House Price Index (HPI)

The price index of second hand houses in Colombia (HPI), registers annually and quarterly the evolution of prices of this type of dwelling. The calculation is based on the repeated sales methodology proposed by Case and Shiller (1989).

In the repeated sales methodology an individual house price is characterized as a stochastic process in which the average variation is measured by a market price index, while the dispersion and volatility respect to the market average are modeled as a process with log-normal distribution (Calhoun, 1996).

The estimation of the index requires to identify houses that have been sold at least twice during the period of analysis. The variables used for identification of the properties are the address and the real estate register number. Then, information of the first and second transactions is merged to compute price variations. Commercial appraisal value is used since the exact house price is not available, this value is used for mortgage acquisitions.

This indicator allows to analyze factors affecting the equilibrium price of assets such as: supply shocks and the existence of speculative bubbles. Following this indicator helps to assess the housing market evolution and related sectors.

Coverage:

Initially the index included information of mortgages in the three major cities: Bogotá (including Soacha), Medellín (including Bello, Envigado and Itagüí) and Cali. In May 2014, information from more cities was considered and time series were updated including information from Barranquilla, Bucaramanga, Cucuta, Manizales, Neiva and Villavicencio.

Periodicity and publication

The base period is 1990=100 and the series are deflated using the aggregate CPI of the cities involved in the index. Both, nominal and real indexes are published.

Annual indexes:

Total housing price index corresponds to the aggregate of house price evolution for the following cities: Bogotá (including Soacha), Medellin (including Bello, Envigado and Itagüi), Cali, Barranquilla, Bucaramanga, Cucuta, Manizales, Neiva and Villavicencio.

Indexes by cities:

- HPI Bogotá (including Soacha).
- HPI Medellín (including Envigado, Bello and Itagüi).
- HPI Cali.
- HPI other cities, includes information from: Barranquilla, Bucaramanga, Cucuta, Manizales, Neiva and Villavicencio.

Indexes by appraisal values:

- HPI LIH: Low income housing.
- HPI ELIH: Excluding low income housing.

Annual indexes series are published four months after the end of the year.

Quarterly index:

In quarterly basis only the total aggregate index is produced.

Total housing price index corresponds to the aggregate of house price evolution for the following cities: Bogotá (including Soacha), Medellin (including Bello, Envigado and Itagüi), Cali, Barranquilla, Bucaramanga, Cucuta, Manizales, Neiva and Villavicencio.

Quarterly index is published four months after the end of the quarter.

Time series are revised each period because as new information is included, the estimates may change.

Source of information:

The information come from disbursements made for housing purchases offered by the main financial institutions that offer mortgages in the country:

1. Av. Villas
2. Bancolombia (includes Conavi)

3. BBVA (includes Granahorrar)
4. BCSC (includes Colmena)
5. Colpatria
6. Davivienda (includes Bancafé)

Methodology

The methodology of this indicator corresponds to weighted repeated sales proposed by Case and Shiller (1989), in which the index is build from a three-stage econometric estimation, selecting the dwelling sold at least twice during the period of analysis and that has not have significant changes in their physical structure during that period. The price dynamic for an individual house, P_{it} is characterized as a stochastic process represented by:

$$\ln(P_{it}) = \beta_t + H_{it} + N_{it} \qquad (1)$$

Where, β_t corresponds to the market price index, $\boldsymbol{H_{it}}$ is a Gaussian random walk, which describe how the change in the price of an individual house deviates from the market index variation. And $\boldsymbol{N_{it}}$ are errors normally distributed that represent the idiosyncratic differences at time *t*.

Equation (1) implies that the total percentage change in price of house *i* transacting in time periods *s* and *t* is given by:

$$\boldsymbol{\Delta V_i = \ln(P_{it}) - \ln(P_{is})} \qquad (2)$$

$$\boldsymbol{\Delta V_i = \beta_t - \beta_s + H_{it} - H_{is} + N_{it} - N_{is}} \qquad (3)$$

The following assumptions are established for the perturbation terms:

$$E(\mathrm{H_{it}} - \mathrm{H_{is}}) = 0 \;\; y \;\; E(\mathrm{H_{it}} - \mathrm{H_{is}})^2 = A(t-s) + B(t-s)^2 \qquad (4)$$

$$\boldsymbol{E(N_{it}) = 0 \;\; y \;\; E(N_{it})^2 = c} \qquad (5)$$

$$\boldsymbol{E(H_{it}N_{it}) = 0} \qquad (6)$$

Equation (6) shows that the dispersion of the length of holding period in the selling price of second hand houses is not related with the dispersion generated by the market price at the time of sale.

From this specification and assumptions, the construction of the index is divided in the following stages:

First stage: initial β estimates and estimation of the errors.
In the sample of repeat sales, the change in price of house *i*, in equation (2) can be expressed as follows:

$$\boldsymbol{\Delta V_i = \sum_{t=0}^{T} \ln(P_{it})\, D_{it}} \quad (7)$$

Where $\boldsymbol{D_{it}}$ is a dummy variable which equals 1 if the price of house *i* is observed for the second time at time *t*, -1 if the price of the house *i* was observed for the first time at time *t*, and zero otherwise.
Substituting (1) in (7) yields,

$$\boldsymbol{\Delta V_i = \sum_{t=0}^{T} \beta_t D_{it} + \varepsilon_i} \quad (8)$$

The parameter $\boldsymbol{\beta_t}$ is estimated by Ordinary Least Squares (OLS). When A or B in equation (4) are statistically different from zero, the variance of $\boldsymbol{\varepsilon_i}$ changes with the distance between repeated sales, the more distant the transactions are, the greater price variation is assumed. Therefore, to obtain more efficient estimates of $\boldsymbol{\beta_t}$ the method of Generalized Least Squares (GLS) is used.

Second stage: Estimation of the variance of the random walk
As quoted by Calhoun (1996) the predicted price of house *i* purchased in period *s* and sold in period *t* is given by:

$$\boldsymbol{\ln\left(\widehat{P_{it}}\right) = \ln(P_{is}) + (\widehat{\beta_t} - \widehat{\beta_s})} \quad (9)$$

The above equation indicates that the estimated price in period t is equal to the initial purchase price plus the expected market appreciation between the two periods. The quadratic error forecast would be given by:

$$\boldsymbol{\varepsilon_i^2} = \left(\mathbf{ln}(\boldsymbol{P_{it}}) - \mathbf{ln}\left(\widehat{\boldsymbol{P_{it}}}\right)\right)^2 = \left(\mathbf{ln}(\boldsymbol{P_{it}}) - \mathbf{ln}(\boldsymbol{P_{is}}) - (\widehat{\boldsymbol{\beta_t}} - \widehat{\boldsymbol{\beta_s}})\right)^2 \qquad (10)$$

Calculating the expected value and using the assumptions in equations (4), (5) and (6) the following equation is obtained:

$$\mathbf{E}\left(\boldsymbol{\varepsilon_i^2}\right) = \mathbf{A}(\mathbf{t}-\mathbf{s}) + \mathbf{B}(\mathbf{t}-\mathbf{s})^2 + \mathbf{C} \qquad (11)$$

The above equation sets the second stage of the procedure. Through this equation we can obtain not only consistent estimations, but also evidence about volatility and dispersion of individual house prices around the market index. The square root of the estimated values by equation (11) is used as weights in the final stage: Generalized Least Squares (GLS).

Third stage: The final estimation of β, considers the following transformation of the series, which suppose not to affect the consistency of the parameters neither the performance of the index.

$$\frac{\boldsymbol{\Delta V_i}}{\sqrt{\widehat{\boldsymbol{\varepsilon_i}}^2}} = \sum_{\boldsymbol{t=0}}^{\boldsymbol{T}} \boldsymbol{\beta_t} \frac{\boldsymbol{D_{it}}}{\sqrt{\widehat{\boldsymbol{\varepsilon_i}}^2}} + \frac{\boldsymbol{\varepsilon_i}}{\sqrt{\widehat{\boldsymbol{\varepsilon_i}}^2}} \qquad (12)$$

Finally, the index is calculated from the obtained parameters by GLS:

$$I_t = 100 \cdot e^{\tilde{\beta}_t} \qquad (13)$$

Unlike conventional indexes, this methodology produces a standard error associated to each estimation, which is given by:

$$\hat{\sigma}_{I_t} = I_t \cdot \hat{\sigma}_{\hat{\beta}_t} \qquad (14)$$

One issue of the geometric index is that it may underestimate the percentage change in the average value of a sample of houses, comprising individual houses of different price values (high medium and low housing prices). Goetzman (1992)

proposed the following correction to the geometric index, to solve the bias that can generate the previous index.

$$\widetilde{I_t} = 100e^{\widehat{\beta_t}+\frac{\sigma_t^2}{2}} \qquad (15)$$

Where $\boldsymbol{\sigma_t^2}$ is an estimate of the variance in house price growth rates associated with the diffusion of house prices after *t* periods**.**

$$\boldsymbol{\sigma_t^2 = At + Bt^2}$$

Database depuration

The repeated sales database excludes dwellings with inconsistent information such as incomplete identification of the house (address or real state identification), or excessive changes in the appraisal value between two sales.

Although the historical information is the same for the annual and quarter indexes, the calculation is independent for each periodicity. For the annual index only those sales of the same dwelling that occur in different years are considered, while for the quarter index dwellings sold in the same year but in different quarters are included in the index. I. e. repeated sales of a particular house in the same period are excluded.

Advantages

The repeated sales methodology highlights because of its econometric simplicity and direct use from data of the variables subject of measurement. Besides, it approximates to a constant quality price index, since it requires that the property has been traded at least twice during the period of analysis without having significant changes. Other advantages are the use of information of values on the same units at two points in time, as well as the possibility of its reproducibility, that is, that different analysts can obtain the same estimates if they use the same information. Additionally, the estimation of average price of individual houses under

this method presents smaller standard error compared to hedonic prices and first transaction (Case, Pollakowski and Wachter, 1991).

Disadvantages

It only uses a portion of the information available on sales transactions (those that have been sold more than once during the period of analysis) without consider information of second hand houses that only have had one owner. On the other hand, several studies have shown that the age of the house is a source of bias, which unfortunately are not controlled by the methodology, affecting the assumption of constant characteristics. Also, the methodology does not consider the influences that certain wanted characteristics of the house may have over time on prices, as the change in the preference for having two or more garages, or bathrooms, being near workplace, parks and other features that affect trends of the market and consequently the final price. Finally, because it is an econometric estimation, the time series can change as new information is available.

Time series can be consulted on the website of the Central Bank:
http://www.banrep.gov.co/es/series-estadisticas/see_precios_ipvu.htm