overige woning* in dit voorbeeld als referentiegroep dient. Elke andere groep zal aangegeven worden met telkens een andere dummyvariabele de waarde 1 te geven en de andere dummyvariabelen de waarde 0 te geven (zie tabel 4.3).

Nr	Dummyvariabelen				Groep van de nominale variabele <i>Woningtype</i>
	d_1	d_2	d_3	d_4	
1	1	0	0	0	Vrijstaande woning
2	0	1	0	0	Hoekwoning
3	0	0	1	0	Tussenwoning
4	0	0	0	1	Twee-onder-een-kap
5	0	0	0	0	Overige woning

Tabel 4.3 Een voorbeeld van dummyvariabelen.

2) *Ratio meetniveau*

Ratio meetniveau is een meetniveau waarbij de verhouding en de verschillen tussen meetwaarden (getallen) van een variabele eveneens de verhouding en de verschillen weergeeft tussen de waarden van de begrippen in werkelijkheid. Deze schaal heeft een natuurlijk of absoluut nulpunt. Ook is er een natuurlijke ordening in de waarden.

Voorbeeld: de variabele *koopsom*; een woning van 150.000 euro kost drie maal zoveel als een woning van 50.000 euro. Op deze manier zou men niet kunnen redeneren over de verhoudingen van waarden bij schoenmaten. Men kan dus niet zeggen dat schoenmaat 44 twee keer zo groot is als schoenmaat 22. Wat wel weer gezegd kan worden is dat het verschil in waarden van woningen van 100.000 euro en 80.000 euro gelijk is aan 20.000 euro. Op deze manier zou men niet over verschil kunnen redeneren bij temperatuur als variabele. Men kan niet zeggen dat 40 graden Celsius minus 30 graden Celsius gelijk is aan 10 graden Celsius. Verder geldt dat koopsom een absoluut nulpunt van 0 euro heeft en vanaf dit nulpunt kan de koopsom hetzelfde blijven of toenemen volgens een natuurlijke ordening. Temperatuur en schoenmaat zouden als variabele nooit onder het Ratio meetniveau kunnen vallen, maar koopsom van de woning wel.

4.4 Verwerking en preparatie van de gegevens

In subparagraaf 4.4.1 wordt wat verteld over de gebruikte programma's in het onderzoek en subparagraaf 4.4.2 gaat in op de in SPSS gebruikte variabelen met de daarbij horende omschrijving en codering.

4.4.1 Programma's

In dit onderzoek wordt gebruik gemaakt van het statistische programma SPSS 12.0.1 for Windows 2003 en het rekenprogramma Microsoft Excel 2000. De data van NBWO bevindt zich in Excel-bestanden. Verwerkingen van de data worden voornamelijk in Excel gedaan. Hierbij kan je denken aan het aanmaken van dummyvariabelen, het opsporen en verwijderen van fouten in de data, het koppelen van verschillende datasets enzovoorts.

Via Excel wordt de verwerkte data ingelezen in SPSS. In SPSS worden de statistische analyses uitgevoerd. SPSS bevat ook over een data editor om gegevens in te voeren en te bewerken. Dus bewerkingen kunnen ook in SPSS worden gedaan voor zover dat niet in Excel wordt gedaan of kan worden gedaan. Tijdens de data-analyse zijn in SPSS bijvoorbeeld restricties meegegeven, zoals dat alleen bepaalde woningtypen geselecteerd mogen worden of dat de verkoopwaarde van de woning minimaal 50.000 euro mag zijn. Grafieken en tabellen worden in SPSS gemaakt voor het overzichtelijk weergeven van de uitkomsten (outputs) van het onderzoek.

4.4.2 Omschrijving en codering variabelen

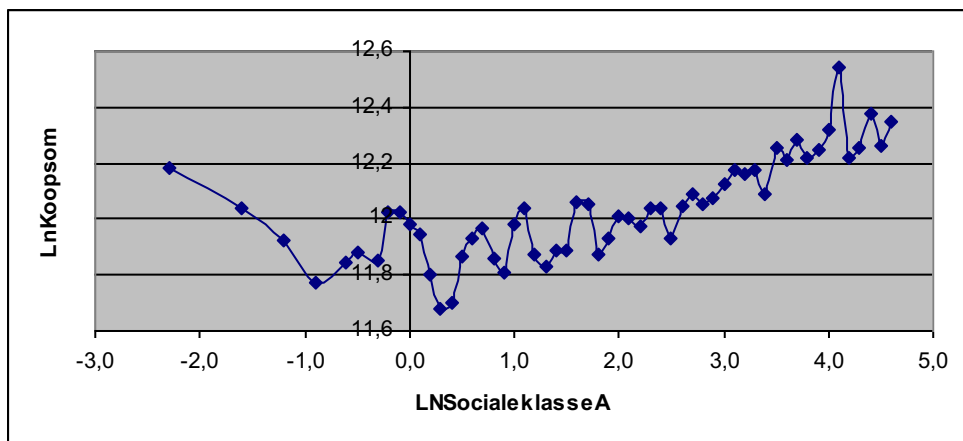
In deze paragraaf worden alle in SPSS gebruikte variabelen besproken. In tabel Tabel 4.4.2b wordt bij elke variabele verteld tot welke soort (kenmerken van de woning, sociale indicatoren, postcode2-gebieden of Residuenclassificatie postcode4-gebieden) het behoort en wat zijn omschrijving en eventuele codering is. Ook is te zien of een variabele een dummyvariabele (D) of een ratiovariabele (R) is.

Kenmerken van de woning

De variabelen uit de indicatorenverzameling kenmerken van de woning, zijn ongeveer dezelfde variabelen als de variabelen die in het huidige (kenmerken)model van NBWO worden gebruikt. Het gaat hier om 15 variabelen, waarbij elke dummyvariabele ook als aparte variabele wordt meegerekend (zie tabel 4.4.2b).

Sociale indicatoren

In het onderzoek worden 63 variabelen uit de indicatorenverzameling sociale indicatoren, gebruikt. Ongeveer de helft hiervan zijn logaritmische getransformeerde (LN) ratiovariabelen en de andere helft zijn dummyvariabelen. In tabel 4.4.2b wordt bij de (logaritmische getransformeerde) variabelen, indien ze uit de literatuur komen, verwezen naar de literatuurbronnen waar ze vandaan komen. Tevens worden in tabel 4.4.2b beschrijvingen gegeven van de sociale indicatoren. Wat betreft de dummyvariabelen, een dummyvariabele maakt hier onderscheid tussen de waarde van een logaritmische getransformeerde sociale variabele die groter dan nul is of een waarde die kleiner of gelijk aan nul is. Wanneer grafieken worden gemaakt met de logaritmische getransformeerde waarden van de sociale variabelen op de horizontale as en op de verticale as de *gemiddelde* logaritmische getransformeerde waarden van de variabele *LNKoopsom*, dan zijn bij de meeste grafieken te zien dat deze zich bij een horizontale waarde kleiner of gelijk aan nul grotere schommelingen vertonen dan bij een waarde groter dan nul. Voor alle sociale variabelen waarvoor dit verschijnsel geldt en waar ook nog voor geldt dat het om aardig wat observaties gaat die bij een horizontale waarde kleiner of gelijk aan nul grotere schommelingen vertonen, is er een dummyvariabele opgenomen. Een voorbeeld hiervan is in figuur 4.4.2 te zien.



Figuur 4.4.2 Grafiek met op de horizontale as de logaritmische getransformeerde waarden van de sociale variabele *LNSocialeklasseA* en op de verticale as de logaritmische getransformeerde gemiddelde waarden van de variabele *LNKoopsom*. Ongeveer 25% van de observaties ligt hier bij een horizontale waarde van kleiner of gelijk aan nul.

Postcode2-gebieden

Variabelen uit de indicatorenverzameling postcode2-gebieden, zijn 10 dummyvariabelen die aangeven of een woning wel of niet in bepaalde postcodegebied staat waarvan de eerste twee cijfers van de postcode van dit gebied gelijk zijn aan een van deze getallen 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 of 20 (zie tabel 4.4.2b). Dit zijn de postcode2-gebieden uit de provincie Noord-Holland.

Residuenclassificatie postcode4gebieden

Wat betreft de variabelen uit de indicatorenverzameling residuenclassificatie postcode4gebieden, het gaat hier om 12 dummyvariabelen die aangeven of een woning wel of niet tot een bepaalde residuenklasse van postcode4-gebieden behoort (zie tabel 4.4.2b). Een postcode4-gebied is een gebied dat gekenmerkt wordt door de eerste vier cijfers van haar postcode. De classificatie van postcode4-gebieden is gebaseerd op de gemiddelde positieve of negatieve residuen die ontstaan wanneer de verkoopwaarden van woningen uit dat postcode4-gebied geschat zouden worden door logaritmisch getransformeerde meervoudige lineaire regressiemodellen.

De residuenclassificatie is tot stand gekomen doordat eerst regressieanalyse heeft plaats gevonden op de traindata. Hierbij zijn de volgende soorten indicatoren gebruikt: kenmerken van de woning, sociale indicatoren, postcode2-gebieden, interacties tussen sociale indicatoren en postcode2-gebieden. Vervolgens zijn de postcode4-gebieden van laag naar hoog gesorteerd op gemiddelde residu en hiervan zijn klassen gemaakt met een bandbreedte van gemiddelde residuen van 0,1. Met behulp van dummyvariabelen wordt aangeven tot welke residuenklasse een postcode4-gebied van woning behoort.

Deze dummyvariabelen zijn samen met de andere variabelen (kenmerken van de woning, sociale indicatoren, postcode2-gebieden, interacties tussen sociale indicatoren en postcode2-gebieden) gebruikt in regressieanalyse op de traindata. Het ontstane regressiemodel is vervolgens toegepast op de waarden van de verklarende variabelen in de testdata. De schattingen die hieruit kwamen, werden vergeleken met de waarden van de te verklaren variabele van de testdata (response). Zodoende is de nauwkeurigheid van het model bepaald. Deze nauwkeurigheid verschilt nauwelijks van de nauwkeurig waarmee het model kan schatten op de traindata. Hiermee kan gezegd worden dat de classificatie voldoende representatief is voor de werkelijkheid.

Het gaat om 13 klassen en deze zijn te zien in tabel 5.4.2a. Woningen in klasse 7 en klasse 8 zijn de woningen die het nauwkeurigst kunnen worden geschat. Naarmate je opklimt van klasse 8 tot en met klasse 13 kom je postcode4-gebieden tegen waarvan de voorspelde verkoopwaarden gemiddeld steeds meer positiever afwijken ten opzichte van de werkelijke gemiddelde verkoopwaarden van de woningen uit die klasse. Naarmate je zakt van klasse 7 tot en met klasse 2 kom je postcode4-gebieden tegen met woningen waarvan de voorspelde verkoopwaarden gemiddeld steeds meer negatiever afwijken ten opzichte van de werkelijke gemiddelde verkoopwaarden van de woningen uit die klasse. Klasse 1 heeft niets te maken met hoe nauwkeurig de woningen uit deze klasse kunnen worden voorspeld, maar dit is een soort van reserveklasse van postcode4-gebieden. Deze reserveklasse bevat te weinig verkooptransacties van woningen (minder dan 5 in de traindata) om representatieve gemiddelde residuen te kunnen berekenen.

Klasse	gemiddelde residu	Postcode4-gebied(en)
1	-	1832, 1482, 1458, 1658, 1489, 1438, 1477, 1061, 1488, 1471, 1683, 1145, 1154, 1654, 1961, 1476, 1646, 1992, 1525, 1095, 1021, 1847, 1093, 1437, 2031, 1031, 1844, 1454, 1967, 1436, 1719, 1546, 1685, 1661, 1968, 1681, 1734, 1536, 1051, 1055, 1052, 1094, 1463, 1451, 1662, 1684, 1184, 1787, 1842, 1736, 1753, 1678, 1643, 1028, 1127, 1019, 1076, 1773, 1757, 1655, 1102, 1472, 1027, 1634, 1641, 1189, 1691, 1674, 1843, 1013
2	<-0,50]	1484
3	[-0,50;-0,40]	1794, 1778, 1175, 1023
4	[-0,40;-0,30]	2032, 1109, 1059, 1056, 1161, 1165, 1475, 1766, 1483, 1456, 1822, 1072, 1461, 1464
5	[-0,30;-0,20]	1033, 1442, 1443, 1486, 1057, 1103, 1795, 1104, 1607, 1107, 1797, 1474, 1505, 1053
6	[-0,20;-0,10]	1074, 1435, 1462, 1951, 1786, 1774, 1752, 1271, 1947, 1034, 1969, 1775, 1744, 1771, 1473, 1767, 1212, 1221, 1676, 1781, 1131, 1608, 2022, 1082, 1966, 1211, 1944, 1068, 1696, 1381, 1731, 1657, 1941, 1079, 1813, 1274, 1445, 1825, 1135, 1441, 1614, 1063, 1841, 1782, 1508, 1276, 1751, 1755, 1015, 1764
7	[-0,10; 0,00]	1943, 1035, 1831, 1826, 1613, 1444, 2021, 1222, 1502, 1971, 1106, 1827, 1636, 1521, 1746, 1273, 2033, 1446, 1097, 1069, 1394, 1065, 1777, 1561, 1566, 1693, 1972, 1383, 1619, 1648, 1433, 1064, 1066, 1647, 1679, 1788, 1834, 1783, 1141, 1244, 2035, 2026, 1697, 1617, 1447, 1214, 1975, 1901, 1132, 1067, 1601, 1749, 1241, 1275, 1452, 1083, 2025, 1823, 1606, 1562, 1687, 1645, 1121, 1531, 1796, 1945, 1964, 1779, 1946, 1215, 1073, 1962, 1541, 1733, 1735, 1732, 1098, 1949, 1824, 1761, 1973, 1115, 1025, 1544, 1026, 1628
8	[0,00; 0,10]	1935, 1223, 2013, 2011, 1942, 1671, 1078, 1974, 2014, 1689, 1756, 1829, 2034, 1504, 1911, 1784, 1851, 1551, 1602, 1171, 1703, 1921, 1611, 1711, 1816, 1112, 1963, 1503, 1715, 1382, 1716, 1509, 1688, 1791, 1616, 1054, 1747, 1424, 1058, 1965, 1701, 1738, 1622, 1871, 1399, 1815, 1511, 1695, 1216, 1012, 1906, 1218, 1501, 1934, 2024, 1448, 1506, 1759, 1625, 1403, 2037, 1092, 1423, 1702, 1507, 1821, 1718, 1852, 1243, 1792, 1873, 1724
9	[0,10; 0,20]	1902, 1432, 1188, 1422, 1183, 1723, 1633, 1811, 1713, 1018, 1704, 1624, 1272, 1431, 1817, 1181, 1981, 1185, 1213, 1402, 1011, 1151, 1722, 1814, 1251, 1721, 1016, 1111, 2015, 1862, 1705, 1631, 1187, 1091, 1741, 1081, 1931, 1261, 1024, 2023, 1421, 1113
10	[0,20; 0,30]	1411, 1785, 1623, 1017, 1991, 1404, 1985, 1191, 1398, 1406, 1768, 1401, 1742, 1156, 1412, 2036, 1186, 1060, 1621, 1277, 1861
11	[0,30; 0,40]	1217, 2041, 2012, 1405, 1075, 1071
12	[0,40; 0,50]	1182, 1865
13	[0,50>	1077

Tabel 4.4.2a Classificatie van postcode4-gebieden.

Interacties tussen sociale indicatoren en postcode2-gebieden

Naast de variabelen in tabel 4.4.2 zijn er variabelen die een transactie vormen tussen de variabelen van de postcode2-gebieden en de variabelen van de sociale indicatoren. NBWO vermoedt dat er mogelijk interacties zijn tussen deze twee soorten variabelen, daarom is gekozen om alle interacties tussen deze twee soorten variabelen te onderzoeken. Het gaat hier om $10 \times 33 = 330$ variabelen. Dit zijn er teveel om in deze tabel op te nemen, maar het is wel handig om te vertellen hoe deze variabelen te herkennen zijn. Deze variabelen zijn te herkennen doordat de naam van de variabele van de postcodegebieden geplakt is aan de naam van de variabele van de sociale indicatoren, bijvoorbeeld *LNAantalpersonenPC2_dummy10*.

Dit is een interactie tussen de variabele *LN*Aantalpersonen (sociale indicator) en de variabele *PC2_dummy10* (postcode2-gebied). Totaal worden dus 110+330= 440 variabelen onderzocht.

Nr	Indicatoren verzameling	D / R	Variabelen in SPSS	Omschrijving en eventuele codering
1	Afhankelijk-e variabele	R	<i>LN</i> Koopsom	LN(de taxeren verkoopwaarde van een eengezinswoning)
2	Kenmerken v/d woning	D	<i>V_dummy</i>	1 = vrijstaande woning: niet aangebouwd aan een ander gebouw en 0 = geen vrijstaande woning.
3		D	<i>T_dummy</i>	1 = tussenwoning: rijtjeshuis grenst aan beide kanten aan ander gebouw en 0 = geen tussenwoning.
4		D	<i>H_dummy</i>	1 = hoekwoning: grenst aan een kant aan een ander gebouw en 0 = geen hoekwoning
5		D	<i>K_dummy</i>	1 = twee-onder-een-kap: twee huizen onder één dak en 0 = geen twee-onder-een-kap.
6		D	<i>O_dummy</i>	1 = overige woning: woning die geen vrijstaande woning, tussenwoning, hoekwoning en twee-onder-een-kap is en 0 = geen overige woning.
8		R	<i>LN</i> Grootte	LN(aantal vierkante meters grond waar de woning op staat)
9		D	<i>Ind</i> Monument	1 = woning is een monument en 0 = anders
10		D	<i>Ind</i> Huurder	1 = de woning werd verhuurd en 0 = anders
11		D	<i>Ind</i> Erfdienstbh	1 = woning waarbij erf gedeeld wordt met derden en 0 = anders
12		D	<i>Ind</i> FamilieTransactie	1 = de woning is aan een familielid verkocht en 0 = anders
13		R	<i>LN</i> Inhoud	LN(aantal kubieke meters inhoud van de woning)
14		R	<i>LN</i> Bouwjaar	LN(het jaar waarin de woningen in een gebied gebouwd zijn)
15		D	<i>Bouwjaar</i> 1900	1 = bouwjaar voor 1900 en 0 = anders
16		D	<i>Bouwjaar</i> 1950	1 = bouwjaar voor 1950 en 0 = anders
17	Sociale indicatoren	R	<i>LN</i> Aantalhuishoudens ^[16]	LN(aantal huishoudens per postcode)
18		R	<i>LN</i> Aantalpersonen ^[15]	LN(aantal personen per postcode)
19		R	<i>LN</i> Alleenstaand ^[11]	LN(percentage alleenstaande per postcode)
20		R	<i>LN</i> Boerderijen	LN(percentage boerderijen per postcode)
21		R	<i>LN</i> Brutoinkomen123000 ^[15]	LN(Percentage huishoudens in bruto jaarinkomenklasse 123000 EUR of meer per postcode)
22		R	<i>LN</i> Brutoinkomen30000 ^[15]	LN(Percentage huishoudens in Bruto jaarinkomenklasse 30000 EUR of minder per postcode)
23		R	<i>LN</i> Brutoinkomen3000045000 ^[15]	LN(Percentage huishoudens in Bruto jaarinkomenklasse 3000-45000 EUR per postcode)
24		R	<i>LN</i> Brutoinkomen4500075000 ^[15]	LN(Percentage huishoudens in Bruto jaarinkomenklasse 45000-75000 EUR per postcode)
25		R	<i>LN</i> Brutoinkomen75000123000 ^[15]	LN(Percentage huishoudens in Bruto jaarinkomenklasse 75000-123000 EUR per postcode)
26		R	<i>LN</i> Flats	LN(percentage flats per postcode)
27		R	<i>LN</i> Geenbaan ^[6]	LN(percentage werklozen per postcode)
28		R	<i>LN</i> Gemiddeldgezinsinkomen ^[15]	LN(gemiddelde gezinsinkomen)
29		R	<i>LN</i> Getrouwd ^[11]	LN(percentage getrouwden per postcode)
30		R	<i>LN</i> HHSmetkinderen ^[11]	LN(percentage huishoudens met kinderen per postcode)
31		R	<i>LN</i> Hoekwoningen	LN(percentage hoekwoningen per postcode)
32		R	<i>LN</i> Hogeopleiding ^[11]	LN(percentage hogeropgeleiden per postcode)
33		R	<i>LN</i> Huurwoningen	LN(percentage huurwoningen per postcode)

Nr	Indicatoren verzameling	D / R	Variabelen in SPSS	Omschrijving en eventuele codering
34	Sociale indicatoren	R	<i>LNKoopwoningen</i>	LN(percentage koopwoningen per postcode)
35		R	<i>LNlageopleiding</i> ^[11]	LN(percentage laagopgeleiden per postcode)
36		R	<i>LNLeeftijd024</i> ^[6]	LN(percentage personen in leeftijdklasse 0-24 jaar per postcode)
37		R	<i>LNLeeftijd2539</i> ^[6]	LN(percentage personen in leeftijdklasse 25-39 jaar per postcode)
38		R	<i>LNLeeftijd4064</i> ^[6]	LN(percentage personen in leeftijdklasse 40-64 jaar per postcode)
39		R	<i>LNLeeftijd65</i> ^[6]	LN(percentage personen in leeftijdklasse 65 jaar en ouder per postcode)
40		R	<i>LMiddelbareopleiding</i> ^[11]	LN(percentage middelaaropgeleiden per postcode)
41		R	<i>LNOverigeburgelijkestaat</i> ^[11]	LN(percentage overige burgerlijke staat per postcode)
42		R	<i>LNOverigewoningtypen</i>	LN(percentage overige woningtypen per postcode)
43		R	<i>LNSamenwonend</i> ^[11]	LN(percentage samenwonenden per postcode)
44		R	<i>LNSocialeklasseA</i> ^[6]	LN(percentage huishoudens uit sociale klasse A per postcode) ¹
45		R	<i>LNSocialeklasseB1</i> ^[6]	LN(percentage huishoudens uit sociale klasse B1 per postcode)
46		R	<i>LNSocialeklasseB2</i> ^[6]	LN(percentage huishoudens uit sociale klasse B2 per postcode)
47		R	<i>LNSocialeklasseC</i> ^[6]	LN(percentage huishoudens uit sociale klasse C per postcode)
48		R	<i>LNSocialeklasseD</i> ^[6]	LN(percentage huishoudens uit sociale klasse D per postcode)
49		R	<i>LVrijstaandhuizen</i>	LN(percentage vrijstaande woningen per postcode)
50		D	<i>LNBoerderijenNUL</i>	1 = LNBoerderijen <= 0 en 0 = anders
51		D	<i>LNBrutoinkomen123000 NUL</i>	1 = LNBrutoinkomen123000 <= 0 en 0 = anders
52		D	<i>LNBrutoinkomen30000 NUL</i>	1 = LNBrutoinkomen30000 <= 0 en 0 = anders
53		D	<i>LNBrutoinkomen3000045000 NUL</i>	1 = LNBrutoinkomen3000045000 <= 0 en 0 = anders
54		D	<i>LNBrutoinkomen4500075000 NUL</i>	1 = LNBrutoinkomen4500075000 <= 0 en 0 = anders
55		D	<i>LNBrutoinkomen75000123000 NUL</i>	1 = LNBrutoinkomen75000123000 <= 0 en 0 = anders
56		D	<i>LNFlatsNUL</i>	1 = LNFlats <= 0 en 0 = anders
57		D	<i>LNGeenbaanNUL</i>	1 = LNGeenbaan <= 0 en 0 = anders
58		D	<i>LVGemiddeldgezinsinkomenNUL</i>	1 = LVGemiddeldgezinsinkomen <= 0 en 0 = anders
59		D	<i>LVGetrouwdNUL</i>	1 = LVGetrouwd <= 0 en 0 = anders
60		D	<i>LVNHHSmetkinderenNUL</i>	1 = LVNHHSmetkinderen <= 0 en 0 = anders
61		D	<i>LVHoekwoningenNUL</i>	1 = LVHoekwoningen <= 0 en 0 = anders
62		D	<i>LVHogeopleidingNUL</i>	1 = LVHogeopleiding <= 0 en 0 = anders
63	D	<i>LVHuurwoningenNUL</i>	1 = LVHuurwoningen <= 0 en 0 = anders	
64	D	<i>LVKoopwoningenNUL</i>	1 = LVKoopwoningen <= 0 en 0 = anders	
65	D	<i>LVNlageopleidingNUL</i>	1 = LVNlageopleiding <= 0 en 0 = anders	
66	D	<i>LVNLeeftijd024 NUL</i>	1 = LVNLeeftijd024 <= 0 en 0 = anders	
67	D	<i>LVNLeeftijd2539 NUL</i>	1 = LVNLeeftijd2539 <= 0 en 0 = anders	
68	D	<i>LVNLeeftijd4064 NUL</i>	1 = LVNLeeftijd4064 <= 0 en 0 = anders	
69	D	<i>LVNLeeftijd65 NUL</i>	1 = LVNLeeftijd65 <= 0 en 0 = anders	
70	D	<i>LVNMiddelbareopleidingNUL</i>	1 = LVNMiddelbareopleiding <= 0 en 0 = anders	
71	D	<i>LVNOverigeburgelijkestaatNUL</i>	1 = LVNOverigeburgelijkestaat <= 0 en 0 = anders	
72	D	<i>LVNOverigewoningtypenNUL</i>	1 = LVNOverigewoningtypen <= 0 en 0 = anders	
73	D	<i>LVNSamenwonendNUL</i>	1 = LVNSamenwonend <= 0 en 0 = anders	
74	D	<i>LVNSocialeklasseA NUL</i>	1 = LVNSocialeklasseA <= 0 en 0 = anders	
75	D	<i>LVNSocialeklasseB1 NUL</i>	1 = LVNSocialeklasseB1 <= 0 en 0 = anders	
76	D	<i>LVNSocialeklasseB2 NUL</i>	1 = LVNSocialeklasseB2 <= 0 en 0 = anders	
77	D	<i>LVNSocialeklasseC NUL</i>	1 = LVNSocialeklasseC <= 0 en 0 = anders	
78	D	<i>LVNSocialeklasseD NUL</i>	1 = LVNSocialeklasseD <= 0 en 0 = anders	
79	D	<i>LVNVrijstaandhuizenNUL</i>	1 = LVNVrijstaandhuizen <= 0 en 0 = anders	

¹ Zie ook bijlage 7.1

Nr	Indicatoren verzameling	D / R	Variabelen in SPSS	Omschrijving en eventuele codering
80	Postcode2-gebieden	D	<i>PC2_dummy10</i>	1 = eerste twee cijfers v/d postcode zijn 10 en 0 = anders
81		D	<i>PC2_dummy11</i>	1 = eerste twee cijfers v/d postcode zijn 11 en 0 = anders
82		D	<i>PC2_dummy12</i>	1 = eerste twee cijfers v/d postcode zijn 12 en 0 = anders
83		D	<i>PC2_dummy13</i>	1 = eerste twee cijfers v/d postcode zijn 13 en 0 = anders
84		D	<i>PC2_dummy14</i>	1 = eerste twee cijfers v/d postcode zijn 14 en 0 = anders
85		D	<i>PC2_dummy15</i>	1 = eerste twee cijfers v/d postcode zijn 15 en 0 = anders
86		D	<i>PC2_dummy16</i>	1 = eerste twee cijfers v/d postcode zijn 16 en 0 = anders
87		D	<i>PC2_dummy17</i>	1 = eerste twee cijfers v/d postcode zijn 17 en 0 = anders
88		D	<i>PC2_dummy18</i>	1 = eerste twee cijfers v/d postcode zijn 18 en 0 = anders
89		D	<i>PC2_dummy19</i>	1 = eerste twee cijfers v/d postcode zijn 19 en 0 = anders
90	Residuen-classificatie Postcode4-gebieden	D	<i>PC2_dummy2</i>	1 = postcode behoort tot klasse 2 en 0 = anders
91		D	<i>PC2_dummy3</i>	1 = postcode behoort tot klasse 3 en 0 = anders
92		D	<i>PC2_dummy4</i>	1 = postcode behoort tot klasse 4 en 0 = anders
93		D	<i>PC2_dummy5</i>	1 = postcode behoort tot klasse 5 en 0 = anders
94		D	<i>PC2_dummy6</i>	1 = postcode behoort tot klasse 6 en 0 = anders
95		D	<i>PC2_dummy7</i>	1 = postcode behoort tot klasse 7 en 0 = anders
96		D	<i>PC2_dummy8</i>	1 = postcode behoort tot klasse 8 en 0 = anders
97		D	<i>PC2_dummy9</i>	1 = postcode behoort tot klasse 9 en 0 = anders
98		D	<i>PC2_dummy10</i>	1 = postcode behoort tot klasse 10 en 0 = anders
99		D	<i>PC2_dummy11</i>	1 = postcode behoort tot klasse 11 en 0 = anders
100		D	<i>PC2_dummy12</i>	1 = postcode behoort tot klasse 12 en 0 = anders
101		D	<i>PC2_dummy13</i>	1 = postcode behoort tot klasse 13 en 0 = anders

Tabel 4.4.2b Codeboek van in SPSS gebruikte variabelen.

4.5 Regressie-analyse

In de probleembeschrijving wordt aangegeven dat er behoefte is aan een onderzoek naar de invloed van sociale indicatoren in combinatie met andere indicatoren op de verkoopwaarde van de woning. Hierbij dienen regressiemodellen te worden ontwikkeld die de verkoopwaarden van woningen zo nauwkeurig mogelijk kunnen schatten. In de literatuur zijn twee verschillende regressiemodellen gevonden die van toepassing kunnen zijn bij het schatten van verkoopwaarden van woningen. Het gaat om tijdreeks-modellen (Time-serie-models) ^[14] ^[15] en (logaritmische getransformeerde) lineaire meervoudige regressiemodellen ^[5]. Verder zijn in de literatuur indexeringsmodellen gevonden ^[10] ^[11].

Bij indexeringsmodellen speelt de factor tijd een belangrijke rol. Met deze modellen kun je alleen interpoleren en niet extrapoleren. Met andere woorden je kan met deze modellen alleen beschrijven hoe de grafiek zich in het verleden heeft gedragen en niet hoe deze zich in de toekomst zal gedragen. De volgende indexeringsmodellen zijn in de literatuur gevonden:

- *Hedonic Regression*: dit model maakt gebruik van allerlei toestand- en omgevingskenmerken van de verkochte panden;
- *Repeat Sales Regression*: dit model vergelijkt herhaalde verkopen van hetzelfde pand om tot schattingen van prijsindexcijfers te komen;

- *Assessed Value method*: dit model vergelijkt werkelijke verkoopprijzen van panden met geschatte waarden van deze panden;
- *Sales Assessment Ratio method*: in dit model worden de benodigde indexcijfers geschat uit de periodieke (maandelijkse) procentuele veranderingen van de gemiddelde verhouding tussen de verkoopprijs en de geschatte waarde.

Hedonic Regression is interessant voor het onderzoek in de zin dat dit model variabelen gebruikt uit indicatorenverzamelingen als kenmerken van de woning en sociale indicatoren etc.^[11]. Een groot aantal van de gebruikte sociale indicatoren in dit onderzoek zijn bij dit model in de literatuur gevonden.

Met een tijdreeks-model (Time-serie-model) kunnen gemiddelde verkoopwaarden van woningen op bepaalde tijdstippen geschat worden. Het voordeel van dit model is dat de factor tijd nauwkeurig wordt gemodelleerd. Een nadeel is dat dit model te complex wordt wanneer rekening wordt gehouden met verschillende factoren die tegelijkertijd invloed uitoefenen op de verkoopwaarden van de woning. Een tijdreeks-model is dus niet aan te raden als alternatief voor het huidige kenmerkenmodel, maar mogelijk kan het wel gebruikt worden als vijfde submodel bij modelmatig waardebepaling. Dit submodel kan de gemiddelde verkoopwaarde van verschillende woningtypen op verschillende lokaliteitsniveaus (postcodegebieden) schatten met behulp van de factor tijd. Dit submodel kan dan voor een corrigerende werking zorgen bij het schatten van de verkoopwaarden van woningen in het eindmodel. In dit onderzoek zullen tijdreeks-modellen niet worden toegepast.

Wat wel wordt toegepast is het (logaritmische getransformeerde) meervoudige lineaire regressiemodel. In dit model kunnen verschillende soorten indicatoren tegelijk opgenomen worden om de verkoopwaarde van de woning te schatten. De volgende indicatoren worden in de literatuur bij dit model gebruikt: kenmerken van de woning, sociale indicatoren, economische factoren etc. Dit maakt het model heel breed in haar toepassing en dat kan alleen maar nuttig zijn bij het zo nauwkeurig mogelijk schatten van de woningwaarde. Dit is ook het model dat het beste aansluit bij het kenmerkenmodel van modelmatig waardebepaling bij NBWO. Logaritmische getransformeerde lineaire meervoudige regressiemodel is het meest gebruikte model in de literatuur. In de praktijk blijkt dat lineair meervoudige regressiemodel met logaritmische transformatie beter fit dan zonder logaritmische transformatie bij het schatten van de verkoopwaarde van een woning aan de hand van onder andere de indicatoren: kenmerken van de woning en sociale indicatoren^[10].

In dit onderzoek zullen twee soorten logaritmische getransformeerde lineaire meervoudige regressiemodel worden gebruikt: een model met en een model zonder interacties tussen bepaalde variabelen uit verschillende soorten indicatorenverzamelingen. Deze twee modellen worden in de subparagrafen 4.5.1 respectievelijk 4.5.2 besproken. In subparagraaf 4.5.3 wordt uitgelegd hoe de parameters van deze modellen in regressieanalyse geschat worden. Subparagraaf 4.5.4 behandelt de gebruikte maten voor het meten van de kwaliteit van het regressiemodel. Uitleg over de analyse van de residuen van schattingen van verkoopwaarden met de regressiemodellen staat in subparagraaf 4.5.5. Tenslotte wordt in subparagraaf 4.5.6 verteld volgens welke methoden de variabelen zijn toegevoegd bij regressieanalyse in SPSS.

4.5.1 Het regressiemodel 1

Het regressiemodel geeft de veronderstellingen weer ten aanzien van de relatie tussen de te verklaren variabele, $\ln Y$, en verschillende verklarende variabelen,

$\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$ in de populatie. Deze relatie wordt voor ieder object n in de populatie beschreven met een logaritmische getransformeerde lineaire meervoudige regressiemodel:

Het model ziet er als volgt uit:

$$\ln Y_n = \mu + \sum_{i=1}^I \alpha_i \ln(x_{in}) + \sum_{j=1}^J \beta_j d_{jn} + \varepsilon_n \quad (1)$$

Y_n : n-de observatie van de te verklaren afhankelijke variabele uit vector Y ;

μ : constante;

α_i : i-de coëfficiënt uit vector α ;

x_{in} : n-de observatie van de i-de ratiovariabele uit matrix x ;

β_j : j-de coëfficiënt uit vector β ;

d_{jn} : n-de observatie van de j-de dummyvariabele uit matrix d ;

ε_n : n-de storingsterm uit vector ε , $\varepsilon_n \sim N(0, \sigma^2)$;

waarbij $i=1, \dots, I; j=1, \dots, J; n=1, \dots, N$.

De ratiovariabelen (x_1, x_2, \dots, x_I) zijn in model (1) logaritmisch getransformeerd doordat de natuurlijke logaritme (\ln) genomen wordt van de waarden van deze variabelen. De dummyvariabelen (d_1, d_2, \dots, d_J) kunnen alleen de waarden 0 of 1 aannemen, dus is het onmogelijk dummyvariabelen logaritmisch te transformeren.

Als Y oplost wordt uit model (1), wordt het volgende regressiemodel verkregen::

$$Y_n = \prod_{i=1}^I (x_{in}^{\alpha_i}) e^{\mu + \sum_{j=1}^J \beta_j d_{jn} + \varepsilon_n} \quad (2)$$

Model (1) valt uiteen in twee delen: het deterministische of verwachte gedeelte,

$\mu + \sum_{i=1}^I \alpha_i \ln(x_{in}) + \sum_{j=1}^J \beta_j d_{jn}$ en het stochastische gedeelte, de storingsterm ε_n . Beide delen worden hieronder toegelicht.

Het deterministische gedeelte

Het deterministische gedeelte van het model geeft een relatie weer tussen de afhankelijke variabelen $\ln Y$ en de verklarende variabelen $\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$, die een directe invloed hebben op $\ln Y$. In deze relatie is μ de constante term, dat wil zeggen de waarde van het deterministische deel van $\ln Y$ als alle verklarende variabelen $\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$ gelijk zijn aan nul. De coëfficiënten, $\alpha_1, \alpha_2, \dots, \alpha_I, \beta_1, \beta_2, \dots, \beta_J$ zijn de directe (ook wel partiele) invloeden van verklarende variabelen $\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$ op $\ln Y$.

Met behulp van regressie-analyse wordt de invloed van afzonderlijke verklarende variabelen op de afhankelijke variabele $\ln Y$ maximaal opgeschoond voor de versturende invloeden van alle andere verklarende variabelen in het model. Dit heeft onder andere als consequentie dat men de invloed van $\ln x_1$ op $\ln Y$ pas goed kan meten, wanneer alle andere variabelen die $\ln Y$ beïnvloeden, $\ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$ in het regressiemodel zijn gespecificeerd. Dan krijgt men een zo zuiver mogelijke kijk op de werkelijke invloed van $\ln x_1$ op $\ln Y$.

Het stochastische gedeelte

Het deterministische gedeelte is in werkelijkheid niet in staat om de te verklaren variabele $\ln Y$ exact te beschrijven. Er treden altijd kleine afwijkingen op. Deze storingen of afwijkingen hebben drie mogelijke oorzaken^[13]:

1. *Niet waarneembare of waargenomen variabelen*; niet waarneembare variabelen veroorzaken zuivere storingen, dat wil zeggen toevallige afwijkingen die niet verklaard kunnen worden. Men kan hier denken aan toevallige gebeurtenissen die naast de andere verklarende variabelen invloed hebben op de te verklaren afhankelijke variabele. Een voorbeeld van niet waargenomen variabelen is het ten onrechte niet opnemen van variabelen in het model, bijvoorbeeld omdat hiervoor geen gegevens verzameld zijn.
2. *Specificatiefouten*; In het regressiemodel wordt een lineaire relatie tussen de afhankelijke variabele en de onafhankelijke variabelen verondersteld. Wanneer deze veronderstelling niet geheel met de werkelijkheid overeenkomt, treden er als gevolg allerlei afwijkingen van de veronderstelde lineaire relatie op.
3. *Meetfouten*; In veel praktische onderzoekssituaties worden meetfouten gemaakt. Deze meetfouten kunnen verschillende oorzaken hebben, zoals onnauwkeurige meetapparatuur, fouten bij het invoeren van de waarnemingsuitkomsten in een databestand of domweg persoonlijke waarnemingsfouten van de onderzoeker.

De storingsterm zou men dus kunnen zien als een vergaarbak van allerlei soorten afwijkingen met een stochastisch karakter: Het is niet op voorhand aan te geven hoe groot de afwijking is en of deze positief of negatief zal uitvallen. Aangezien dit feit is het met het oog op het schatten en toetsen nodig een aantal veronderstellingen te maken over de verdeling van mogelijke uitkomsten van deze storingsterm. Deze veronderstellingen worden ook wel aangeduid als de basisveronderstellingen van het lineaire meervoudige regressiemodel^[13].

1. De verwachting van de residu is voor alle observaties gelijk aan nul:
 $E\varepsilon_n = 0, n = 1, \dots, N$.
2. De variantie van de residu is voor alle observaties gelijk aan σ^2 (homoskedasticiteit):
 $E\varepsilon_n^2 = \sigma^2, n = 1, \dots, N$.
3. Er is geen samenhang tussen de residuen onderling (onafhankelijkheid):
 $E\varepsilon_n \varepsilon_p = 0, n, p = 1, \dots, N, n \neq p$.
4. Er is geen samenhang tussen de residuen en de verklarende variabelen $\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$:
 $E\varepsilon_n \ln x_{in} = 0, E\varepsilon_n \ln d_{jn} = 0, n = 1, \dots, N, i = 1, \dots, I$.
5. De verdeling van mogelijke uitkomsten van de residuen is voor elk element i normaal: ε_n is normaal verdeeld.
6. De verklarende variabelen $\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$ mogen niet onderling lineair afhankelijk zijn. Wanneer deze eis *bijna* wordt geschonden ontstaat er

een zogenaamd multicollineariteitsprobleem. Dit betekent dan dat deze verklarende variabelen vrijwel dezelfde variatie van de te verklaren variabele ‘verklaren’. Er is geen absolute regel wanneer multicollineariteit 'ernstig' wordt. Een oplossing is om één van de twee onderling correlerende onafhankelijke variabelen uit de analyse te verwijderen bij een in de praktijk gebruikte aanname: $|r| \geq 0,9$ ^[19].

4.5.2 Het regressiemodel 2

Het regressiemodel geeft de veronderstellingen weer ten aanzien van de relatie tussen de te verklaren variabele, $\ln Y$, en verschillende verklarende variabelen,

$\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$ in de populatie. Deze relatie wordt voor ieder object n in de populatie beschreven met een logaritmische getransformeerde lineaire meervoudige regressiemodel met interacties:

Het model ziet er als volgt uit:

$$\ln Y_n = \mu + \sum_{i=1}^I \alpha_i \ln(x_{in}) + \sum_{j=1}^J \beta_j d_{jn} + \sum_{i=1}^I \sum_{j=1}^J \gamma_{ij} \ln(x_{in}) d_{jn} + \varepsilon_n \quad (3)$$

Y_n : n-de observatie van de te verklaren afhankelijke variabele uit vector Y ;

μ : constante;

α_i : i-de coëfficiënt uit de vector α ;

x_{in} : n-de observatie van de i-de ratiovariabele uit matrix x ;

β_j : j-de coëfficiënt uit de vector β ;

d_{jn} : n-de observatie van de j-de dummyvariabele uit matrix d ;

γ_{ij} : i-j-de coëfficiënt uit de matrix γ

ε_n : n-de storingsterm uit vector ε , $\varepsilon_n \sim N(0, \sigma^2)$;

waarbij $i=1, \dots, I$; $j=1, \dots, J$; $n=1, \dots, N$.

De ratiovariabelen (x_1, x_2, \dots, x_I) zijn in model (1) logaritmisch getransformeerd doordat de natuurlijke logaritme (\ln) van de waarden van deze variabelen wordt genomen. De dummyvariabelen (d_1, d_2, \dots, d_J) kunnen alleen de waarden 0 of 1 aannemen, dus is het onmogelijk dummyvariabelen logaritmisch te transformeren.

Als Y oplost wordt uit model (3), wordt het volgende regressiemodel verkregen:

$$Y_n = \prod_{i=1}^I x_{in}^{\alpha_i} \prod_{i=1}^I \prod_{j=1}^J (x_{in}^{\gamma_{ij} d_{jn}}) e^{(\mu + \sum_{j=1}^J \beta_j d_{jn} + \varepsilon_n)} \quad (4)$$

In het model is ook te zien dat er rekening gehouden wordt met invloeden van interacties tussen verklarende variabelen uit twee verschillende indicatorenverzamelingen, vector X en

vector V . Een interactie-effect van variabele x_1 met variabele d_1 wil zeggen dat het effect van variabele x_1 beïnvloed wordt door variabele d_1 ^[12].

Om een voorbeeld te geven laat de variabele x_1 perceelgrootte van een woning zijn en d_1 is een dummyvariabele voor de nominale variabele Indmonument (1=monument en 0= geen monument). Dan geldt dat wanneer de perceelgrootte toeneemt de verkoopwaarden van monumenten relatief sneller groeien dan de verkoopwaarden van niet monumenten. Dit betekent dat er interactie is tussen variabele perceelgrootte van de woning en de variabele Indmonument, omdat anders de groei van de verkoopwaarden evenredig zou moeten zijn.

Model (3) valt uiteen in twee delen: het deterministische of verwachte gedeelte,

$$\mu + \sum_{i=1}^I \alpha_i \ln(x_{in}) + \sum_{j=1}^J \beta_j d_{jn} + \sum_{i=1}^I \sum_{j=1}^J \gamma_{ij} \ln(x_{in}) d_{jn}$$
 en het stochastische gedeelte, de storingsterm ε_n . Beide delen worden hieronder toegelicht.

Het deterministische gedeelte

Het deterministische gedeelte van het model geeft een relatie weer tussen de afhankelijke variabelen $\ln Y$ en de verklarende variabelen $\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$, die een directe invloed hebben op $\ln Y$. In de deze relatie is μ de constante term, dat wil zeggen de waarde van het deterministische deel van $\ln Y$ als alle verklarende variabelen $\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$ gelijk zijn aan nul. De coëfficiënten, $\alpha_1, \alpha_2, \dots, \alpha_I, \beta_1, \beta_2, \dots, \beta_J, \gamma_{11}, \gamma_{22}, \dots, \gamma_{IJ}$ zijn de directe (ook wel partiële) invloeden van verklarende variabelen $\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$ op $\ln Y$.

Met behulp van regressie-analyse wordt de invloed van afzonderlijke verklarende variabelen op de afhankelijke variabele $\ln Y$ maximaal opgeschoond voor de versturende invloeden van alle andere verklarende variabelen in het model. Dit heeft onder andere als consequentie dat men de invloed van een $\ln x_1$ op $\ln Y$ pas goed kan meten, wanneer alle andere variabelen die $\ln Y$ beïnvloeden, $\ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J$ in het regressiemodel zijn gespecificeerd. Dan krijgt men een zo zuiver mogelijke kijk op de werkelijke invloed van $\ln x_1$ op $\ln Y$.

Het stochastische gedeelte

Ook hier geldt dat het deterministische gedeelte in werkelijkheid niet in staat is om de te verklaren variabele $\ln Y$ exact te beschrijven. De storingen en afwijkingen die optreden hebben dezelfde redenen als bij regressiemodel 1.

De veronderstellingen bij dit regressiemodel zijn:

1. De verwachting van de residu is voor alle observaties gelijk aan nul:
 $E\varepsilon_n = 0, n = 1, \dots, N$.
2. De variantie van de residu is voor alle observaties gelijk aan σ^2 (homoskedasticiteit):
 $E\varepsilon_n^2 = \sigma^2, n = 1, \dots, N$.

3. Er is geen samenhang tussen de residuen onderling (onafhankelijkheid):
 $E\varepsilon_n \varepsilon_p = 0, n, p = 1, \dots, N, n \neq p.$
4. Er is geen samenhang tussen de residuen en de verklarende variabelen
 $\ln x_1, \ln x_2, \dots, \ln x_I, \ln d_1, \ln d_2, \dots, \ln d_J :$
 $E\varepsilon_n \ln x_{in} = 0, E\varepsilon_n \ln d_{jn} = 0, n = 1, \dots, N, i = 1, \dots, I, j = 1, \dots, J.$
5. De verdeling van mogelijke uitkomsten van de residuen is voor elk element i normaal: ε_n is normaal verdeeld.
6. De verklarende variabelen $\ln x_1, \ln x_2, \dots, \ln x_I, \ln v_1, \ln v_2, \dots, \ln v_J$ mogen niet onderling lineair afhankelijk zijn. Wanneer dit wel het geval is, ontstaat er een zogenaamd multicollineariteit-probleem. Dit betekent dan dat deze verklarende variabelen vrijwel dezelfde variatie van de te verklaren variabele ‘verklaren’. Om dit probleem te voorkomen wordt één van de twee onderling correlerende onafhankelijke variabelen uit de analyse te verwijderd bij de aanname: $|r| \geq 0,9$ ^[19].

4.5.3 Schatten van de parameters

De Kleinste-kwadraten-methode is de methode die bij regressieanalyse wordt gebruikt om voor zowel regressiemodel 1 als regressiemodel 2 parameterschattingen te krijgen. In deze subparagraaf zal uitgelegd worden hoe de parameters van regressiemodel 1 worden geschat. De parameterschatting van regressiemodel 2 gaat op dezelfde manier. De gedachte van de Kleinste kwadraten methode is dat de parameterschattingen $\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_I, \hat{\beta}_1, \dots, \hat{\beta}_J$ en $\hat{\sigma}^2$, en daarmee de ligging van de regressielijn, zodanig worden geconstrueerd dat de gemiddelde afstand van de waarnemingen tot die lijn zo klein mogelijk is ^[20]. De regressielijn wordt dan ook wel aangeduid als de “best passende rechte lijn door de puntenwolk”. De spreiding van de residuen is dan minimaal. Meer concreet worden $\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_I, \hat{\beta}_1, \dots, \hat{\beta}_J$ en $\hat{\sigma}^2$ gevonden uit de volgende minimalisatie-procedure:

$$\min_{\substack{\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_I, \\ \hat{\beta}_1, \dots, \hat{\beta}_J}} Q = \sum_{n=1}^N \left(\ln y_n - \left(q + \sum_{i=1}^I r_i \ln(x_{in}) + \sum_{j=1}^J s_j d_{jn} \right) \right)^2 \quad (5)$$

Hierin is Q de som van de gekwadrateerde afstanden van uitkomsten $\ln y_n$ van $\ln Y_n$ tot een vlak gedefinieerd in termen van $\hat{\alpha}_{n1}, \dots, \hat{\alpha}_{In}, \hat{\beta}_{1n}, \dots, \hat{\beta}_{Jn}$ met $q, r_1, \dots, r_I, s_1, \dots, s_J$ als algemene aanduiding voor de constante en richtingscoëfficiënten van het vlak. Q meet dus het verschil tussen de werkelijke waarde van de observaties en de verwachte waarden wanneer $q, r_1, \dots, r_I, s_1, \dots, s_J$ de echte parameters zouden zijn. Het gelijk stellen van de eerste afgeleiden van Q naar de argumenten $q, r_1, \dots, r_I, s_1, \dots, s_J$ aan nul en het oplossen van de resulterende normaalvergelijkingen, leidt tot uitdrukkingen voor $\hat{\mu}, \hat{\alpha}_1, \dots, \hat{\alpha}_I, \hat{\beta}_1, \dots, \hat{\beta}_J$ als kleinste kwadratenschatters voor de onbekende $\mu, \alpha_1, \dots, \alpha_I, \beta_1, \dots, \beta_J$.

4.5.4 Evaluatiematen van het model

De data wordt opgesplitst in een traindata en een testdata. De verhouding in aantal records tussen deze datasets is respectievelijk 67% en 33%. Op de traindata vindt regressieanalyse plaats, waarbij een regressiemodel wordt ontwikkeld. Om te controleren hoe goed dit model werkt, wordt dit model toegepast op de testdata. Het toepassen gebeurt op de verklarende variabelen in de testdata. De hierbij verkregen schattingen worden dan vergeleken met de waarden van de te verklaren variabele (respons) in de testdata.

Aan de hand van een aantal evaluatiematen wordt het regressiemodel gemeten op hoe goed het zich gedraagt op de traindata en de testdata. Als het regressiemodel goed genoeg is, zullen de waarden van de evaluatiematen bij de traindata niet veel afwijken van die bij de testdata. Een belangrijke oorzaak van dat een model niet goed genoeg is, is *Overfitting*. Er is sprake van Overfitting wanneer een model zich te veel aanpast aan de waarnemingen in de trainingsverzameling, waardoor de schattingskracht afneemt. Dit betekent dat het model in het geval van Overfitting zich niet alleen aan de structuur, maar ook aan de ruis (noise) aanpast.

Drie evaluatiematen worden gebruikt om wat te zeggen over de kwaliteit van een regressiemodel:

- A. Determinatiecoëfficiënt (R^2)
- B. Percentage objecten binnen x %
- C. Standaarddeviatie van residuen

A. Determinatiecoëfficiënt (R^2)

De determinatiecoëfficiënt, R^2 , wordt gedefinieerd als ^[8]:

$$R^2 = \frac{SST - SSE}{SST} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2 - \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (6)$$

waarbij

SST : de totale kwadraatsom van de afhankelijke variabele;

SSE : de kwadraatsom van het verschil van de werkelijk gemeten waarden Y_i en de aan de hand van het model voorspelde waarden \hat{Y}_i ;

n : het aantal observaties

\bar{Y} : het gemiddelde van de geobserveerde data Y_i

SST geeft de fout weer wanneer het gemiddelde van de observaties als voorspelling van de waarnemingen genomen wordt. SSE geeft de fout weer indien voor de voorspelling van de waarnemingen het regressiemodel gebruikt wordt. R^2 geeft dus het percentage van de reductie van de fouten weer indien, in plaats van het gemiddelde van de waarnemingen, het regressiemodel gebruikt wordt. De waarde van R^2 ligt steeds tussen 0 en 1. Hoe dichterbij 1 ligt, hoe beter de werkelijke waarden van de afhankelijke variabele benaderd worden door het model. Als R^2 gelijk is aan 0, dan wil dit zeggen dat het model geen enkele toegevoegde waarde heeft.

B. Percentage objecten binnen x %

Het percentage woningen waarvoor de schatting van hun verkoopwaarde, op basis van het regressiemodel, minder dan x % van de gerealiseerde verkoopwaarde afwijkt; voor x worden de waarden 5, 10, 15 of 20 genomen. Bij deze maat kan gezegd worden hoe groter het aantal voorspelde objecten dat binnen 5%, 10%, 15% of 20% ligt, des te beter het model is.

C. Standaarddeviatie van residuen

De standaarddeviatie van de residuen is een maat voor de afwijking van alle waarnemingen ten opzichte van het rekenkundige gemiddelde. Hoe kleiner de standaarddeviatie, des nauwkeuriger de voorspelling. Verder geldt uit de statistiek dat als een verdeling van waarnemingen enigszins lijkt op een zogenoemde normale verdeling dan gelden de volgende vuistregels: 68 %, 95% en 99,9% van de woningen hebben residuen die respectievelijk niet meer dan een, twee en drie maal de standaarddeviatie(s) naar links en rechts afwijkt of afwijken van het gemiddelde^[10].

4.5.5 Analyse residuen

Door de residuen te analyseren kunnen de veronderstellingen van regressieanalyse met betrekking tot de lineariteit van het model en ten aanzien van de residuen (een normale verdeling; een gemiddelde 0 en een constante variantie σ^2) worden gecontroleerd. Ook kunnen de cases met de grootste residuen, en dus met de grootste afwijkingen van de door het model voorspelde Y-waarde, worden opgespoord. Deze zogenaamde extreme waarden zijn van invloed op de uiteindelijke regressievergelijking. Het is meestal verstandig om extreme waarden buiten de analyse te laten.

Controle op lineariteit en homoscedasticiteit

Om de lineariteit en homoscedasticiteit van het model te controleren wordt een grafiek gemaakt van de gestandaardiseerde residuen op de Y-as tegen de gestandaardiseerde voorspelde waarden op de X-as. Als aan de aanname van lineariteit en homoscedasticiteit is voldaan, zullen de punten in deze plot volstrekt willekeurig verdeeld zijn. De aanwezigheid van een bepaald patroon is een aanwijzing voor niet-lineaire regressie, voor heteroscedasticiteit of voor beide.

Controle op normaliteit

In het histogram van de gestandaardiseerde residuen wordt de theoretische normale verdeling ingetekend. Sterke afwijkingen van de geobserveerde frequenties (kolommen) en de theoretische normale verdeling zijn een aanwijzing voor non-normaliteit van de errortermen, misspecificatie van het model (b.v. er zijn relevante onafhankelijke variabelen vergeten) of heteroscedasticiteit.

Localiseren van extreme waarden

Hoewel dit niets te maken heeft met de schending van de assumpties, kan voor de interpretatie van een regressie-analyse het opsporen van extreem grote residuen van belang zijn. Het gaat daarbij immers om slecht voorspelbare waarden op de afhankelijke variabele. Een extreme waarde wordt gedefinieerd als een waarde met een gestandaardiseerde residu die meer dan drie standaarddeviaties afwijkt van nul.

4.5.6 Directe en stapsgewijze regressie

Bij meervoudige regressieanalyse kunnen de onafhankelijke variabelen allemaal tegelijk of stapsgewijs ingevoerd worden in het regressiemodel. Bij de stapsgewijze methoden veranderen de regressieparameters bij iedere stap: de uitvoer is dan veel uitgebreider. Bij de stapsgewijze methoden worden alleen variabelen opgenomen, die een statistisch significante rol spelen in het regressiemodel. De variabelen worden toegevoegd op volgorde van hun relatieve invloed op de afhankelijke variabele. Het model is voltooid, wanneer de significanties van alle nog niet opgenomen variabelen groter zijn dan 0,05. De grenswaarde 0,05 wordt PIN (Probability of F) genoemd en komt overeen met F-waarde of FIN (F-to-enter) van 3,84.

Binnen dit onderzoek worden in SPSS drie verschillende methoden gebruikt voor het invoeren van variabelen in de regressievergelijking^[19]:

1. ENTER: de standaardmethode, waarin alle onafhankelijke variabelen tegelijk worden ingevoerd. Hierbij kunnen dus zowel significante als niet significante variabelen worden opgenomen.
2. FORWARD-methode: Bij deze methode worden de onafhankelijke variabelen een voor een in het model opgenomen op basis van significantie. Als eerste wordt de variabele met de laagste PIN opgenomen. Dit is de variabele met de hoogste enkelvoudige correlatie met de afhankelijke variabelen. Bij de volgende stappen wordt steeds de variabele met de laagste PIN toegevoegd, onder constant houding van de invloed van de variabelen die al zijn opgenomen. Als er geen significanties meer voorkomen die aan het PIN-criterium ($\leq 0,05$) voldoen, is het model voltooid.
3. STEPWISE-methode: deze methode is een combinatie van Forward en Backward. Eveneens bij Forward wordt eerst de meest significante variabele opgenomen op basis van de PIN. Bij elke volgende stap kan een reeds opgenomen variabele weer worden verwijderd (Backward), indien door toevoeging van een nieuwe variabele de significantie van de opgenomen variabele niet meer aan het eliminatiecriterium, een maximale significantie van 0,10 (POUT) voldoet. Deze komt overeen met een F-waarde of FOUT (F-to-remove) van 2,71. Om te voorkomen dat een variabele steeds opnieuw wordt opgenomen en weer wordt verwijderd, dient de PIN altijd kleiner te zijn dan de POUT.

De procedure die doorlopen wordt bij het opstellen van regressiemodellen in dit onderzoek gaat als volgt:

Stap 0:

Gebruik Enter-methode om te kijken wat de verklaarde variantie is van de afhankelijke variabele door alle variabelen tegelijk. Als het hier om veel te onderzoeken variabelen gaat, ga naar stap 2, anders ga naar stap 1.

Stap 1:

Gebruik Stepwise-methode; kies hier het model die de maximale verklaarbaarheid van de variantie heeft met zo min mogelijk verklarende variabelen. Ga naar stap 3

Stap 2:

Gebruik tegelijkertijd Enter-methode met significante variabelen uit stap 0 inclusief hoofdvariabelen (variabelen die je bij voorkeur erbij wilt hebben) en Forward-methode met andere niet significante variabelen uit stap 0. Vervolgens vindt er een uitwisseling van variabelen plaats tussen Enter en Forward: significante variabelen gaan van Forward naar Enter en niet-significante variabelen gaan van Enter naar Forward. De niet-significante variabelen bij Forward gaan uit het proces en de significante variabelen bij Enter blijven hier.

Dit proces gaat net zolang door totdat er geen variabelen zijn bij Forward. De variabelen die bij Enter overblijven behoren tot het beste model. Ga naar stap 1.²

Stap 3:

Gebruik Enter-methode; je krijgt hier een model die zoveel mogelijk variantie verklaard met zo min mogelijke variabelen. De verklaarde variantie zal nauwelijks afwijken van de totale verklaarde variantie bij stap 0. STOP

² Stap 2 is nodig om het grote aantal variabelen op een *snelle* manier te reduceren tot een kleiner aantal, waarbij dit niet ten koste gaat van de nauwkeurigheid waarmee voorspeld kan worden.

5. Resultaten onderzoek

“Vertrouwelijk”

5.1 Beschrijvende statistiek

“Vertrouwelijk”

5.2 Regressieanalyse

“Vertrouwelijk”

5.3 Beste regressiemodellen

“Vertrouwelijk”

5.3.1 Regressiemodel 6A

“Vertrouwelijk”

5.3.2 Regressiemodel 6B

“Vertrouwelijk”

5.4 Analyse van residuen

“Vertrouwelijk”

5.4.1 Analyse van residuen regressiemodel 6A

“Vertrouwelijk”

5.4.2 Analyse van residuen regressiemodel 6B

“Vertrouwelijk”

6 Vergelijking alternatieve en huidige modellen

“Vertrouwelijk”

7. Conclusies & aanbevelingen

“Vertrouwelijk”

.

8. Bijlagen

“Vertrouwelijk”

8.1 Bijlage: Classificatie sociale klassen

“Vertrouwelijk”

8.2 Bijlage: Output SPSS regressiemodel 6A

“Vertrouwelijk”

8.3 Bijlage: Output SPSS regressiemodel 6B

“Vertrouwelijk”

8.4 Bijlage: Grafieken residuenanalyse regressiemodel 6A

“Vertrouwelijk”

8.5 Bijlage: Grafieken residuenanalyse regressiemodel 6B

“Vertrouwelijk”

9. Literaturopgave

- [1] Dr. A. Buijs, *Statistiek om mee te werken*, zesde druk, 1997, Educatieve Partners Nederland BV, Houten
- [2] J.A. van Buuren, *Methoden en technieken van onderzoek 1*, derde druk, 1997, Open Universiteit Heerlen
- [3] John M. Clapp, *The influence of Economic Variables on Local House Price Dynamics*, *Journal of Urban Economics* 36, 161-183 (1994)
- [4] Fred A. von Dewart, Dimitry J.C. Fleming en Fred W.M. Pallada, *Een geïntegreerde prijsindex voor de markt van koopwoningen*
- [5] Thomas A. Downes and Jeffrey E. Zabel, *The impact of school characteristics on house prices: Chicago 1987-1991*, *Journal of Urban Economics* 52 (2002) 1-25
- [6] Edward L. Glaeser, *The social Consequences of housing*, *Journal of Housing Economics* 9, 1-23 (2000)
- [7] G. Groenewegen, *Modelmatig waarderen/Platform WOZ*, 2004, Kluwer bv, Deventer
- [8] Dr. M.C.M de Gunst, *Statistical Models* Collegedictaat, Vrije Universiteit, (Najaar 2004)
- [9] George G.M. ten Have, *Taxatieleer vastgoed 1*, derde druk, 2002, Wolters-Noordhoff bv, Groningen/Houten
- [10] Martin Hoesli, *Three New Real Estate Price Indices for Geneva, Switzerland*, *Journal of Real Finance and Economics* 15:1, 93-109 (1997)
- [11] Yan Kestens, *Land use, accessibility and profile of the households: effects on the residential choice and the value of the properties*, 2004, Laval university
<http://www.theses.ulaval.ca/2004/21646/21646.html>
- [12] prof. dr. A. van Knipperberg en dr. F.W. Siero, *Multivariate analyse*, tweede herziene druk, 1994, Bohn Stafleu Van Loghum, Houten
- [13] E. de Leede, J. van Dale, *In & Uit Statistisch onderzoek met SPSS for Windows*, 1996, Eburon bv, Delft
- [14] Michael C. Lovell, *Seasonal Adjustment of Economics Time Series and Multiple Regression Analysis*, *Journal of the American Statistical Association*, Vol. 58, No. 304 (Dec., 1963), 993-1010.
- [15] Geoffrey Meen, *The Time Series behaviour of House Prices: A Transatlantic Divide?* *Journal of Housing Economics* 11, 1-23 (2002)
- [16] James M. Poterba, *House Price Dynamics: The Role of Tax Policy and Demography*, *Brookings Papers on Economic Activity*, 2:1991.
- [17] dr. M.A.J. Theebe, *Samenvatting proefschrift "Housing Market Risks"*, 2002, Universiteit van Amsterdam

- [18] Olympia Bover en Pilar Velilla, *Hedonic house prices without characteristics: the case of new multiunit housing*, www.cepr.org/pubs/dps/DP3161.asp, (2002)
- [19] Alphons de Vocht, *Basishandboek SPSS 6.1 voor Windows*, derde druk, 1998, Bijleveld Press, Utrecht