The hedonic house price index for Poland – modelling on NBP BaRN data

Narodowy Bank Polski International Workshop, Zalesie Górsne, 14 - 15 November 2013
Motivation

- Unprecedented house price boom (2006 - 2008) due to the fast mortgage growth showed the necessity of a closer monitoring of the housing market.
- The new trend in economic macromodelling stresses the importance of microfoundations.
- New task of central banks – macroprudential policy in the aftermath of the financial crises.
- Lack of reliable official statistics for the housing market.
- Until today no official House Price Index (HPI) exists for Poland (not even based on the simple average).
- Lack of appropriate, reliable data to construct HPI.

=> NBP decides to conduct its own basic survey of the housing market and compile hedonic HPI.
House prices in Poland in 16 cities and mortgage loans

- 9 cities RW trans.
- 7 cities RW trans.
- Warsaw RW trans.

Motivation
I focus on:

Which hedonic method to choose to calculate HPI for Poland? What practical problems occur during hedonic modelling and how to solve them?

- Short information of the NBP’s survey and BaRN database
- Hedonic index
- Warsaw as a special case to analyse different hedonic HPI approaches
- HPI for 16 voivodioship cities
- Conclusions
BaRN database

- The survey started in the 3q 2006 and is conducted in 16 capital cities of Polish voivodeships.
- Source: real estate agencies, developers, notary acts (municipal offices)
- Asking and transaction data on dwellings in multiunit buildings, new construction and existing stock. Since 2011 also the segments of single family houses and plots are monitored.
- Frequency: quarterly
- 17 main characteristics + 2 assessments of location + address (without the building number) + 28 proxies for location variables added using GIS (6 biggest cities; accurate for mid points of the streets)
- Data is gathered by local market analysts from the NBP Branches. They have gained knowledge and experience in their local housing markets, that enables them to correct the data and select true information. They also correct for repetitions of records and very often complete the data.

Thus our dataset is probably most complex in terms of housing attributes information as for the whole country. The week point is lack of randomized sample and lack of exact address of the real estate.
BaRN database

Number of transaction data in BaRN - higher since the survey became mandatory in 2013. The sample covers around 15 - 20% of total population. Due to the lack of correct official statistics on housing transactions samples are not random.
Quality adjustment seems to be one of the most important issue to address when measuring true price dynamics

Mean price, hedonic time - dummy and characteristics structure indices (Warsaw secondary housing market, transactions, Q/Q)
The hedonic index is any index that makes use of a hedonic function (Triplett 2006). There are 6 different methods to compute a hedonic index (ILO 2004). In my work I consider 4 direct methods (which do not require match - models)

**Diagram: Metody konstrukcji hedonicznych indeksów cen**

- **POŚREDNIE** (direct)
  - Posrednia metoda imputacji (imputation method)
  - Metoda hedonicznego dopasowania jakości (hedonic quality adjustment method)

- **BEZPOŚREDNIE** (indirect)
  - Bezposrednie metoda imputacji (imputation method)
  - Metoda cen charakterystyk (characteristic price method)
  - Metoda ze zmienymi zero-jedynkowymi czasu (ZZC) (time dummy variable method)

- **Metoda ZZC sąsiadujących okresów** (adjacent period approach)
- **Metoda ZZC wszystkich okresów** (pooled period approach)
**Imputation index**

- One hedonic model for the reference period used to calculate shadow prices in all periods and then we compare shadow with real prices – approach applied by NBP since 2009, separate models for 16 voivodeship cities

\[
\text{Index} = \frac{1}{N} \sum_{i=1}^{N} \frac{p_i^{t+1}(z_i^{t+1})}{\hat{p}_i^t(z_i^{t+1})}
\]

- or separate hedonic models for every period to compare shadow prices relevant for period \(t+1\) with real prices from period \(t\)

\[
\text{Index} = \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{p}_i^{t+1}(z_i^t)}{p_i^t(z_i^t)}
\]

\(p\) – real house price, \(z\) – vector of characteristics, dashed \(p\) – shadow price
Characteristic price index

- Laspeyres type
  \[ \text{Index} = \frac{\exp \sum_{j=0}^{\infty} a_j^{t+1} q_j^t}{\exp \sum_{j=0}^{\infty} a_j^t q_j^t} \]

- Pasche type
  \[ \text{Index} = \frac{\exp \sum_{j=0}^{\infty} a_j^{t+1} q_j^{t+1}}{\exp \sum_{j=0}^{\infty} a_j^t q_j^{t+1}} \]

\(q\) – weight for \(j\) characteristic; \(a\) – hedonic coefficient for \(j\) characteristic

**BUT** (1) experience reveals lack of statistical significance of coefficients (characteristics’ prices) in each period hedonic model; (2) the same model specification is not appropriate for every period. **How to address these issues?**
Time – dummy index

\[
\ln p_i^t = a_0 + \sum_{j=1}^{J} a_j z_{ij}^t + \sum_{t=2}^{T} b^t D^t + \epsilon^t
\]

D – time - dummies for T periods (\textit{pooled approach}) or for two adjacent periods (\textit{adjacent approach})

\[
\text{Index} = \exp(b^t) + 0.5\sigma^2
\]

Critisized for the assumption of stable price characteristics over all pooled periods (demand and supply for housing characteristics remain unchanged which is not consistent with the theoretical model of Rosen 1974).
Hedonic model for Warsaw

- 6 structural characteristics + dummies for district + 14 location variables

- Two step of outliers elimination – firstly expert assessment (area, total price, price per sq. m.); secondly according to studentized residuals.

\[
\text{Log (price sq.m.)} = f(\text{area, sq\_area, rooms, finishing standard, ownership law, construction year, district, center dist., boundary dist., metro dist., tram noise, industrial noise, cemetery, discounter, medical centers, hospitals, No green areas, green dist., protected green points dist., industrial points, trade centers dist.})
\]
HPI for Warsaw (Q/Q)
HPI for Warsaw (Q/Q)
Hedonic HPIs for Warsaw (3 q 2006 = 100)
Pooled time – dummy index for Warsaw (1 q 2009 = 100)

- No much need for revisions of the index when adding new quarter data
In case of models for one quarter data (characteristic price approach) most of coefficients are statistically insignificant on 10% level.

P-values of hedonic models for characteristic price index.
... and changes substantially from period to period and have wrong signs although the RESET test proves a good specification for most of the models.
implicit prices of construction year < 1945

implicit prices of construction year 1945 - 1970

implicit prices of construction year 1970 - 1979

implicit prices of construction year > 2005
Characteristics price models - Implicit prices - Warsaw

**Implicit prices of flats under construction**

**Implicit prices of cooperative ownership**

**Implicit prices of Rembertów, Wesoła, Wawer district**

**Implicit prices of Mokotów**
Characteristics price models - Implicit prices - Warsaw
Hedonic model for 6 and 10 cities

- 2 separate models for 6 bigest cities and the group of 10 cities (small differences in specifications)

- 9 structural characteristics + dummies for cities

- Two step of outliers elimination – firstly expert assessment (area, total price, price per sq. m.); secondly according to studentized residuals.

\[ \log(\text{price sq.m.}) = f(\text{No of rooms (6 cities) or area (10 cities), sq_area (10 cities), location assessment, finishing standard, building technology, floor, No of floors, kitchen type, ownership law, construction year, dummy for each city}) \]

- Problem with specification test Ramsey’s RESET but all structural variables significant at 1% level…. [No of obs. - 10 583 for 6 cities and 14 861 for 10 cities]
Pooled time – dummy index for Warsaw (Q/Q)
Pooled time – dummy index for 16 cities (3 q 2006 = 100)
Conclusions and questions for discussion

- Need to deal with data selection bias (randomization of samples needed but unknown population of transactions)
- Hedonic approach gives more reliable index than simple average or median but which hedonic approach to apply?
- Detailed analysis of the Warsaw market shows that imputation and pooled time–dummy indices are preferable. They give similar results despite differences in assumptions (!) Why? What’s your experience? Why most preferable characteristic price approach is not valid? Can we suspect any systematic error in our econometric job?

- Pooled time-dummy approach chosen for composite index due to operational reasons (no need for re-specification and re-estimation of models on quarterly basis).
- I am not aware of stability of implicit price coeff. but should the structure of characteristics should be somehow controlled in this case?
Conclusions and questions for discussion

- „Big” hedonic models (pooled TD Warsaw, 6 cities, 10 cities) – significant coeff., reasonable values and good signs but problem with specification tests (Ramsey’s RESET or linktest).
- „Small” hedonic models (characteristic price approach Warsaw) – in reverse … Why? How specify such models? Are the choosen specification tests appropriate?

- Examining more sophisticated econometric modeling tools (GAM approach and spatial modelling)
- Work on statistical formula of the index (Fisher type? weighted estimator for time-dummy method?)
- HPI for primary market and single family houses (one composite HPI for all segments)
- Other suggestions from the floor most welcome!
We protect the value of money