

# House Price Indexes: Hedonic, Repeat Sales and SPAR Methods

Jan de Haan  
Statistics Netherlands

December 15, 2009

# Introduction

## Statistics Netherlands (CBS)

Independent government organization

More than 2000 employees

Five Divisions

- Macro-economic statistics (national accounts, prices, etc.)
- Business statistics
- Social and spatial statistics
- Methodology
- IT services

Two establishments (The Hague, Heerlen)

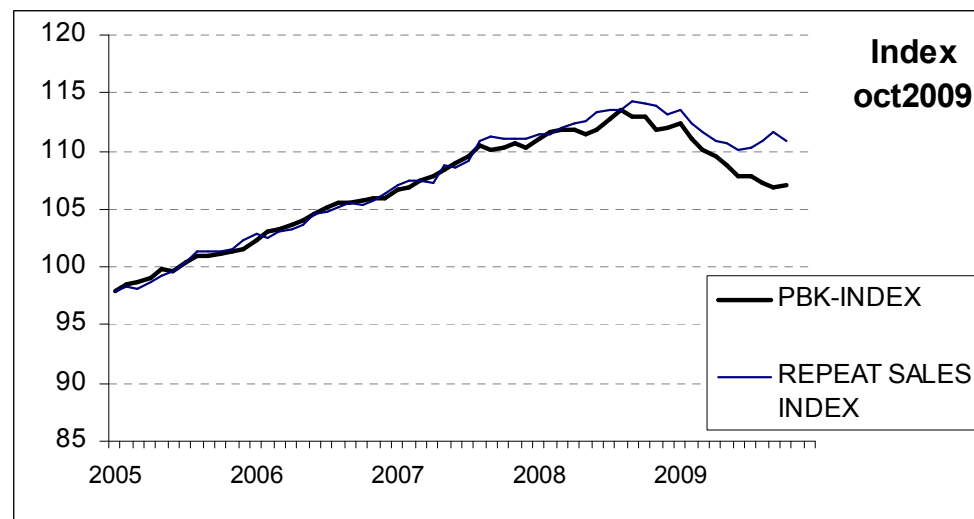
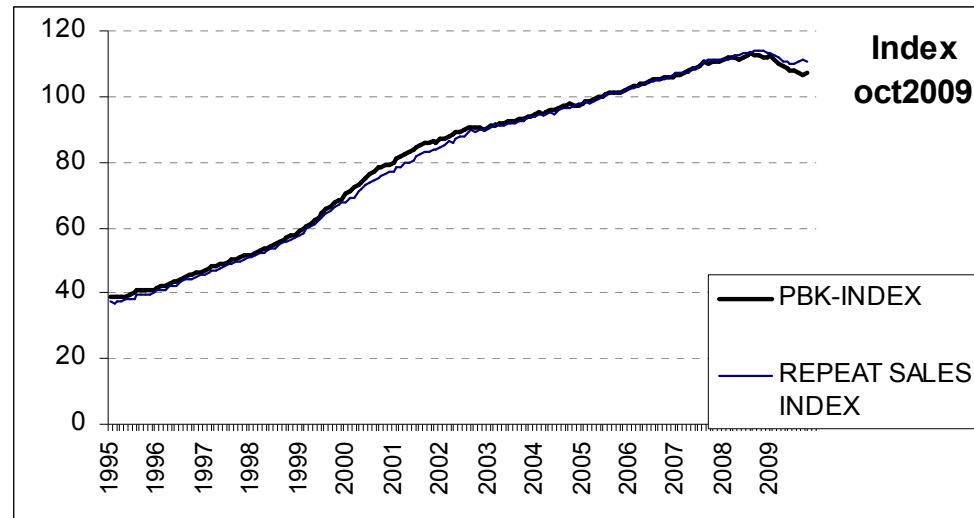
# Our new buildings

Heerlen



The Hague

# House price change in the Netherlands



# Outline

- Importance of House Price Indexes
- NVM index
- Eurostat pilot project
- Measurement problems
- A simple method: stratification
- Hedonic method
  - hedonic modeling
  - time dummy index
- Repeat sales method
- SPAR method (CBS index)
- Hedonic and SPAR index numbers using NVM data

# Importance of House Price Indexes

Housing represents a **substantial part of wealth** in a country;  
can have a major impact on the economy

Note: HPI's are not part of the Dutch national CPI

Banks: house price indexes needed for **mortgage portfolio management**

General public: monitoring the **value change** of houses

# NVM Index

Until January 2008 CBS did not publish an HPI (for reasons that are not clear)

Important source: NVM index (Dutch Association of Real Estate Agents)

Some **drawbacks**:

- Median price index; suffers from quality mix changes
- Does not account for quality changes of houses
- Does not cover all transactions
- Is not evenly spread across the country
  
- **Strong point**: very timely

# Eurostat Pilot Project

Eurostat: Statistical agency of the EU (part of the European Commission)

Aim of the pilot project: **compilation and publication of HPI's across EU** in a more or less harmonized way

**Differences in methods remain** as they depend on data availability (prices and house characteristics)

Since January 2008: publication of HPI's for the Netherlands jointly by CBS and the Land Registry

Method makes use of **sale prices and appraisals** (official house valuations); to be explained later on

## Measurement Problems

- Each house is a **unique item** (location!)
- Houses are **sold very infrequently**, so ....
- the value of a house cannot be tracked over time and the '**matched-model principle**' **breaks down**
- Change in average house prices suffers from **quality mix bias**
- **Sample selection bias** when houses sold do not represent the housing stock well
- **Quality changes** of individual houses (depreciation and renovations)

## A Simple Method: Stratification ('Mix adjustment')

- i) Subdivide (stratify) all sales into a number of cells or strata, e.g. according to type of house, region
- ii) Compute HPI's for all cells using means or medians
- iii) Compute a weighted average of those indexes

Controls for quality mix changes between strata; easy to explain

### Drawbacks:

- Mix changes within strata
- Quality changes of individual houses
- Sample selection bias (partially controlled for)

# Hedonic Regression

Hedonics - stems from hedonism, something that gives you pleasure

## Basic idea

Price of a house is determined by its characteristics, both physical (number of rooms etc.) and non-physical (location)

$$p_h^t = f(z_{1h}^t, \dots, z_{Kh}^t, e_h^t)$$

$p_h^t$  price of house h in period t

$z_{1h}^t, \dots, z_{Kh}^t$  set of characteristics

$e_h^t$  error term

## Hedonic Modeling: Pooled Regression

Semi-logarithmic model:

$$\ln(p_h^t) = \alpha + \sum_{t=1}^T \delta^t D_h^t + \sum_{k=1}^K \beta_k z_{hk}^t + e_h^t$$

$D_h^t$  **Time dummy variable**, which measures the effect of 'time' on price  
(equal to 1 if observation comes from period t, 0 otherwise)

Estimating by OLS on the **pooled data** of periods  $t=1, \dots, T$ :  $\hat{P}_{TD}^{0t} = \exp(\hat{\delta}^t)$

is a price index which measures aggregate price change with respect to the  
base period 0

What is the **index number formula** behind this implicit price index?

## Time Dummy Hedonic Method

**Index number theory:** What is the 'best' way to measure aggregate (price) changes over time?

**Price index number formula:** how best to average the price changes of different items

In this case: **no price changes available** (houses sold in period t are seldom sold in period 0)

**Implicit (OLS) Time Dummy index**

$$\hat{P}_{TD}^{0t} = \frac{\prod_{h \in S^t} (p_h^t)^{1/n^t}}{\prod_{h \in S^0} (p_h^0)^{1/n^0}} \exp \left[ \sum_{k=1}^K \tilde{\beta}_k (\bar{z}_k^0 - \bar{z}_k^t) \right]$$

## Time Dummy Hedonic Method (2)

$$\hat{P}_{TD}^{0t} = \frac{\prod_{h \in S^t} (p_h^t)^{1/n^t}}{\prod_{h \in S^0} (p_h^0)^{1/n^0}} \times \exp \left[ \sum_{k=1}^K \tilde{\beta}_k (\bar{z}_k^0 - \bar{z}_k^t) \right]$$

ratio of geometric mean prices

quality-adjustment factor

Quality-adjustment factor adjusts the ratio of mean prices for changes in the quality mix and for quality changes of individual houses

(OLS) Time Dummy index is unweighted and geometric

## Time Dummy Hedonic Method (3)

### Drawbacks

- **Revision**: adding new data changes all coefficients and thus all previously computed index numbers
- **Sample selection bias**
- Underlying assumption: **constant coefficients** (shadow prices) of characteristics
- **Data on all relevant characteristics needed** (including location)

The last disadvantage has led to the development of a method that needs no characteristics: repeat sales approach

## Repeat Sales Method

Uses only observations (price changes) of houses that have been repeatedly sold – twice or more during  $0, \dots, T$

Suppose  $s$  denotes the period that a house was sold for the first time and  $t$  the period it was sold for the second time ( $s < t$ ). The price change then is  $p_h^t / p_h^s$

Problem: the observed **price changes relate to different periods**. How to combine them into an index going from period 0 to period  $t$ ?

Starting point: time dummy hedonic model for periods  $t$  and  $s$

$$\ln(p_h^t) = \alpha + \sum_{t=1}^T \delta^t D_h^t + \sum_{k=1}^K \beta_k z_{hk}^t + e_h^t$$

## Repeat Sales Method (2)

$$\ln(p_h^s) = \alpha + \sum_{s=1}^T \delta^s D_h^s + \sum_{k=1}^K \beta_k z_{hk}^s + e_h^s$$

Assumption: **no quality change** of a house (characteristics stay the same;  $z_{hk}^s = z_{hk}^t$ )

Subtract time dummy model pertaining to s from model pertaining to t:

$$\ln(p_h^t) - \ln(p_h^s) = \ln(p_h^t / p_h^s) = \sum_{t=1}^T \delta^t D_h^t + \sum_{s=1}^T \delta^s D_h^s + (\varepsilon_h^t - \varepsilon_h^s)$$

Can be written as

$$\ln(p_h^t / p_h^s) = \sum_{t=1}^T \gamma^t D_h^{*t} + \mu_h^t$$

## Repeat Sales Method (3)

$D_h^{*t}$  is a time dummy with the value 1 in the period that the resale occurs, -1 in the period that the previous sale occurs, and 0 otherwise

Estimating by OLS regression on the **pooled, repeat sales data** of  $t=0, \dots, T$

$P_{RS}^{0t} = \exp(\hat{\gamma}^t)$  is the **Repeat Sales price index**

**Advantages** compared with Time Dummy method:

- No characteristics data needed
- Controls automatically for location (only 'matched houses')

## Repeat Sales Method (4)

**Disadvantages** compared to Time Dummy method:

- No quality adjustment for individual houses
- Possibly larger sample selection bias

Also

- Revision when new data is added to the sample
- Assumption of constant characteristics parameters
- Difficult to explain ('black box')

Most important house price indexes in the USA (Standard and Poor's/OFHEO) are based on the Repeat Sales approach

Land Registry published a Repeat Sales index for the Netherlands between 2006 and 2008

# SPAR Method

Revision of Repeat Sales index numbers problematic for official statistics

CBS and University of Delft developed a house price index using the **Sale Price Appraisal Ratio** method, applied to Land Registry data (similar to QV's method)

Land Registry data:

- **sale prices** for all transactions
- only a few characteristics, including type of house and address

Appraisals (official valuations of houses, for tax purposes):

- **WOZ data**
- available for every real estate object

## SPAR Method (2)

WOZ-value: **estimate of the market value** ('price') of a house in January of some year (the base period 0)

**For each house** sold in period  $t$  we have a sale price  $p_h^t$ , an appraisal value  $a_h^0$ , and thus **a price change**  $p_h^t / a_h^0$

This makes it possible to construct matched-model price indexes

Arithmetic index: 
$$\frac{\sum_{h \in S^t} p_h^t / n^t}{\sum_{h \in S^t} a_h^0 / n^t} = \sum_{h \in S^t} s_h^0 (p_h^t / a_h^0)$$

Geometric index 
$$\prod_{h \in S^t} (p_h^t / a_h^0)^{1/n^t}$$

## SPAR Method (3)

Problem: index differs from 1 in base period 0 if appraisals differ from sale prices

Solution: rescaling

Arithmetic SPAR index (CBS)

$$\hat{P}_{A,SPAR}^{0t} = \frac{\sum_{h \in S^t} p_h^t / n^t}{\sum_{h \in S^t} a_h^0 / n^t} \left[ \frac{\sum_{h \in S^0} p_h^0 / n^0}{\sum_{h \in S^0} a_h^0 / n^0} \right]^{-1} = \frac{\sum_{h \in S^t} p_h^t / n^t}{\sum_{h \in S^0} p_h^0 / n^0} \left[ \frac{\sum_{h \in S^0} a_h^0 / n^0}{\sum_{h \in S^t} a_h^0 / n^t} \right]$$

Geometric SPAR index

$$\hat{P}_{G,SPAR}^{0t} = \prod_{h \in S^t} (p_h^t / a_h^0)^{1/n^t} \left[ \prod_{h \in S^0} (p_h^0 / a_h^0)^{1/n^0} \right]^{-1} = \frac{\prod_{h \in S^t} (p_h^t)^{1/n^t}}{\prod_{h \in S^0} (p_h^0)^{1/n^0}} \left[ \frac{\prod_{h \in S^0} (a_h^0)^{1/n^0}}{\prod_{h \in S^t} (a_h^0)^{1/n^t}} \right]$$

## SPAR Method (4)

Like the hedonic time dummy indexes, SPAR indexes can be written as the product of the ratio of (arithmetic or geometric) mean prices and a quality (mix) adjustment factor

**Advantages** compared with Repeat Sales method:

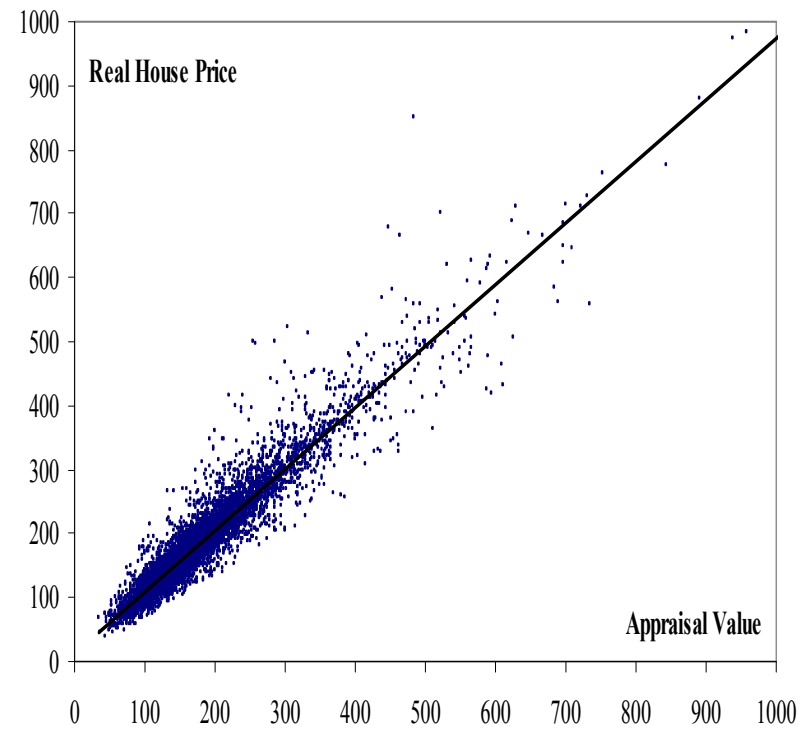
- uses all transaction data (not only repeat sales)
- no revisions
- possibly smaller sample selection bias
- transparent

Same drawback as Repeat Sales:

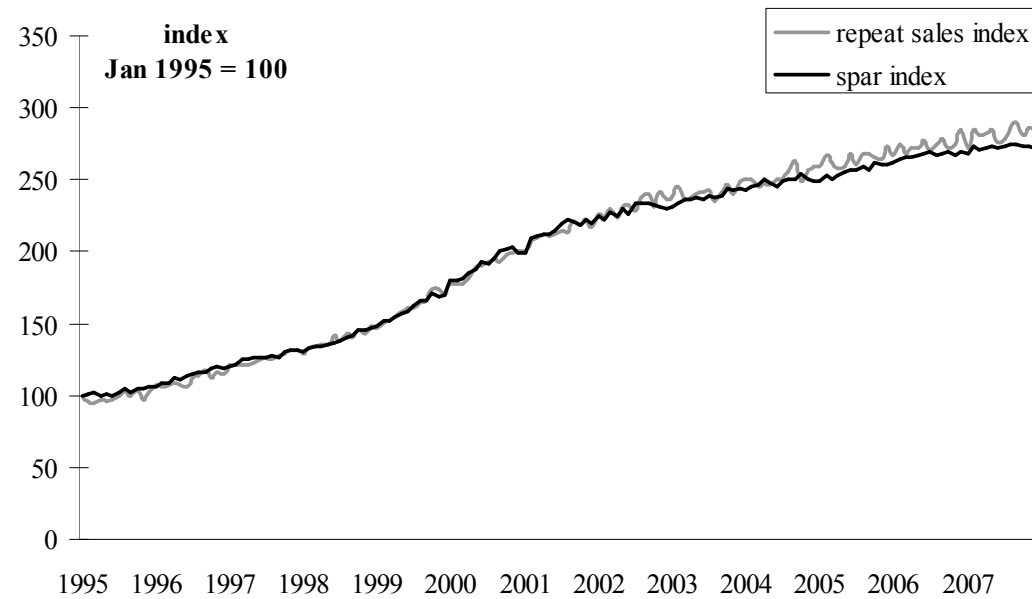
- No adjustment for quality changes of individual houses (although ....)

Also: Appraisal values should be of **sufficient quality**

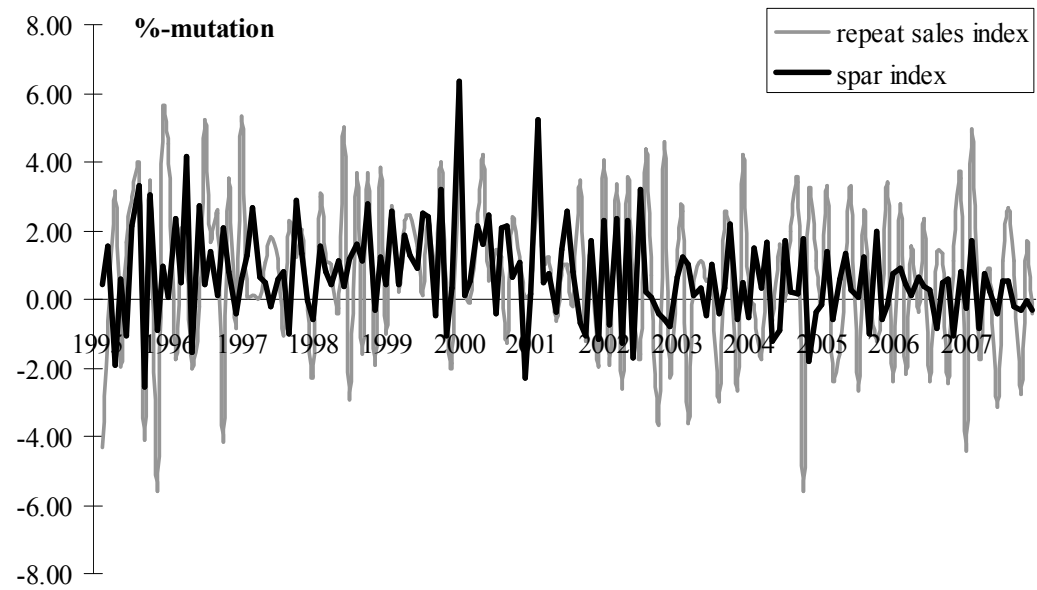
## Sale prices versus appraisals (January 2003)



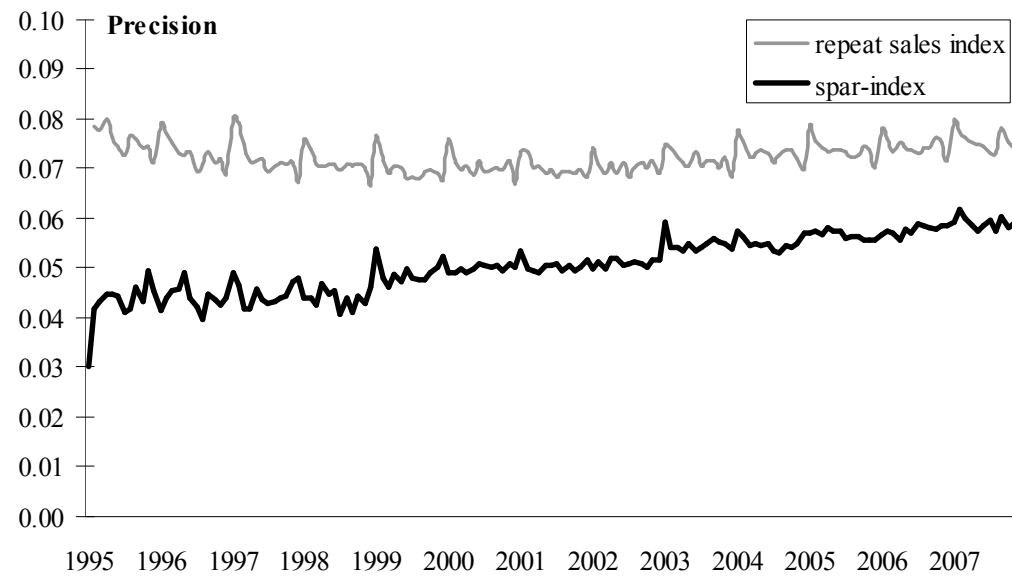
## Repeat sales and SPAR index numbers (province of Overijssel, January 1995=100)



## Monthly % changes in repeat sales and SPAR price index numbers

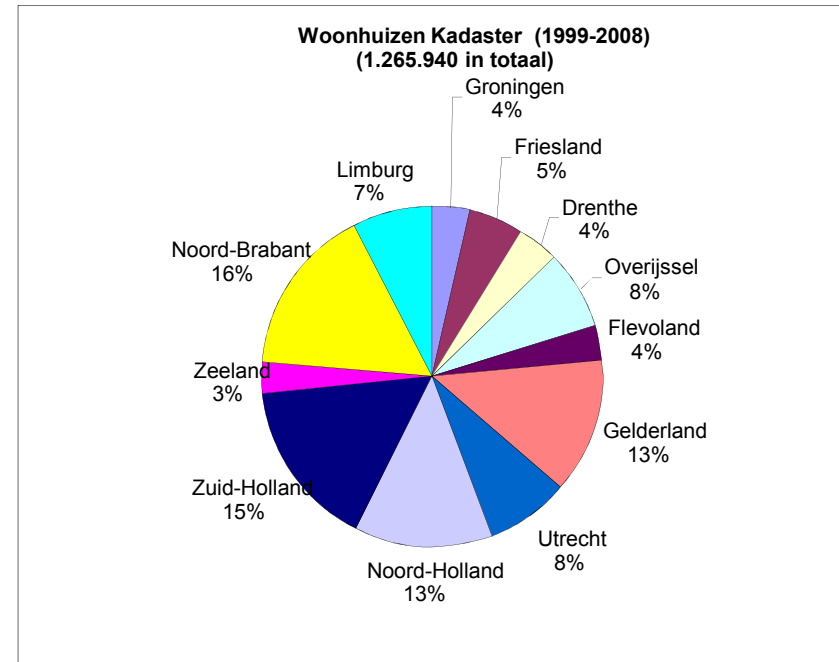
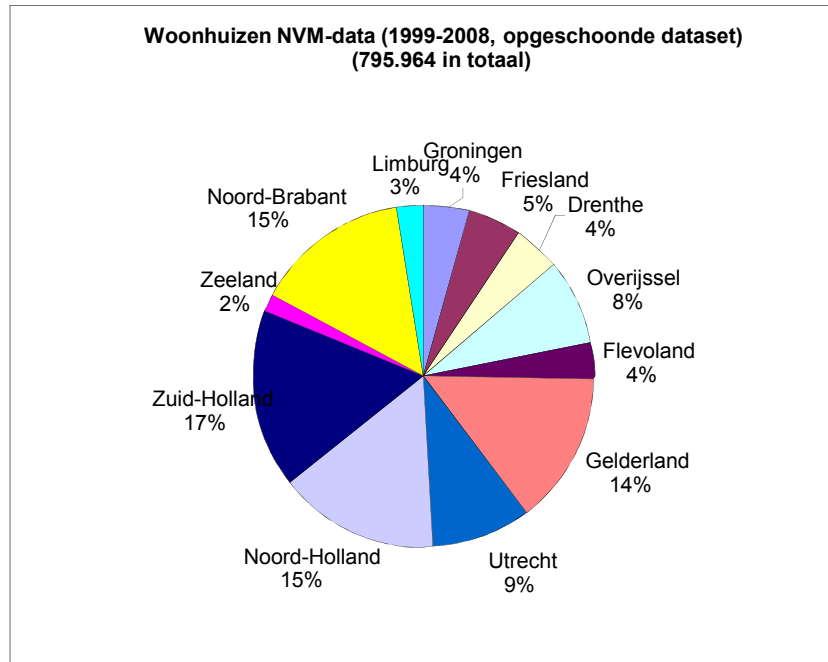


## Precision of house price index numbers (inverse measure)



# Hedonic and SPAR Index Numbers Based on NVM data

Number of transactions: NVM (63%) versus Kadaster (100%), 1999-2008



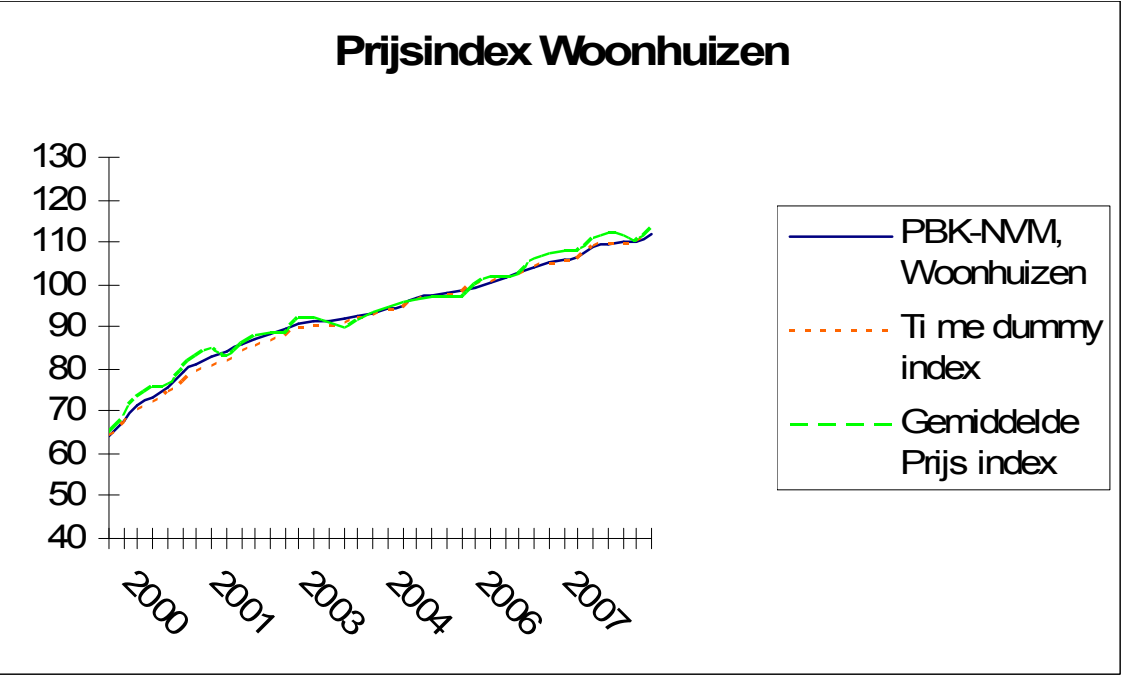
## NVM data: OLS regression results (excluding time dummies)

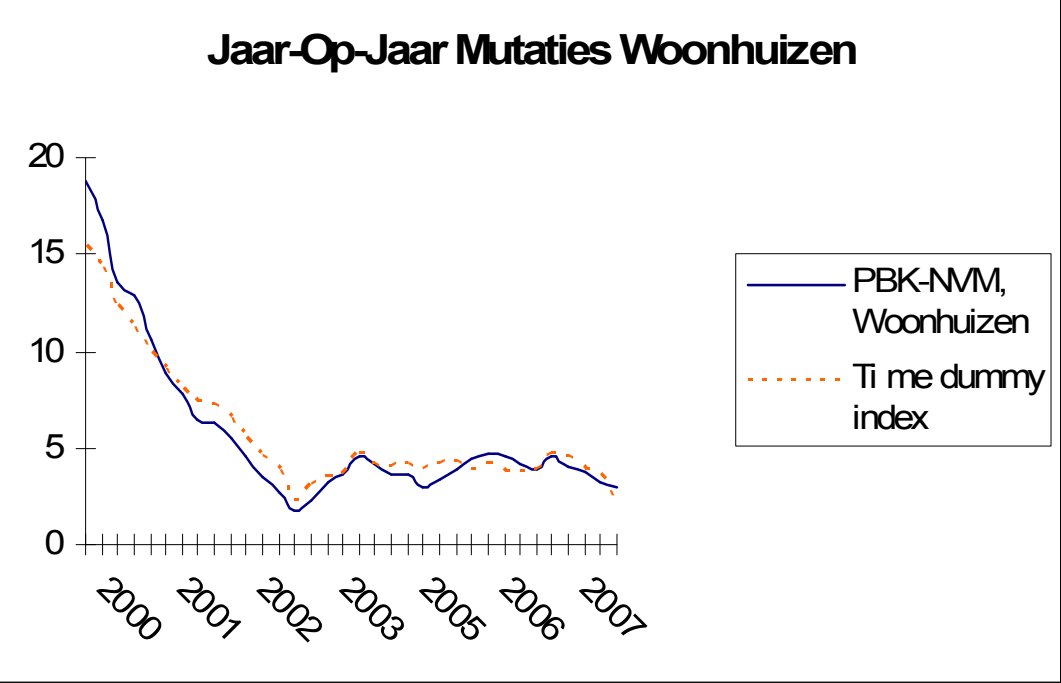
|                       | B      | Std. Error | t       |
|-----------------------|--------|------------|---------|
| (Constant)            | 8,227  | 0,012      | 674,803 |
| ln_m2                 | 0,747  | 0,002      | 358,469 |
| LN_Hoogte             | 0,167  | 0,008      | 22,231  |
| LN_TUIN_min_median_Ap | 0,016  | 0,001      | 12,888  |
| TUIN                  | 0,073  | 0,001      | 48,814  |
| LooptijdHalfJaar      | 0,020  | 0,001      | 14,671  |
| LooptijdJaar          | 0,036  | 0,002      | 20,541  |
| LooptijdLangerDanJaar | 0,036  | 0,003      | 11,451  |
| IsErfpacht            | -0,014 | 0,002      | -8,861  |
| HeeftMonument         | 0,193  | 0,004      | 51,170  |
| Maisonnette           | -0,136 | 0,002      | -64,236 |
| Portieflat            | -0,044 | 0,001      | -32,189 |
| Galerijflat           | -0,076 | 0,002      | -47,988 |
| Verzorgingsflat       | -0,642 | 0,005      | 121,506 |
| BenedenEnBovenWoning  | -0,031 | 0,008      | -3,672  |

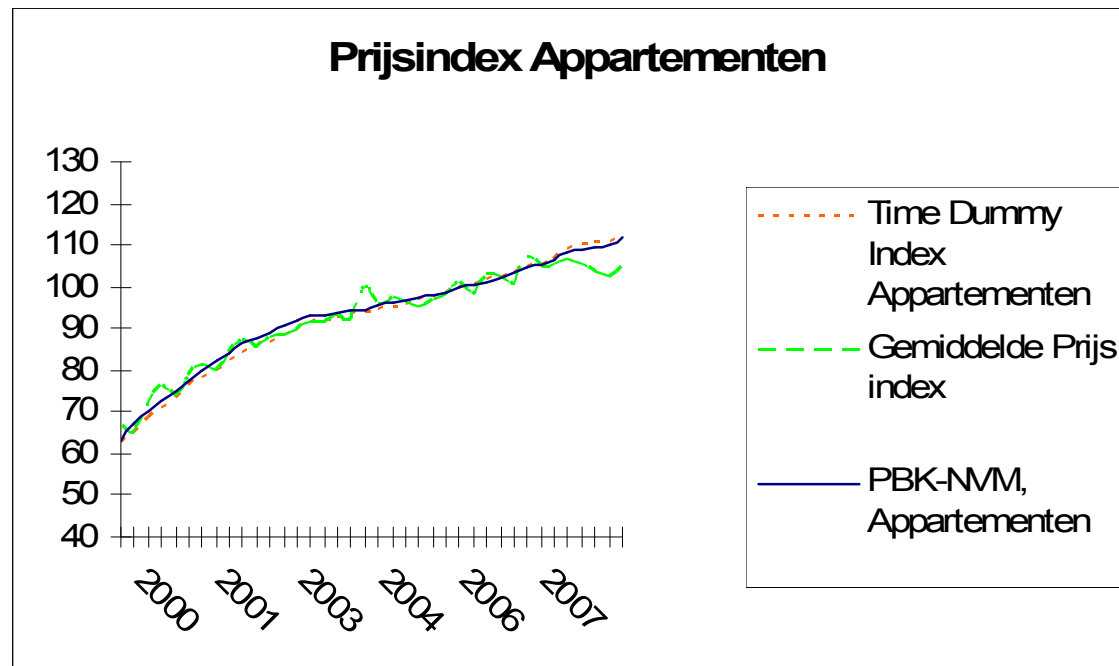
|                                  |              |              |               |
|----------------------------------|--------------|--------------|---------------|
| <b>HeeftIsolatie</b>             | <b>0,018</b> | <b>0,001</b> | <b>15,553</b> |
| <b>HeeftDakterras</b>            | <b>0,090</b> | <b>0,002</b> | <b>43,569</b> |
| <b>MeerDanEenBadKamer</b>        | <b>0,148</b> | <b>0,004</b> | <b>35,116</b> |
| <b>OnderhoudBinnenGoed</b>       | <b>0,108</b> | <b>0,002</b> | <b>54,453</b> |
| <b>OnderhoudBinnenUitstekend</b> | <b>0,199</b> | <b>0,003</b> | <b>78,678</b> |
| <b>OnderhoudBuitenGoed</b>       | <b>0,159</b> | <b>0,003</b> | <b>56,531</b> |
| <b>OnderhoudBuitenUitstekend</b> | <b>0,231</b> | <b>0,003</b> | <b>66,927</b> |
| <b>ParkeerPlaats</b>             | <b>0,161</b> | <b>0,002</b> | <b>74,477</b> |
| <b>CarportGeenGarage</b>         | <b>0,247</b> | <b>0,003</b> | <b>96,070</b> |
| <b>GarageGeenCarportNietIn</b>   | <b>0,173</b> | <b>0,003</b> | <b>49,555</b> |
| <b>GarageGeenCarportWelIn</b>    | <b>0,213</b> | <b>0,004</b> | <b>59,557</b> |
| <b>GarageEnCarportNietIn</b>     | <b>0,265</b> | <b>0,022</b> | <b>11,782</b> |
| <b>GarageEnCarportWelIn</b>      | <b>0,285</b> | <b>0,012</b> | <b>24,413</b> |
| <b>GarageMeerAutosNietIn</b>     | <b>0,295</b> | <b>0,015</b> | <b>19,175</b> |
| <b>GarageMeerAutosWelIn</b>      | <b>0,336</b> | <b>0,013</b> | <b>25,750</b> |
| <b>TweeWC</b>                    | <b>0,101</b> | <b>0,002</b> | <b>58,074</b> |
| <b>MeerDan2WC</b>                | <b>0,124</b> | <b>0,009</b> | <b>14,519</b> |

|                            |               |              |                 |
|----------------------------|---------------|--------------|-----------------|
| <b>OostGroningen</b>       | <b>-0,485</b> | <b>0,012</b> | <b>-39,268</b>  |
| <b>DelfzijlEnOmgeving</b>  | <b>-0,604</b> | <b>0,015</b> | <b>-41,054</b>  |
| <b>OverigGroningen</b>     | <b>-0,355</b> | <b>0,003</b> | <b>-112,844</b> |
| <b>NoordFriesland</b>      | <b>-0,443</b> | <b>0,006</b> | <b>-70,455</b>  |
| <b>ZuidWestFriesland</b>   | <b>-0,384</b> | <b>0,013</b> | <b>-30,652</b>  |
| <b>ZuidOostFriesland</b>   | <b>-0,428</b> | <b>0,007</b> | <b>-58,463</b>  |
| <b>NoordDrenthe</b>        | <b>-0,474</b> | <b>0,006</b> | <b>-73,060</b>  |
| <b>ZuidOostDrenthe</b>     | <b>-0,394</b> | <b>0,008</b> | <b>-47,307</b>  |
| <b>ZuidWestDrenthe</b>     | <b>-0,336</b> | <b>0,007</b> | <b>-44,880</b>  |
| <b>NoordOverijssel</b>     | <b>-0,193</b> | <b>0,005</b> | <b>-35,714</b>  |
| <b>ZuidWestOverijssel</b>  | <b>-0,291</b> | <b>0,006</b> | <b>-45,128</b>  |
| <b>Twente</b>              | <b>-0,398</b> | <b>0,004</b> | <b>-91,484</b>  |
| <b>Veluwe</b>              | <b>-0,087</b> | <b>0,003</b> | <b>-27,177</b>  |
| <b>Achterhoek</b>          | <b>-0,158</b> | <b>0,007</b> | <b>-23,681</b>  |
| <b>ArnhemNijmegen</b>      | <b>-0,132</b> | <b>0,003</b> | <b>-47,715</b>  |
| <b>ZuidWestGelderland</b>  | <b>-0,097</b> | <b>0,009</b> | <b>-10,683</b>  |
| <b>KopVanNoordHolland</b>  | <b>-0,175</b> | <b>0,008</b> | <b>-23,210</b>  |
| <b>AlkmaarOmgeving</b>     | <b>-0,019</b> | <b>0,004</b> | <b>-4,266</b>   |
| <b>IJmond</b>              | <b>-0,057</b> | <b>0,005</b> | <b>-12,128</b>  |
| <b>AgglomeratieHaarlem</b> | <b>0,023</b>  | <b>0,003</b> | <b>7,214</b>    |
| <b>Zaanstreek</b>          | <b>-0,102</b> | <b>0,005</b> | <b>-19,251</b>  |

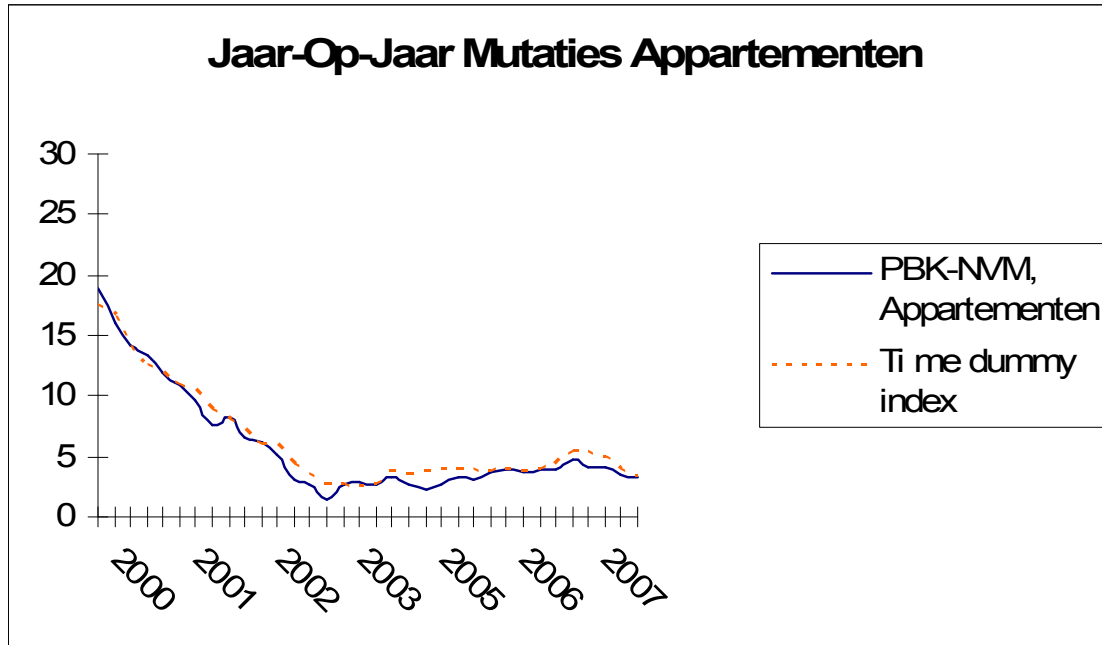
|                                   |               |              |                |
|-----------------------------------|---------------|--------------|----------------|
| <b>GrootAmsterdam</b>             | <b>0,246</b>  | <b>0,002</b> | <b>113,910</b> |
| <b>HetGooienVechtstreek</b>       | <b>0,012</b>  | <b>0,004</b> | <b>3,264</b>   |
| <b>AgglomeratieLeidenEnBollen</b> | <b>0,068</b>  | <b>0,003</b> | <b>22,940</b>  |
| <b>sGravenhage</b>                | <b>-0,206</b> | <b>0,002</b> | <b>-99,992</b> |
| <b>DelftEnWestland</b>            | <b>-0,049</b> | <b>0,005</b> | <b>-10,751</b> |
| <b>OostZuidHolland</b>            | <b>-0,071</b> | <b>0,004</b> | <b>-17,282</b> |
| <b>GrootRijnmond</b>              | <b>-0,214</b> | <b>0,002</b> | <b>-98,241</b> |
| <b>ZuidOostZuidHolland</b>        | <b>-0,225</b> | <b>0,004</b> | <b>-57,315</b> |
| <b>ZeeuwschVlaanderen</b>         | <b>-0,313</b> | <b>0,016</b> | <b>-19,435</b> |
| <b>OverigZeeland</b>              | <b>-0,223</b> | <b>0,008</b> | <b>-28,659</b> |
| <b>WestNoordBrabant</b>           | <b>-0,053</b> | <b>0,004</b> | <b>-13,970</b> |
| <b>MiddenNoordBrabant</b>         | <b>-0,046</b> | <b>0,005</b> | <b>-8,943</b>  |
| <b>NoordoostNoordBrabant</b>      | <b>0,030</b>  | <b>0,004</b> | <b>8,011</b>   |
| <b>ZuidoostNoordBrabant</b>       | <b>-0,041</b> | <b>0,004</b> | <b>-11,388</b> |
| <b>NoordLimburg</b>               | <b>-0,306</b> | <b>0,011</b> | <b>-29,009</b> |
| <b>MiddenLimburg</b>              | <b>-0,182</b> | <b>0,009</b> | <b>-19,221</b> |
| <b>ZuidLimburg</b>                | <b>-0,300</b> | <b>0,006</b> | <b>-47,312</b> |
| <b>CoropFlevoland</b>             | <b>-0,164</b> | <b>0,006</b> | <b>-26,742</b> |







### Jaar-Op-Jaar Mutaties Appartementen



## Publications

J. de Haan, E. van der Wal and P. de Vries (2009), The Measurement of House Prices: A Review of the SPAR Method, *Journal of Social and Economic Measurement*, 34, 51-86.

S.T. Jansen, P. de Vries, H.C.C.H. Coolen, C. Lamain, P. Boelhouwer, Developing a House Price Index for the Netherlands; A Practical Application of Weighted Repeat Sales, *Journal of Real Estate Finance and Economics* 37, 163-186.

P. de Vries, J. de Haan, E. van der Wal and G. Mariën (2009), A House Price Index Based on the SPAR Method, *Journal of Housing Economics* 18, 214-223.

---

B.M. Balk, E.W. Diewert, D. Fenwick, J. de Haan and M. Prud'homme, International Handbook on Residential Property Price Indexes (in preparation)