Paper

Price Index for Newly Built Dwellings

Method description

October 2018

Contents

1. Introduction

1.1 Reason for the report

Statistics Netherlands (CBS) calculates a Price Index for Newly Built Dwellings (PND) pursuant to an European regulation. In European Commission Regulation [1] stipulates that all member states must calculate a PND. This report describes how the new PND is calculated for the Netherlands as a whole in accordance with guidelines from Eurostat [1], the statistical office of the European Union.

1.2 Objective PND

The Price Index of Newly Built Dwellings (PND) measures quarterly price developments of newly constructed dwellings. The dwelling must be built on Dutch territory and must be sold to a private individual. It is also important that the property is, or will be founded on a new foundation and is added to the housing stock for the first time. Rent, renovation and transformation of homes are therefore excluded of the population. Private commissioning (also known as diy) is also excluded. The PND is part of the OOH (Owner Occupied Housing) and in the guidelines of Eurostat for OOH, a distinction is made between newly built dwellings and self-built dwellings. The ground is usually not included in the price for the property by self-built dwellings, because the land is often already in possession of the private party or bought by another party.[1)]

Usage of PND:
- Inclusion in the OOH. The OOH measures the price development of homeowners.
- Inclusion in the House Price Index (HPI). The HPI measures the price development of existing and newly built dwellings purchased by households.
- An independent publication on Statline. The table is complemented by values and number of newly built transactions.

1.3 Abbreviations

The next abbrevations are used in this report:

HDIF	Hedonic Double Imputation Fisher
HDIL	Hedonic Double Imputation Laspeyres
HDIP	Hedonic Double Imputation Paasche
HPI	House Price Index
OOH	Owner Occupied Housing
PND	Price Index for Newly Built Dwellings

1) For more information about the distinction between newly built dwellings and self-built dwellings, see also Technical manual on Owner-Occupied Housing [2]

2. Calculation PND

2.1 Hedonic regression method

Based on the hedonic regression method, the PND is calculated. This is a multiple regression analysis in which several independent variables influence the dependent variable. For the PND, this means that the coherence is calculated between the sale price of the dwelling and different house features. In this way, the price development can be corrected for quality differences between the number of sold dwellings in consecutive periods. The following regression model is used to calculate a hedonic price index:

$$\log(p_i^t) = \beta_0^t + \beta_1^t x_{i1} + \ldots + \beta_k^t x_{ik} + \varepsilon_i^t$$

In this case is $\log(p_i^t)$ in period t the natural logarithm of the sales price and x_{ik} is the k^e characteristic of the i^e dwelling, sold in period t. Which characteristic is included in this model is determined on the basis of their contribution to explaining the price. Only significant contributed characteristics are included.

Based on the coefficients that the regression analysis yields, a selling price is estimated. The estimated sale price is then the input for the calculation of a price development. In this research it is chosen to estimate the sales price of the reference period and the reporting period. This method of hedonic regression is also called the double imputation-method [1].

The double imputation method had three variants of hedonic regression: Laspeyres, Paasche and Fisher. All three are based on the same above regression, but differ in how the price index is constructed. Given the results of the research the double imputation method of Fisher has been proved to be the most suitable to calculate an index for the Netherlands total. This index is the geometric average of the Laspeyres index and the Paasche index. The following formula is a representation of the Hedonic Double Imputation Laspeyres Index (I_{HDIL}):

$$I_{HDIL}(t,0) = \frac{\exp\left(\sum_{k=1}^{K} \hat{\beta}_k^t \bar{x}_k^0\right)}{\exp\left(\sum_{k=1}^{K} \hat{\beta}_k^0 \bar{x}_k^0\right)}$$

The next formula is a representation of the Hedonic Double Imputation Paasche Index (I_{HDIP}):

$$I_{HDIP}(t,0) = \frac{\exp\left(\sum_{k=1}^{K} \hat{\beta}_k^t \bar{x}_k^t\right)}{\exp\left(\sum_{k=1}^{K} \hat{\beta}_k^0 \bar{x}_k^t\right)}$$

The next formula is a representation of the Hedonic Double Imputation Fisher Index (IHDIF):

$$I_{HDIF}(t,0) = \left(I_{HDIL}(t) I_{HDIP}(t)\right)^{[1/2]}$$

In these formulas shows $\hat{\beta}_k^t$ the estimated coefficient for the k^e characteristic in period *t*, $\bar{x}_k^0$ the average of the k^e characteristic in period 0, $\bar{x}_k^t$, is the average of the k^e characteristic in period t and *K* shows the number of house characteristic that is used in the hedonic model.

Using the hedonic method the selling price is estimated. To check whether this estimated price is approximately equal to the actual price and whether the multiple regression model is therefore suitable to calculate the index, different assumptions are tested. For instance it is examined whether there is a linear relationship between selling price and house characteristic and if there is no/weak correlation between the house characteristic. Using Cook's distance, outliers are determined to help continue improvements in the index. The price index is calculated on the basis of the data without these outliers.

2.2 Characteristics of the hedonic regression method

An important characteristic of a double imputation hedonic regression method is that a price development can be calculated without a dwelling to follow over time or to compare identical dwellings. In the first case, a dwelling is not a newly constructed dwelling if the dwelling is sold again. Then it is an existing dwelling. In the second case, it is almost impossible to compare a dwelling with the same characteristic over time, such as the same location, dwelling type, and so on. In a hedonic price index, adjustments have been made for differences in quality between dwellings.

2.3 Data

2.3.1 Primary perception

To calculate the PND, a process of primary perception has been set up. There is a random sample drawn from an estimated population of developers and developing companies. The sample was drawn per stratum (region X type of dwelling) and it is frequently (partially) refreshed. The companies in the sample are requested to deliver the data of the sold new dwellings each quarter. The PND is available from the first quarter of 2015.

2.3.2 Variables

To calculate the PND, the address and guarantee certification data (to identify the dwellings) and several characteristics of the dwelling are used.

Selling price (dependent variable)

The selling price is the agreed (total) price including cost of land, cost of transferring ownership of the property and any other additional cost, including VAT. If the price of the additional work is known at the time of sale, these cost are included in the sale price.

Sales period (stratification variable)

The date of sale is the moment when the last party signed the purchase agreement. The sales date is chosen above the date when the dwellings is completed, because the price is already determined and agreed at the time of drawing. Also the date when the dwelling is completed can differ per dwelling and within a project. The time between the sales date and when the dwelling is completed can really add up in contract to existing dwellings where this term is fairly stable.

The following characteristics of dwellings are used to calculate the PND:

Characteristic 1: living area (independent variable)

The living area is the usable floor area of the dwelling according to NEN 2580, consisting the usable floor area of the residential function complemented with the usable area of other indoor spaces. The living area is specified in square meters.

Characteristic 2: type of dwelling (independent variable)

One of the independent variables is the type of dwelling. There are five types of dwellings:

- Corner dwelling
- Between or semi detached dwelling
- Two-under-one-roof dwelling
- Detached dwelling
- Apartment

Characteristic 3: location (independent variable)

Companies note on the survey where the dwellings is located and this place is used to find the province. The name of the province is used in the calculation of the PND.

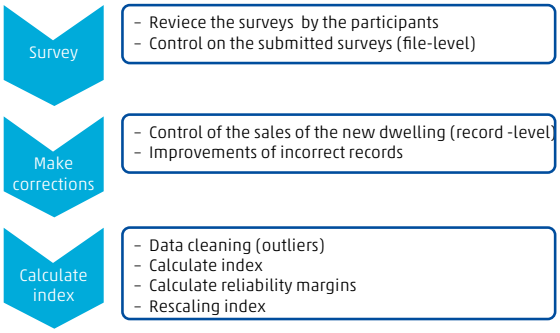
Characteristic 4: garage (independent variable)

Companies give up if there is a garage present with the dwelling and if this garage is indoor or not. Depending of answer the area of the garage is included of excluded of the living area.

The other variables of the survey are used for validation of the results.

2.4 Calculation PND

The calculation of the PND based on the hedonic regression method is conducted in three steps. The following diagram gives a clear overview of the various processes that are passed through to calculate the PND. In the first phase the data is collected and made ready for reading the records. In the second phase, different actions carried out to increase the quality of the data. The third phase refers to the calculation of the price index. The final step in this phase, the rescaling of the index, is important because the reference period still needs to be adjusted. The average of the four quarters of 2015 is determined and this level is in the index equated to 100. This is the reference period (2015=100).



In addition to the price index the average selling price of new built dwellings is calculated. This is done after cleaning the data with the following calculation:

$$\bar{p} = \frac{\sum p}{n}$$

In this formula $\sum p$ shows the sum of all observed prices, n is the number of observed prices and $\bar{p}$ for the calculated average.

3. Calculation method number of sold newly built dwellings

3.1 Cooperation CBS-Land register

The PND uses data of selling developers. Because this observation is based on a sample, not all transactions are included in this research. For mapping the number of sold new built dwelling a method has been developed in cooperation with the Land registry. This method is explained below.

3.2 Data

The basis of the method is a combination between Key Land-Register and the basic registration Addresses and Buildings (BAG). The transaction file of the Land registry office, built up from the Key Land-registry, contains all real estate transactions in the Netherlands since 1993. The database is filled with mandatory information provided by notaries. The BAG contains information about all buildings in the Netherlands. Including the intended use of the object, year of construction and the status of the property (for example planning permission had been granted, construction starts, in use, and so on)

3.3 Method for determining the number of newly built dwellings

There is no suitable variable available to filter transactions of newly built dwellings from the transaction file of the Land registry. The sale of a new dwelling is not (always) registered. For this reason, there is a step-by-step filter which was developed to find transactions of newly built dwellings as accurately as possible. Even before the data is filtered, the transaction data is enriched with data from the BAG.

The following questions will be asked for each transaction during the filtering:
- Is the date of plot before or after the start of the registration by Land registry in 1993?
- Is it the first transaction of the plot to a private individual.
- Is it the first transaction of the object (dwelling) to a private individual?
- How large is the difference between the year of construction and the time of sale?
- Was the property formerly a rental?

A combination of the answers to these questions lead to the indication of whether it is a transaction of a newly built dwelling or not.

3.4 Calculation of the number and total value purchase price of sold dwellings

If the newly built dwellings are derived from the Land register file, the sum of the number of transactions will be calculated per period. This is the number that is displayed in the table on Statline. The numbers are then multiplied by the average sales prices, which may be calculated from the primary perception (chapter 2) with the following formula:

$$P = N \cdot \bar{p}$$

In these formula $\sum p$ shows the sum of all observed prices, n is the number of observed prices and $\bar{p}$ for the calculated average.

In this formula $\bar{p}$ shows the average sales price from the primary perception, N is the calculated number of Land registry transaction and P for the calculated total sales price.

References

[1] Eurostat (2013). Handbook on Residential Property Prices Indices (RPPIs).

[2] Eurostat (2012). Technical manual on Owner-Occupied Housing for Harmonised Index of Consumer Prices (version 2.0).

Explanation of figures

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2017–2018	2017 to 2018 inclusive
2017/2018	Average for 2017 to 2018 inclusive
2017/'18	Crop year, financial year, school year, etc., beginning in 2017 and ending in 2018
2015/'16–2017/'18	Crop year, financial year, etc., 2015/'16 to 2017/'18 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress: Statistics Netherlands
Design: Edenspiekermann

Information
Telephone +31 88 570 70 70
Via contact form: www.cbs.nl/information