Continuous Time Hedonic Methods

A new way to construct house price indices

Sofie Waltl
University of Graz

August 20, 2014

OVERVIEW

Overview

1. Hedonic methods to construct house price indices
2. Categories of hedonic methods
3. Continuous time hedonic method

HEDONIC METHODS TO CONSTRUCT HOUSE PRICE INDICES

"The theory of hedonic indexes is built on the proposition that the characteristics [of a product] are the variables that the buyer [...] want, and that the characteristics of the product also are costly to produce."

Triplett (2006)

Historical overview

Early works by **Waugh (1928)** about vegetables' prices, **Vial (1932)** about prices of mix fertilizers, **Court (1939)** constructing commodity prices and **Stone (1956)** analysing liquor prices.

Griliches (1961) revived hedonic approaches and investigated the relationship of auto mobile prices in the US to the various dimensions of an *auto mobile*.

Lancaster (1966) and **Rosen (1974)** constituted the conceptional basis of hedonic methods.

In 1968 the US **Bureau of the Census** constructed a hedonic price index for new single family houses.

Overview

1. Hedonic methods to construct house price indices
2. Categories of hedonic methods
3. Continuous time hedonic method

Categories of Hedonic Methods

Following the taxonomy of Hill (2012) there are three categories of hedonic methods: Time-dummy, imputation and characteristics methods.

Time-dummy method

1.) Run regression explaining (logged) house prices, p, via a vector of characteristics, X, and time dummy variables, D:

$$\log p = D\delta + X\beta + \varepsilon, \quad \varepsilon \overset{iid}{\sim} N(0, \sigma^2).$$

2.) Construct price index through estimated period-specific shadow prices:

$$\hat{P}_t = \exp(\hat{\delta}_t) \quad \text{or} \quad \hat{P}_t = \exp\left(\hat{\delta}_t - \frac{1}{2}\hat{\sigma}^2((\tilde{X}^t\tilde{X})^{-1})_{tt}\right)^1,$$

where $\tilde{X} = (D, X)$ and $t \in \{1, \cdots, T\}$ with T the number of periods.

[1]This bias-correction results from the properties of a log-normal distribution. For a general discussion for semi-logarithmic equations consult Kennedy (1981).

Imputation methods

Every house is different and the price of a specific dwelling is usually not observed in every period. This is why standard price formulae can not directly be applied in a housing context. To overcome this problem, period-wise models are estimated and used to predict the price of a specific dwelling for every period where the true price was not observed.

Using this semi-calculated data, a standard price formulae such as the Paasche, Laspeyres or Fisher formula can be applied.

Characteristics methods

1.) Construct a hypothetical dwelling that is *average* in its characteristics.

2.) Estimate period-wise hedonic models and predict the price of the hypothetical dwelling in every period.

3.) Calculate bilateral price comparisons by using a standard price formula and chain them together delivering a price index.

Overview

1. Hedonic methods to construct house price indices
2. Categories of hedonic methods
3. Continuous time hedonic method

CONTINUOUS TIME HEDONIC METHOD

- Extension of time-dummy method
- Why time-dummy?
  Single-model approach, no artificial changes in variance structure, standard errors readily available.
- Why continuous?
  Time is a continuous variable and shall be treated as such.
- Why are periods *bad*?
  - discretization of time (discretization error)
  - Introducing periods leads to averaging over a time interval. Changes in the index might be averaged out.
  - It is not clear how to choose an appropriate period length (competing goals: *long enough* to guarantee sufficient number of observations vs. *short enough* to guarantee precise estimation).
  - A priori selection of starting points of periods can have significant influence.

THE BASIC MODEL

1.) Calculate continuous time scale, e.g.,

$$\mathtt{TIME}_i = \mathtt{YEAR}_i + \frac{\mathtt{MONTH}_i - 1 + \frac{\mathtt{DAY}_i - 1}{30}}{12}.$$

2.) Use continuous variable instead of time dummies and estimate the time effect smoothly[2], i.e.,

$$\log p = f(\mathtt{TIME}) + X\beta + \varepsilon.$$

3.) Establish the price index[3]via

$$\hat{P}_t = (\exp \circ \hat{f})(t) = exp(\hat{f}(t)), \quad t \in [\min(\mathtt{TIME}), \max(\mathtt{TIME})].$$

[2]Estimation is based on penalized least squares using thin plate regression splines introduced by Wood (2003) and Wood (2006). Optimal basis dimension according to GCV criterion.

[3]To gain an unbiased estimator $\hat{P}_t$ has to be adapted as for the original time-dummy method.

Application

Data set

- data set by Australian Property Monitors
- transaction prices, dates, house characteristics (including number of bed- and bathrooms, land area and exact longitudes and latitudes) of sold houses between 2001 and 2011
- after cleaning: $435,295$ observations
- huge problem: missing data
  - $201,571$ observations are incomplete, i.e., 46.3%
  - $46,747$ incomplete observations can be refilled through simple reconstruction approach
  - final sample size: $280,471$[4]
  - incompleteness is a problem of the variables `BED` and `BATH` only

[4] Further approaches to handle missing observations in this context are under way.

Hedonic model

$$\log(p) = \beta_0 + \beta^{\texttt{AREA}} \log(\texttt{AREA}) + \sum_{j=2}^{6} \beta_j^{\texttt{BATH}} \mathbb{I}_{\{j\}}(\texttt{BATH}) + \sum_{j=2}^{6} \beta_j^{\texttt{BED}} \mathbb{I}_{\{j\}}(\texttt{BED})$$
$$+ f_1(\texttt{LONG}, \texttt{LAT}) + f_2(\texttt{TIME}) + \varepsilon,$$

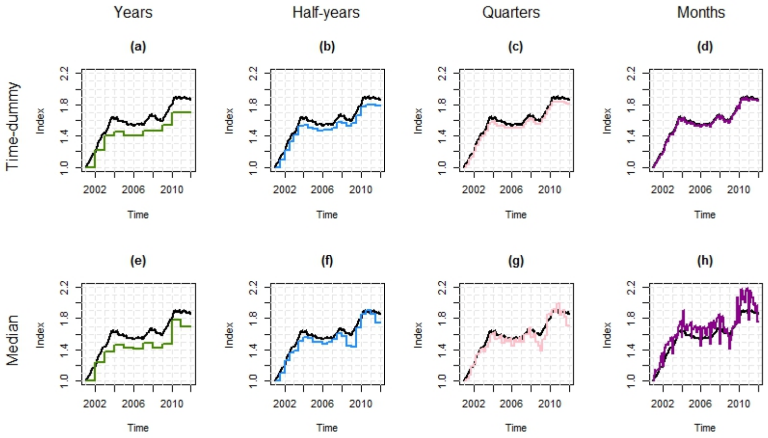
where

$$\mathbb{I}_{\{j\}}(\texttt{VARIABLE}) = \begin{cases} 1, & \texttt{VARIABLE} = j, \\ 0, & \texttt{VARIABLE} \neq j. \end{cases}$$

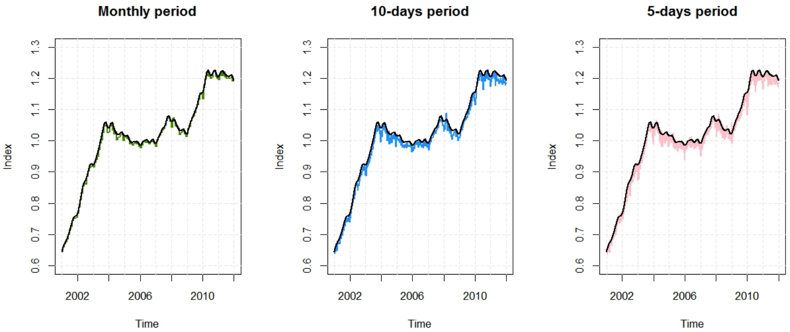
Locational effects: two-dimensional surface defined on longitudes and latitudes (see Hill and Scholz, 2014)

logged area

RESULTS


Years
Half-years
Quarters
Months
Time-dummy
Median
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Index
1.0
1.4
1.8
2.2
2002
2006
2010
Time

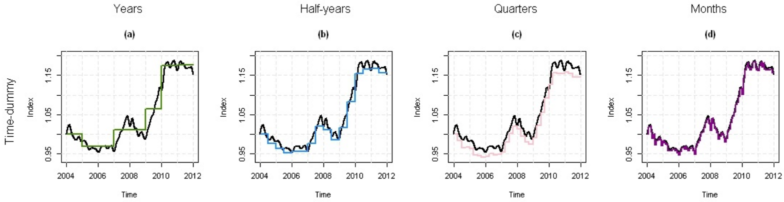
Results



Average number of observations per period:

Monthly period	3,300
10-days period	1,100
5-days period	660

RESULTS



- Truncated period: 2004-2011
- Normalization with respect to January 1, 2008

Pros and Cons

- ⊕ Single model approach
- ⊕ Circumvents problem of choosing appropriate period length and starting points
- ⊕ Peaks and troughs are not averaged out
- ⊕ Turning points are detected precisely
- ⊕ Discrete indices approach continuous index
- ⊖ Computationally more costly than time-dummy method
- ⊖ Higher complexity due to smooth estimation
- ⊗ Basic model does not allow changing shadow prices $\rightarrow$ interactions of smooth time function and characteristics as well as regular updates of geographical spline

References I

Court, A. T. (1939). *The Dynamics of Automobile Demand*, chapter Hedonic Price Indexes with Automotive Examples, pages 99–117. The General Motors Corporation, New York.

Griliches, Z. (1961). *The Price Statistics of the Federal Goverment*, chapter Hedonic Price Indexes for Automobiles: An Econometric of Quality Change, pages 173–196. National Bureau of Economic Research.

Hill, R. J. (2012). Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of Economic Surveys*, pages 879–914.

Hill, R. J. and Scholz, M. (2014). Incorporating geospatial data in house price indexes: A hedonic imputation approach with splines. Graz Economics Papers 2014-05, University of Graz, Department of Economics.

Kennedy, P. E. (1981). Estimation with correctly interpreted dummy variables in semilogarithmic equations. *The American Economic Review*, 71:801.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2):132.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34.

Stone, R. (1956). *Quantity and Price Indexes in National Accounts*. Organisation for European Economic Cooperation, Paris.

REFERENCES II

Triplett, J. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*. Organization for Economic Co-operation and Development, Paris.

Vial, E. E. (1932). Retail prices of fertilizer materials and mixed fertilizers. Cornell University.

Waugh, F. V. (1928). Quality factors influencing vegetable prices. *Journal of Farm Economics*, 10(2):185.

Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, 65(1):95–114.

Wood, S. N. (2006). *Generalized Additive Models: An introduction with R*. Chapman and Hall / CRC.

RECONSTRUCTION ALGORITHM

- There are many multiply traded dwellings in the data set (that can be identified through a unique identification number).
- If a dwelling is sold for instance twice and the number of bedroom is available at one transaction date but not at the other, the missing observation can be refilled.
- However, house characteristics might change over time due to renovation! $\rightarrow$ Refilling is subject to the following constraints:

1.) **Constancy constraint:** Available numbers of bedrooms (or bathrooms respectively) are constant.

2.) **Time constraint:** The time span between two transactions is greater than six month, i.e.,

$$\mathtt{TIME}_{\mathtt{diff}} = \mathtt{TIME}_2 - \mathtt{TIME}_1 > 0.5 \text{ years.}$$

3.) **Price growth constraint:** The average annual price growth is less than 25%, i.e.,

$$\left(\frac{\mathtt{PRICE}_2}{\mathtt{PRICE}_1}\right)^{1/\mathtt{TIME}_{\mathtt{diff}}} - 1 < 25\%.$$

Results of reconstruction algorithm

The share of incomplete recordings has been reduced from 46.3% to 35.6%. $37,657$ missing records of `BED` and $47,293$ missing records of `BATH` were reconstructed.

	BED	BATH
Constancy constraint	$4,979$	$5,188$
Time constraint	0	0
Price growth constraint	$2,119$	$2,081$

Table: Number of non-replacements due to restrictions.

back

Changing shadow prices

To the basic model

$$\log P = \beta_0 + f_1(\texttt{TIME}) + f_2(\texttt{LONG}, \texttt{LAT}) + \sum_{i=2}^{6} \beta_i^{\texttt{BED}} \mathbb{I}_{\{i\}}(\texttt{BED}) + \sum_{i=2}^{6} \beta_i^{\texttt{BATH}} \mathbb{I}_{\{i\}}(\texttt{BATH}) + \beta^{\texttt{AREA}} \log(\texttt{AREA}) + \varepsilon$$

interaction terms are added,

$$\begin{aligned} \log P =& \beta_0 + f_1(\texttt{TIME}) + f_2(\texttt{LONG}, \texttt{LAT}) + \sum_{i=2}^{6} \beta_i^{\texttt{BED}} \mathbb{I}_{\{i\}}(\texttt{BED}) + \sum_{i=2}^{6} \beta_i^{\texttt{BATH}} \mathbb{I}_{\{i\}}(\texttt{BATH}) + \beta^{\texttt{AREA}} \log(\texttt{AREA}) \\ &+ \sum_{i=2}^{6} f_i^{\texttt{BED}}(\texttt{TIME}|\texttt{BED} = i) \mathbb{I}_{\{i\}}(\texttt{BED}_t) + \sum_{i=2}^{6} f_i^{\texttt{BATH}}(\texttt{TIME}|\texttt{BATH} = i) \mathbb{I}_{\{i\}}(\texttt{BATH}) \\ &+ \sum_{i=2}^{5} f_i^{\texttt{AREA}}(\texttt{TIME}|\texttt{AREA}_{cat} = i) \mathbb{I}_{\{i\}}(\texttt{AREA}_{cat}) + \varepsilon. \end{aligned}$$

In the interaction term the continuous variable AREA is transformed to a discrete variable of five categories, $\texttt{AREA}_{cat}$.

back

ESTIMATED LAND AREA EFFECT