



Methodebeschrijving Outputprijsindexcijfer van nieuwbouwwoningen

1. Inleiding

Dit is een methodebeschrijving van de statistiek Outputprijsindexcijfer van nieuwbouwwoningen (O-PINW). De beschrijving heeft alleen betrekking op de methodologische aspecten van de O-PINW. De focus zal daarom vooral gericht zijn op een beschrijving van het regressiemodel en onderdelen daarvan zoals de uitbijterselectiemethode.

Voor een beter begrip van de achtergrond van de O-PINW wordt begonnen met een algemene inleiding. Hierin wordt toegelicht op welke informatie het indexcijfer O-PINW gebaseerd is. In de volgende paragraaf wordt dieper ingegaan op het regressiemodel. De variabelen en formules worden hier verder uitgewerkt en beschreven.

2. Algemeen

De O-PINW is gebaseerd op de statistiek Verleende bouwvergunningen (Vb). Deze statistiek geeft informatie over alle door de gemeenten verstrekte bouwvergunningen voor nieuwbouwwoningen. Een bouwvergunning bestaat uit één bouwproject, waarbij een dergelijk bouwproject uit meerdere woningen kan bestaan. Dit betekent dat de waargenomen prijs betrekking heeft op het gehele project. Een prijs per woning kan worden berekend omdat het aantal woningen waaruit een project bestaat ook bekend is.

Het prijsindexcijfer wordt berekend door middel van een regressiemodel, wat een hedonische methode wordt genoemd. Ieder kwartaal wordt een gemiddelde prijs van de basiswoning geschat aan de hand van een aantal prijsbepalende variabelen. De geschatte prijs wordt vervolgens gekoppeld aan de gemiddelde geschatte prijs in het referentiejaar (zie subparagraaf 3.4.2). Op deze wijze wordt een prijsindexcijfer verkregen.

Het regressiemodel is gebaseerd op vijf verklarende variabelen, te weten :

- De inhoud per woning in kubieke meters
- Het aantal woningen in het project
- Het type opdrachtgever
- De grondsoort
- De eigendomsvorm van de woning (huur/eigen).

3. Methodologische beschrijving

In bovenstaande paragraaf is een algemene beschrijving gegeven. Nu kunnen we overgaan op een nadere beschrijving van het regressiemodel en de gebruikte variabelen. Eerst zal stapsgewijs beschreven worden welke han-



delingen worden uitgevoerd om te komen tot een indexcijfer, welke als leidraad wordt gebruikt bij de methodebeschrijving.

3.1 De berekening – stap voor stap

Voor het bepalen van de gemiddelde waarden van de variabelen in het basisjaar¹ en de daar bijhorende referentieprijzen in het referentiejaar² worden éénmalig de volgende stappen doorlopen:

- A1. Het regressiemodel wordt *ongewogen* berekend om de residuen te berekenen. Op basis hiervan worden de uitbijters bepaald.
- A2. De uitbijters worden verwijderd uit het databestand.
- A3. Op basis van het met het aantal in de bouwvergunning genoemde woningen *gewogen* nieuwe databestand worden de gemiddelde waarden van de variabelen in het basisjaar berekend.
- A4. Het regressiemodel wordt nogmaals *ongewogen* berekend met het nieuwe databestand. Hieruit volgt per variabele een regressiecoëfficiënt.
- A5. De regressiecoëfficiënten worden gekoppeld aan de gemiddelde waarden van de variabelen in het basisjaar (stap A3). Het exponent van de optelsom van deze koppelingen is de geschatte prijs voor het betreffende kwartaal.
- A6. Het rekenkundig gemiddelde van de vier geschatte kwartaalprijzen van het referentiejaar geeft de referentieprijzen.

Om tot een prijsindex te komen, worden ieder kwartaal de volgende stappen doorlopen (stappen B1 en B2 zijn gelijk aan de stappen A1 en A2 en de stappen B3 en B4 zijn gelijk aan A4 en A5):

- B1. Het regressiemodel wordt *ongewogen* berekend om de residuen te berekenen. Op basis hiervan worden de uitbijters bepaald.
- B2. De uitbijters worden verwijderd uit het databestand.
- B3. Het regressiemodel wordt nogmaals *ongewogen* berekend met het nieuwe databestand. Hieruit volgt per variabele een regressiecoëfficiënt.

¹ Het basisjaar is het jaar waarin de waarde van de variabelen van de (gemiddelde) basiswoning wordt vastgesteld; in dit geval 2007.

² Het referentiejaar is het jaar waarin het prijsindexcijfer (rekenkundig) op 100 is vastgesteld; in dit geval 2005.



B4. De regressiecoëfficiënten berekend in stap B3 worden gekoppeld aan de gemiddelde waarden van de variabelen in het basisjaar (stap A3). Het exponent van deze optelsom van deze koppelingen is de geschatte prijs voor het betreffende kwartaal.

B5. Deze geschatte kwartaalprijs wordt gedeeld door de geschatte prijs in het referentiejaar (i.e. de referentieprijs, stap A6). Vermenigvuldigd met 100 levert dit een indexcijfer op.

3.2 De variabelen

In paragraaf 2 is een globale beschrijving gegeven van de variabelen die gebruikt worden in de regressievergelijking. Hieronder wordt verder beschreven hoe de variabelen zijn opgebouwd en op welke wijze de afzonderlijke variabelen in de regressievergelijking worden opgenomen.

De afhankelijke variabele in de vergelijking is de logaritme van de bouwkosten per woning.

Er zijn vijf onafhankelijke variabelen in de vergelijking. De eerste variabele is de inhoud per woning. Deze inhoud in m^3 wordt berekend door de totale woninginhoud in een project te delen door het aantal woningen in het betreffende project. Vervolgens wordt hiervan de logaritme genomen. De projectgrootte, i.e. het aantal woningen in het project is de tweede variabele in het model. Ook deze variabele wordt omgerekend naar logaritmen. De overige drie variabelen worden omgezet in dummyvariabelen. Hiertoe worden de variabelen Type opdrachtgever, Grondsoort en Eigendomsvorm uitgesplitst naar meerdere variabelen, analoog aan de verdeling van de variabelen in categorieën. De variabele Opdrachtgever is onderverdeeld in drie categorieën, te weten: DO1 (Overheid en woningbouwcorporaties), DO2 (Bouwers voor de markt en institutionele beleggers) en DO3 (Particulieren en overige). De variabele Grondsoort is de enige variabele die niet rechtstreeks in het bestand van de Vb bekend is. Voor deze variabele is eenmalig vastgesteld welke grondsoort in welke gemeente het meest voorkomt. Vervolgens wordt deze informatie gekoppeld aan het onderdeel Gemeente in het Vb-bestand. De variabele Grondsoort is onderverdeeld in vijf categorieën, te weten: DG1 (zand, waddenzee, heuvelland en duinen), DG2 (laagveen), DG3 (rivierengebied), DG4 (zeeklei) en DG5 (getijdengebied, Zeeland en afgesloten zearmen). Van de variabele Eigendomsvorm (H/E) wordt slechts één categorie opgenomen. Bij afwezigheid van de een is namelijk vanzelf de ander aanwezig. Voor ieder van deze laatste drie variabelen wordt één van de dummies niet in de regressie opgenomen omdat deze als referentie dient. Gekozen is om de meest voorkomende categorie als referentie te gebruiken. In de regressievergelijking zijn de variabelen DO1 t/m DO2 en DG2 t/m DG5 en H/E te vinden.

Tot slot wordt er ieder kwartaal een constante toegevoegd aan de vergelijking. Deze constante is de coëfficiënt van een kwartaaldummy die de

waarde 1 aanneemt voor alle records van het desbetreffende kwartaal en de waarde 0 voor de records van de overige kwartalen. Doordat er expliciet voor ieder kwartaal een constante wordt opgenomen is het overbodig (en onmogelijk) om ook nog een algemene constante voor alle records in een jaar op te nemen. Het regressiemodel wordt daarop in de SPSS-syntax gespecificeerd als een model zonder constante term.

3.3 De uitbijterselectiemethode (stap A1, A2, B1 en B2)

Voordat de regressiecoëfficiënten worden geschat, vindt een uitbijterselectie plaats. De gebruikte methodiek is gebaseerd op de residuen van het regressiemodel, waarbij de projecten met de grootste residuen worden aangemerkt als uitbijter. Het betreft hier 5 % van alle projecten, waarvan 2,5 % met de grootste positieve residuen en 2,5 % met de grootste negatieve residuen. De residuen worden bepaald door eerst een regressieanalyse uit te voeren op het bestand met alle projecten. Na verwijdering van de uitbijterprojecten wordt het model opnieuw geschat volgens de bovengeschreven methode.

3.4 Het referentie- en basisjaar

Bij een basisverlegging worden de gemiddelde waarden van de variabelen in het basisjaar en de referentieprijzen van het referentiejaar bepaald.

3.4.1 Basiswoning in basisjaar (A3)

De gemiddelde waarden van de variabelen worden berekend op basis van het met het aantal woningen van de bouwvergunning gewogen jaarbestand zonder uitbijters. Het zijn dus gemiddelde waarden van het gehele basisjaar en vormen samen een gemiddelde basiswoning. Let wel, het gaat hier nadrukkelijk om de gemiddelde waarden van de regressievariabelen, dus het gemiddelde van de logaritme van de inhoud, van de logaritme van de projectgrootte, en van de dummyvariabelen. Zij worden ieder kwartaal gekoppeld aan de geschatte regressiecoëfficiënten. Voor ieder kwartaal is het gemiddelde voor de betreffende kwartaaldummy gelijk aan 1 en van de overige kwartaaldummies gelijk aan 0.

3.4.2 Referentieprijzen in referentiejaar (A4 t/m A6)

De referentieprijzen worden berekend als een gemiddelde van de geschatte prijzen van de vier kwartalen in het referentiejaar. Dit houdt in dat in het referentiejaar eerst de volledige berekening wordt doorlopen tot het punt waar per kwartaal de prijs is geschat. Het gemiddelde van deze vier prijzen is de referentieprijzen. In een formule ziet dit er als volgt uit :

$$(1) \hat{P}_r = (\hat{P}_{1r} + \hat{P}_{2r} + \hat{P}_{3r} + \hat{P}_{4r}) / 4$$

waarbij

\hat{P}_r is de referentieprijzen, i.e. de gemiddelde geschatte prijs in het referentiejaar

\hat{P}_{kr} is de geschatte prijs per kwartaal in het referentiejaar

Voor de uitleg van de berekening van \hat{P}_{kr} wordt verwezen naar paragraaf 3.5. Deze berekening is namelijk identiek aan de berekening van de geschatte prijs in een willekeurig kwartaal van een willekeurig jaar \hat{P}_{kj} .

3.5 De berekening (B3 tm B5)

Na het bepalen van de referentieprij van de basiswoning (stap A1 t/m A6) en het verwijderen van de uitbijters (stap B1 en B2) kan ieder kwartaal het regressiemodel worden berekend. Kort gezegd komt het erop neer dat er voor alle verklarende variabelen regressiecoëfficiënten worden bepaald (formule 2). Deze regressiecoëfficiënten worden vervolgens gekoppeld aan de gemiddelde waarden uit het basisjaar (formule 3). Dit levert een totaalcoëfficiënt op waarvan het exponent de geschatte bouwkosten van het kwartaal is. Tot slot worden deze geschatte bouwkosten vergeleken met de referentieprij om tot een index te komen (formule 4).

Omgezet in formules ziet dit er als volgt uit:

$$(2) \ln(P_{kj}) = c_{kj} + \beta_{1kj} \ln(x_{1kj}) + \beta_{2kj} \ln(x_{2kj}) + \beta_{3kj} x_{3kj} + \beta_{4kj} x_{4kj} + \beta_{5kj} x_{5kj} + \beta_{6kj} x_{6kj} + \beta_{7kj} x_{7kj} + \beta_{8kj} x_{8kj} + \beta_{9kj} x_{9kj} + \varepsilon_{kj}$$

waarbij

P_{kj} is de bouwkosten per woning voor kwartaal k in jaar j

c_{kj} is de coëfficiënt van de constante

β_{xkj} is de regressiecoëfficiënt van variabele x voor kwartaal k in jaar j

x_1 is de variabele inhoud

x_2 is de variabele projectgrootte

x_3 x_4 x_5 en x_6 zijn dummyvariabelen voor de categorieën 2, 3, 4 en 5 van de variabele Grondsoort

x_7 en x_8 zijn dummyvariabelen voor de categorieën 1 en 2 van de variabele Type opdrachtgever

x_9 is de dummyvariabele voor de categorie H/E van de variabele Eigendomsvorm

ε_{kj} is de storingsterm

Het regressiemodel schat ongewogen de coëfficiënten van de variabelen per kwartaal voor een heel jaar tegelijkertijd. Dit betekent dat voor ieder kwartaal apart regressiecoëfficiënten worden berekend op basis van de bouwvergunningen die in dat betreffende kwartaal worden uitgegeven³. De regressiecoëfficiënten worden vervolgens gekoppeld aan de gemiddel-

³ Er wordt onderscheid gemaakt tussen enerzijds meldjaar en – maand en anderzijds rekenjaar en – maand. De bouwvergunningen worden namelijk door gemeenten meestal in een andere maand gemeld bij het CBS dan in de maand waarvoor de bouwvergunning geldt. Sinds 2001 worden deze bouwvergunningen teruggelegd. Dit betekent dat de bouwvergunningen in de berekening dus niet langer worden meegenomen in het kwartaal waarbinnen de maand (meldmaand) valt dat de bouwvergunning wordt aangemeld, maar in het kwartaal waarbinnen het rekenjaar/maand geldt.

de waarden van de variabelen in het basisjaar. Omdat de logaritme van de bouwkosten is berekend, wordt de exponent genomen van het totaal van de coëfficiënten maal de basiswaarden.

$$(3) \hat{P}_{kj} = \exp \left(c_{kj} * \bar{c}_b + \hat{\beta}_{1kj} * \bar{x}_{1b} + \hat{\beta}_{2kj} * \bar{x}_{2b} + \hat{\beta}_{3kj} * \bar{x}_{3b} + \hat{\beta}_{4kj} * \bar{x}_{4b} + \hat{\beta}_{5kj} * \bar{x}_{5b} + \hat{\beta}_{6kj} * \bar{x}_{6b} + \hat{\beta}_{7kj} * \bar{x}_{7b} + \hat{\beta}_{8kj} * \bar{x}_{8b} + \hat{\beta}_{9kj} * \bar{x}_{9b} \right)$$

waarbij

\hat{P}_{kj} is de geschatte prijs voor kwartaal k in jaar j

\bar{x}_{xb} is de gewogen gemiddelde waarde per variabele x in het basisjaar b

$\hat{\beta}_{xkj}$ zijn de regressiecoëfficiënten per variabele

Om tot een indexcijfer per kwartaal te komen, wordt vervolgens de geschatte kwartaalprijs \hat{P}_{kj} gedeeld door de zogenoemde referentieprijs \bar{P}_r . Vermenigvuldigd met 100 leidt dit tot het indexcijfer.

$$(4) I_{kj} = 100 * \left(\hat{P}_{kj} / \bar{P}_r \right)$$

waarbij

\bar{P}_r is de referentieprijs, i.e. de gemiddelde geschatte prijs in het referentiejaar

I_{kj} is het prijsindexcijfer van kwartaal k voor jaar j