



***Methodenbeschrijving Prijsindexcijfer van de bouwkosten van
nieuwe woningen (PINW)***

Sanne Elfering

Samenvatting: Dit is een methodebeschrijving van de statistiek Prijsindexcijfer van de Bouwkosten van Nieuwe Woningen. De focus ligt bij de methodologische aspecten van de berekeningswijze.

Trefwoorden: procesbeschrijving, methodologie, PINW, bouwkosten



Inleiding

Dit is een methodebeschrijving van de statistiek Prijsindexcijfer van de Bouwkosten van Nieuwe Woningen (PINW). De beschrijving heeft alleen betrekking op de methodologische aspecten van de PINW. De focus zal daarom vooral gericht zijn op een beschrijving van het regressiemodel en onderdelen daarvan als de uitbijterselectiemethode.

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

Algemeen

De 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 1 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.

De PINW is een kwartaalstatistiek en is in gebruik genomen op 1 januari 2001. Het prijsindexcijfer wordt berekend door middel van een regressiemodel, wat een hedonische methode wordt genoemd. Ieder kwartaal wordt een gemiddelde prijs geschat aan de hand van een aantal prijsbepalende variabelen. De geschatte prijs wordt vervolgens gekoppeld aan de gemiddelde geschatte prijs in het basisjaar (zie paragraaf 3.3). Op deze wijze wordt een prijsindexcijfer verkregen.

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

- De inhoud per woning in kubieke meters m^3
- Het aantal woningen in een project
- Het type opdrachtgever
- De grondsoort
- Het soort woning : een huurwoning of een koopwoning (huur/eigen).

De variabele Opdrachtgever is onderverdeeld in drie categorieën, te weten ‘Overheid en woningbouwcorporaties’ (DO1), ‘Bouwers voor de markt en institutionele beleggers’ (DO2), en ‘Particulieren en overige’ (DO3). 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 : zeelei

DG5 : getijdengebied, Zeeland en afgesloten zeearmen.

Methodologische beschrijving

In bovenstaande paragraaf is een algemene beschrijving gegeven van de data waarop de PINW is gebaseerd. Nu kunnen we overgaan op een nadere beschrijving van het regressiemodel en de gebruikte variabelen. Eerst zal echter stapsgewijs besproken worden welke handelingen worden uitgevoerd om te komen tot een indexcijfer.

De berekening – stap voor stap

Om tot een prijsindex te komen, worden voor ieder kwartaal van een willekeurig jaar 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. Het regressiemodel wordt nogmaals *ongewogen* berekend met het nieuwe databestand. Hieruit volgt per variabele een regressiecoëfficiënt.
- A4. De regressiecoëfficiënten worden gekoppeld aan de gemiddelde waarden van de variabelen in het basisjaar (stap B3). Het exponent van deze optelsom is de geschatte prijs voor het betreffende kwartaal.
- A5. Deze geschatte kwartaalprijs wordt gedeeld door de geschatte prijs in het basisjaar (i.e. de basisprijs, stap B6). Vermenigvuldigd met 100 levert dit een indexcijfer op.

Voor de gemiddelde waarden van de variabelen in het basisjaar en de basisprijs worden voor de kwartalen van het basisjaar de volgende stappen doorlopen (stappen B1 en B2 zijn gelijk aan de stappen A1 en A2 en stappen B4 en B5 zijn gelijk aan de stappen A3 en A4) :

- 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. Op basis van het *gewogen* nieuwe databestand worden de gemiddelde waarden van de variabelen berekend.
- B4. Het regressiemodel wordt nogmaals *ongewogen* berekend met het nieuwe databestand. Hieruit volgt per variabele een regressiecoëfficiënt.
- B5. De regressiecoëfficiënten worden gekoppeld aan de gemiddelde waarden van de variabelen in het basisjaar (uit stap B3). Het exponent van deze optelsom is de geschatte prijs per kwartaal.



B6. Het gemiddelde van de geschatte prijzen van de vier kwartalen uit het basisjaar is de basisprijs.

De variabelen

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

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

Er zijn vijf onafhankelijke variabelen in de vergelijking. De eerste variabele is de inhoud per woning m^3 . Deze inhoud m^3 wordt berekend door de totale woninginhoud in een project te delen door het aantal woningen in het betreffende project. Vervolgens 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 derde, vierde en vijfde variabele worden omgezet in dummyvariabelen. Hiertoe worden de variabelen Type opdrachtgever en Grondsoort uitgesplitst naar meerdere variabelen, analoog aan de verdeling van de variabelen in categorieën. Voor ieder van deze twee variabelen is één van de dummies overbodig (maakt niet uit welke). In de regressievergelijking zijn daarom de variabelen DO1 t/m DO2 en DG2 t/m DG5 te vinden. De variabele huur/eigen wordt een dummyvariabele omdat voor een huurwoning een 1 wordt genoteerd en voor een eigen woning een 0.

Tot slot wordt er ieder kwartaal een constante toegevoegd aan de vergelijking. Deze constante is de coëfficiënt van een kwartaal dummy 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 daarom in de SPSS-setup gespecificeerd als een model zonder constante term.

Het basisjaar

Bij een basisverlegging worden eenmalig de basisprijs en de gemiddelde waarden van de variabelen in het basisjaar bepaald.

De gemiddelde waarden van de variabelen worden berekend op basis van het gewogen jaarbestand zonder uitbijters. Het zijn dus gemiddelde waarden van het gehele basisjaar. Let wel, het gaat hier nadrukkelijk om de gemiddelde waarden van de regressievariabelen, dus het gemiddelde van de logaritme van de inhoud m^3 , 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 van de desbetreffende kwartaal dummy gelijk aan 1 en van de overige kwartaal dummies gelijk aan 0.

De basisprijs wordt berekend als een gemiddelde van de geschatte prijzen van de vier kwartalen in het basisjaar. Dit houdt in dat in het basisjaar eerst de volledige berekening wordt doorlopen tot het punt waar per kwartaal de prijs is geschat.



Centraal Bureau voor de Statistiek

Het gemiddelde van deze vier prijzen is de basisprijs. In een formule ziet dit er als volgt uit :

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

waarbij

\hat{P}_b is de basisprijs, i.e. de gemiddelde geschatte prijs in het basisjaar

\hat{P}_{kb} is de geschatte prijs per kwartaal in het basisjaar

Voor de uitleg van de berekening van \hat{P}_{kb} wordt verwezen naar paragraaf 3.4. Deze berekening is namelijk identiek aan de berekening van de geschatte prijs in een willekeurig kwartaal van een willekeurig jaar \hat{P}_{kj} .

De berekening

Vervolgens wordt ieder kwartaal het regressiemodel berekend. Kort gezegd komt het erop neer dat er voor alle verklarende variabelen regressiecoëfficiënten worden bepaald. Deze regressiecoëfficiënten worden vervolgens gekoppeld aan de gemiddelde waarden uit het basisjaar. Dit levert een totaalcoëfficiënt op waarvan het exponent de geschatte prijs van het kwartaal is.

Uitgebreid levert dit het volgende regressiemodel op :

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

waarbij

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

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

c_{kj} is de coëfficiënt van de betreffende kwartaal dummy.

x_1 is de variabele inhoud m^3

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 variabele Huur/eigen

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 zijn uitgegeven¹. De regressiecoëfficiënten worden vervolgens gekoppeld aan de gemiddelde waarden van de variabelen in het basisjaar. Omdat de logaritme van de prijs is berekend, wordt de exponent genomen van het totaal van de coëfficiënten maal de basiswaarden.

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

waarbij

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

¹ 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 waarin de bouwvergunning is afgegeven. 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 waarin de vergunning is afgegeven (rekenjaar/maand).



$\bar{x}_{x,b}$ 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 basisprijs \hat{P}_b . Vermenigvuldigd met 100 leidt dit tot het indexcijfer.

$$(3) I_{kj} = 100 * \left(\hat{P}_{kj} / \hat{P}_b \right)$$

waarbij

\hat{P}_b is de basisprijs, i.e. de gemiddelde geschatte prijs in het basisjaar

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

De uitbijtersselectiemethode

Voordat de regressiecoëfficiënten worden geschat, vindt een uitbijtersselectie 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.