WORKING PAPER

Using Different Administrative Data Sources to Develop House Price Indexes for Portugal

March 2014

R. Evangelista, Â. Teixeira

Statistics Portugal, National Accounts Department [1] [2]

Abstract:

The most recent economic and financial crises highlighted the lack of important information to monitor the performance of the housing market, whose imbalances could, through wealth and other effects, impact significantly on economic growth. This gap has prompted statistical institutes across Europe to develop strategies to improve the already existing house price statistics. This paper presents the work and experience of Statistics Portugal in this area, which involved the analysis of prices information taken at different stages of the buying and selling process and supported the compilation of appraisals-based and transactions-based house price indexes.

Keywords:

Accessing Administrative Data, Policy Indicators, Hedonic Regression, House Price Index

JEL Classification:

C81, R31, E31, C43

[1] The views expressed in this paper are solely those of the authors and do not necessarily reflect the position of Statistics Portugal or of any other institution mentioned in it.

[2] This is a working paper covering some topics that are being developed. The results and the conclusions presented here should, for this reason, be regarded as preliminary. Comments about the methodology and obtained results are welcome. Contact email: rui.evangelista@ine.pt.

Introduction

The most recent economic and financial crises highlighted the lack of some important information on the housing market, whose imbalances could, through wealth and other effects, impact significantly on the economic performance of countries. This information gap has prompted statistical institutes across Europe to develop strategies to improve the already existing house price statistics and to the production of new official real estate market indicators.

In the case of Portugal, data on transaction prices were generally not available for price index compilers, and the only indication about the evolution of housing market prices was provided by a private producer using asking prices collected from a real estate portal (CI, 2006).

Statistics Portugal's chosen strategy to fill this information gap was based on the exploration of bank appraisals and fiscal administrative data sources rather than in the design of new (and costly) surveys[3]. The application of this "double data source approach", which resulted in the production of house price indexes (HPIs) based on bank appraisals and fiscal administrative data, was essentially imposed by an evaluation of the potentialities that each source of information had to produce good-quality statistics without jeopardising the need to comply with the new European legal requirements in this area[4]. In fact, while considered from the outset as probably the most suitable data source for the derivation of a (transactions-based) HPI, fiscal administrative data was not, and contrary to bank appraisals data, ready to be used at the start of the project on real estate indicators.

This paper provides an account of the work that was carried out in developing new quarterly HPIs for Portugal. The paper will be focused not only on the way data have been incorporated in the production of the appraisals- and transactions-based HPIs but also on the comparison of HPI results based on house price information taken from different stages of the buying and selling process (i.e., asking prices, appraisal values and transaction values).

The present paper is divided into four sections. Section one revolves around the data and compilation methodology that is used in the production of the appraisals-based HPI. Section two provides a description of the data that are available for the compilation of the transactions-based HPI and presents, in addition, a summary of the outputs of the hedonic regressions used in the derivation of this indicator. Section three presents and compares the results of the asking-, appraisals-, and transactions-based HPIs for the 1Q2009 - 4Q2013 period. The comparisons will not only be concerned with the evolution of prices but also with the coherency of the volume information that could be extracted from appraisals and fiscal data sources. Finally, some conclusions and final remarks are presented at the end of this paper.

[3] Administrative data simply refers to information collected by public or private institutions for administrative (not statistical) purposes.

[4] These include Regulation No 93/2013, which defines the third quarter of 2012 as the start for the regular production and transmission of HPIs to Eurostat. Another obligation stems from the fact that the HPI is included in the scoreboard of indicators that is used for the early detection of macroeconomic imbalances in member states of the European Union (European Commission, 2011).

1. Appraisals-based HPI

1.1. Data description

Bank appraisals data are collected through a monthly survey that covers the most important commercial banks providing mortgage credit [5], and include information that is typically generated before any actual purchase takes place.

The next table provides some descriptive statistics for the bank appraisals value (APPRVAL) and gross floor area (GRFLOORA). The table refers to the 2009-2013 period and is presented with a division between new and existing dwellings and between apartments and houses.

Table 1: Descriptive statistics of available bank appraisals data

		N	*Mean APPRVAL*	*Median APPRVAL*	*Stdev APPRVAL*	*Mean GRFLOORA*	*Median GRFLOORA*	*Stdev GRFLOORA*
Existing	*Apartments*	158,267	132,335	103,900	283,581	132	121	135
	Houses	82,715	176,442	145,000	212,326	236	196	331
New	*Apartments*	52,404	175,640	150,900	157,536	162	153	88
	Houses	33,057	235,084	204,220	157,219	298	263	305

Notes: *APPRVAL in €; GRFLOORA in m2.*

The number of bank appraisals available for the compilation of the appraisals-based index exceeds the 326 thousand (an average of more than 16.3 thousand observations per quarter.

1.2. Advantages and disadvantages of the bank appraisals data

The main advantages of the bank appraisals data are the large number of observations, which were considered enough to be used in the compilation of a house price index.

The main disadvantage of this source is the fact that the survey does not include information on transactions prices. As such, appraisals have to be considered as a proxy of transaction prices. In addition, the survey does not include many variables characterizing the quality of appraised properties and the sub-universe of cash transactions is not covered.

1.3.Compilation methodology

The appraisals-based HPI is an index that is compiled using a stratification approach. The strata are defined using the following basic design:

- Location of appraised dwelling: as defined by the 7 NUTS II regions for Portugal;

[5] In 2011, sampled banks accounted for nearly 95% of all mortgage loans provided to private households by financial institutions.

- Dimension of appraised dwelling: 2 categories, which are based on the number of rooms;
- Type of dwelling: house or apartment; and
- Occupancy status of dwelling: 2 categories; i.e. "new" and "existing" dwellings, which is built using the age of dwelling as a proxy of occupancy status.

The application of this scheme yields 56 elementary indexes (7x2x2x2), which are compiled using the Jevons formula.

In each quarter, appraisals are checked against pre-established upper and lower (cut-off) values. For each dwelling category/cluster (35 in total), cut-off values are defined on the basis of the top/bottom 1% appraised figures found in the previous complete year of data. The choice and definition of these 35 clusters was based on the results taken from a cluster analysis study aimed at identifying the most important factors explaining appraisals variability.

2. Transactions-based HPI

2.1.Data description

The information used in the compilation of the transactions-based HPI is derived from the combination of two different fiscal administrative data sources, which are transmitted to Statistics Portugal on a monthly basis.

The records of the Municipal Tax on Real Estate Transfer (IMT), which provide information on transaction prices, constitute the first data source of the transactions-based HPI. The IMT is a tax levied on property transfers, which is calculated based on the declared value of the transaction or on the fiscal appraisal value of the property, depending on which is higher[6]. The IMT is paid by the buyer immediately before the property changes hands[7]. Available IMT data go as far back as the beginning of 2009.

The records of the Local Property Tax (IMI), which provide information on the characteristics of transacted dwellings, compose the second data source of the transactions-based HPI. The IMI is a municipal tax levied on the value of the property. The value relevant for IMI purposes is an appraisal of the value of the property (and not its market price), which is based on an evaluation of its attributes (e.g., area, quality of location, etc)[8]. At present, Statistics Portugal has information about the characteristics of nearly 5.2 million dwellings[9].

[6] It is generally accepted that under this system, which takes into account declared and appraised values, the incentive to underdeclare is reduced and that the IMT generates information on transaction values that are the same (or very close) to real market transaction values.

[7] A proof of payment of the IMT has to be shown before sale. Typically, the payment of the IMT is done some days before the moment the transference of the property takes place.

[8] More on the way fiscal appraisals are carried out in Portugal can be found in DSA (2011).

[9] According to the last population and housing Census, Portugal had a stock of 5.8 million conventional dwellings for residential purposes. The difference of around 600 thousand dwellings between the Census and the IMI figures can be explained (at least partly) by the fact that Statistics Portugal did not receive fiscal appraisals that were carried out from the start of the IMI and IMT tax system (December 2003) and the date in which the first data transmission agreement started to produce effects (beginning of 2005).

The access to fiscal administrative information is supported by data transmission agreements, which have been signed between fiscal authorities and Statistics Portugal[10]. Although sometimes sent days delayed in relation to agreed data transmission dates, IMT and IMI information is available regularly within a time period that allows the compilation of a quarterly price index in a timely manner[11]. The development of the transactions-based HPI, benefited from the existence of a good institutional and technical collaboration between the tax authorities and Statistics Portugal.

In order to use the IMT and IMI data, the raw information provided by the fiscal authorities had to have suffered several transformations and restrictions. First, it was necessary to interpret and convert the flow of IMT and IMI information into a list of meaningful variables which could be used in posterior statistical analysis[12]. Second, as the index aims at measuring price inflation of dwelling purchases for residential purposes, information concerning transactions of agricultural land, commercial dwellings, parking facilities and plots of land (for later construction) was excluded from HPI compilation. Third, as the index is restricted to transactions carried out in a market context, all dwelling sales signalled to have been carried out with a zero price, were excluded[13]. However, barter-like deals, which involve the exchange of dwellings (*contratos de permuta*), were, as long as the purchasing price was not equal to zero, not ruled out from subsequent HPI calculation phases[14]. Finally, transactions of parts of dwellings were also excluded. Thus, only the dwellings that were transacted as a whole were considered for later stages of the production of the HPI[15].

For the compilation of the transactions-based HPI, the above-mentioned data sources were merged using a property (cadastral) register identification number, which is used as the key variable for the match of prices and characteristics. The match of the IMT and IMI is carried out on a monthly basis. As a rule, the production of a HPI whose reference quarter ends in month *m* makes use of the information that was transmitted (and matched) in months *m+1* and *m+2*. Around 86% of all dwelling transactions are matched and available for the compilation of the quarterly HPI[16].

[10] A first agreement, which covered the annual transmission of IMI data, was signed at the beginning of 2005. A first amendment to this initial legal document, encompassing the monthly transmission of IMT data, was signed in the last quarter of 2007. A second and final amendment, allowing the monthly transmission of IMI data, was signed between the Portuguese fiscal authorities and Statistics Portugal in the first quarter of 2012. The legal framework supporting the compilation of the transactions-based HPI is essentially based upon these two amendments to the initial data transmission agreement.

[11] The irregularity to which the transmission of IMT and IMI data is presently subject precludes, at least for the time being, the production of a monthly HPI.

[12] Annex 1 provides examples of these transformations.

[13] The exclusion of zero-priced dwelling transactions resulted in an elimination of around 9,300 observations from the original dataset (slightly more than 2% of all transactions included in the IMT/IMI dataset). These cases can result from non-harm's length transactions (i.e. situations in which there are a relationship or a business affiliation between the seller and the buyer).

[14] These transactions were identified by a dummy variable (DBARTERTRANS).

[15] These transactions, which include transactions of two-equal parts of the same dwelling on the same day (which is a typical situation involving the acquisition of a home by a couple), represent slightly more than 90% of all available data.

[16] This percentage covers only those transactions that have been carried out for residential housing purposes and in which the dwellings are classified (by the IMI) as being of a residential nature. The percentage of matched transactions increases to 90-95% if no restriction is imposed on the IMI side (i.e., if dwellings classified as, for instance, commercial are merged with the transactions data extracted from IMT records). However, in order to avoid potential "noise", it was chosen to impose a restriction on IMI

The end product of this matching and data cleaning process was a unique dataset, which is fed on a monthly basis, with information characterizing both the prices and attributes of residential property transactions in Portugal (from 1Q2009 to 4Q2013, the number of transactions available in the database exceeded the 455 thousand (an average of more than 22.7 thousand observations per quarter).

2.2.Statistical description of fiscal administrative data

The table below presents the summary statistics for transaction value (TRANSVAL) and gross floor area (GRFLOORA). As in table 1, the information is presented with a division between new and existing dwellings and between apartments and houses.

Table 2: Descriptive statistics of available data

		N	*Mean TRANSVAL*	*Median TRANSVAL*	*Stdev TRANSVAL*	*Mean GRFLOORA*	*Median GRFLOORA*	*Stdev GRFLOORA*
Existing	*Apartments*	212,863	95,701	78,000	76,635	96	92	37
	Houses	94,134	96,111	60,000	168,966	128	107	91
New	*Apartments*	110,085	153,411	129,000	120,380	115	111	41
	Houses	38,709	157,442	135,000	175,645	164	163	77

Notes: *TRANSVAL in €; GRFLOORA in m2.*

As seen from above, existing properties represent two thirds of the transactions and apartments slightly more than 67% of all available data. As expected, median values for TRANSVAL and GRFLOORA are lower than mean values, which is a typical outcome in positively skewed distributions[17]. The observed mean prices are also in accordance with what one would expect from a dataset containing prices and characteristics of dwellings (e.g., new houses show the highest price and area mean values and existing apartments the lowest). Finally, it is interesting to note that the dispersion of prices and areas is higher for houses than for apartments (thus suggesting the existence of a higher heterogeneity in the former dwelling category).

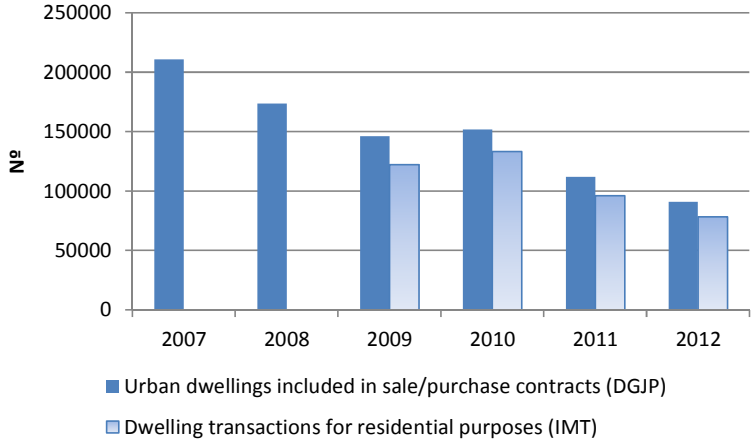
The coverage of the universe of residential property transactions by the dataset is deemed to be very high. Although there is not a straightforward way of assessing its representativeness, it can nevertheless be inferred through a comparison between the counts of dwelling transactions, which are obtained using the IMT data, and the number of urban properties included in purchase and sell contracts (*Contratos de Compra e Venda*), which are compiled by the Directorate-General for Justice Policy of the Ministry of Justice (DGJP). The next chart depicts these two sources of information for the 2007 – 2012 period[18].

data and to consider only those transactions whose dwellings are defined by the IMI as being of a residential type.

[17] The same happens with the data depicted in table 1.

[18] There are DGJP data available from 2007 onwards. As mentioned in the previous section, IMT data is only available from 2009 onwards.

Chart 1.: Dwelling transactions and urban properties in contracts



There are two issues that are worthwhile to point out. First, it should be mentioned that the number of transacted dwellings for residential purposes, as measured by the IMT data, represents nearly 86% of the number of urban properties included in purchase and sell contracts over the 2009-2012 period[19]. Second, it is interesting to note that DGJP and IMT data reveal the same annual evolution (i.e., both show an increase in contracts and transactions from 2009 to 2010 and a decrease in these figures from that point onwards).

2.3.Compilation methodology

The transactions-based HPI is a hedonic price index that is derived from the application of the adjacent time dummy approach[20]. For all $q=(Q-1,Q)$ and $i=1,...,n_q$, this approach can be described as follows:

$$\eta_{i,q} = a + \sum_{K=1}^{K} \beta_k X_{i,q;k} + \theta D_{i,q} + \varepsilon_{i,q} \qquad (2.1)$$

where,

[19] It should be borne in mind that DGJP figures include not only residential dwellings but also other type of urban property transactions (e.g. garages, industrial facilities, non-onerous transactions).

[20] A hedonic price index is simply an index that uses information taken from (what was dubbed as) a hedonic function (Triplett, 2006). The hypothesis underlying the building of hedonic functions is based on the reasoning that goods, such as houses, are purchased as bundles of characteristics and that price differences, found among similar varieties of the same good, are simply explained by the different quantities of quality attributes characterizing each one of these varieties (Rosen, 1974). The aspects underlying the econometric estimation o hedonic functions have been the focus of many studies. A classical reference is Berndt (1991; chapter 4) and a more recent reference is Triplett (2006). This method has been used in the construction of constant-quality official real estate price statistics since the 1960's (e.g., Musgrave, 1969) and its use is recommended by international technical manuals (see, for instance, Eurostat, 2011).

$\eta_{i,q}$, is the price level (or some transformation of it) of the *ith* dwelling transaction in quarter *q*;

$X_{i,q}$, stands for the value of the *kth* characteristic of the *ith* transacted dwelling in quarter *q*;

D_q, is a temporal indicator, which is defined as:

$$\text{for all } q=(Q-1,Q) \text{ and all } i=1,...,n_q, \quad D_{i,q}=\begin{cases}1, \text{ if } q=Q, \text{ and}\\ 0, \text{ if otherwise.}\end{cases}$$

θ, is the parameter associated to the temporal indicator D_q; and

$\varepsilon_{i,q}$, corresponds to an error term.

As noted elsewhere[21], when the dependent variable (η) is expressed as the natural logarithm of price levels ($\eta = \ln(Y)$), then the value of the price index in quarter Q is obtained by the exponentiation of θ[22].

The adjacent time dummy method is presently used in the production of the French HPI for new dwellings (INSEE, 2013)[23]. Koev (2003), INE (2009) and ISTAT (2013) constitute other examples of the application of the hedonic method on administrative data for the compilation of official HPI figures.

2.4. Model specification, estimator, estimation process and diagnostic checks

As noted in the literature[24], economic theory sheds little light as to the correct specification of the hedonic prices-characteristics relationship for housing markets. In this context, the choice of the functional form is not insensitive to the data at hand and empirical studies and critical reviews provide valuable additional guidance to price index compilers[25].

The choice of the regressand rested on the natural logarithm of the transaction value (LNTRANSVAL)[26].

In relation to the model's covariates, its choice involved the categorisation of available IMT/IMI information into main groups of price-determining attributes and

[21] See, for instance, Halvorsen and Palmquist (1980).

[22] For the estimation of $\exp(\theta)$, Kennedy (1981) proposes the use of $\exp\left(\hat{\theta}-0.5\hat{\delta}^2\right)$, where $\hat{\delta}^2$ is an estimate of the variance of $\hat{\theta}$, instead of $\exp\left(\hat{\theta}\right)$. The results given by the two estimators are not expected to differ much in large samples and Kennedy's correction is usually disregarded in hedonic time dummy studies. This situation was confirmed in the present study and $\exp\left(\hat{\theta}\right)$ is used to estimate $\exp(\theta)$.

[23] For existing dwellings, INSEE applies the hedonic repricing method. See Gouriéroux and Laferrère (2009).

[24] See, *inter alia*, Butler (1982) and Cropper et al. (1988).

[25] A good review focusing on the use of the hedonic house price model in housing is given in Chin and Chau (2003).

[26] Other variables, such as the transaction price level, were also tried. However, the models and tests that were carried out supported the use of LNTRANSVAL in regression analysis.

was mainly driven by empirical testing. Following this idea, available variables were grouped into five main categories:

- Area / dimension variables;
- Age / time variables;
- Location variables;
- Quality/comfort variables; and
- Variables characterising the type of transaction.

In the specification process, special attention was given to the inclusion of location, area and age effects in the models. This had essentially to do with the recognition that, on the one hand, the models needed to incorporate both macro and micro location attributes and that, on the other, the correct specification of area and age effects could be complex[27].

The transactions-based HPIs are estimated by ordinary least squares (OLS) using the adjacent time dummy method. Thus, for each pair of contiguous quarters, four regressions covering the following basic strata, are run[28]:

- Existing apartments;
- Existing houses;
- New apartments; and
- New houses.

In order to investigate the influence of outliers in OLS results, least square absolute deviation (LAD) estimates were also computed.

Robust statistics, testing individual and joint significance of parameters, were applied in the specification and estimation process[29]. Moreover, the Variance Inflation Factor (VIF) was applied to detect excess of collinearity and a (robust) version of the Reset (Ramsey, 1969) was used to detect possible functional form problems.

A list and description the variables that were included in the models are available in annexes 2 and 3.

[27] The importance of the various location levels in hedonic models is expressed in Kiel and Zabel (2008). Goodman and Thibodeau (1995) provide an example of a study on the relationship between age and market price of properties.

[28] The separation between new and existing dwellings was directly imposed by Commission Regulation (EC) No 93/2013, which requires the provision of HPIs for these two subdivisions. The use of type of dwelling as a further stratification variable was based on the analysis of the data, which suggested that the apartments and house market segments should be dealt separately.

[29] The presence of heteroskedasticity was checked through (a robust version of) the Breush-Pagan test (Wooldridge, 2003). The homoskedasticity assumption was rejected in all the tests.

2.5. Outliers and influential observations outliers

A number of criteria are applied before the data enters in regression analysis. These included, in particular, the following general rules that exclude the transactions that:

- are transacted more than twice on the same day;
- are transacted at a price that is below or equal to 5,000 €;
- have a price per square meter that is higher than 9,750€;
- have a gross floor area that exceeds the 800 square meters;
- have a plot area that exceeds the 3,000 square meters;
- have a number of rooms that is greather or equal to 16.

As a result of the application of the criteria, 5.1% of the data is not included in regression analysis.

Unusual observations are also spotted during regression analysis. In particular, the observations that have large residuals (i.e. studentized residuals that lie outside the [-3,3] interval) and that are simultaneously influential (i.e., that, in addition to large residuals, produce a high leverage in the regression) are droped out from final regressions. The number of "influential outliers" found in regression analysis does not exceed the 1% the data.

2.6. Regression results

This section provides the results of the hedonic regressions that were used to derive the hedonic price index. For the sake of space, the analysis shown in this section refers to the regression output that was extracted using data from the first and second quarters of 2010 for existing apartments, a stratum which represents around 40% of the HPI weight[30].

For this stratum, the existing apartment that has been chosen to serve as the base model is characterised as follows:

- existing apartment, which has not been traded in a barter-like transaction, with no central heating system, with no access to a swimming pool or parking facilities, located outside a private condominium, with no negatively appraised quality and comfort characteristic, located outside the capital district area, belonging to the Lisbon (metropolitan) area, in a parish with no front coast, with no visual prominence to offer and with no particular good micro-location and access to public goods/amenities.

The results for this regression output are illustrated in the next table. The table includes, together with parameter point estimates, p-, VIF and other statistics, a column with the *a priory* expected sign of the coefficients.

[30] The main conclusions presented in this section can be extended to other index strata.

Table 3: Regression results for *Existing Apartments* (OLS estimator)

(Dependent variable; natural logarithm of transaction price)

Parameter	*Expected Sign*	*Point Estimate*	*Heteroskedasticity Robust Statistics*			*VIF*
			Stdev	*t-value*	*P>|t|*	
Intercept	.	9.984	0.05	202.3	<.0001	0.0
DBARTERTRANS	*(-)*	-0.464	0.02	-24.8	<.0001	1.0
GRFLOORA	*(.)*[(a)]	-0.001	0.00	-1.8	0.079	49.5
SQROOTGRFLOORA	*(.)*[(a)]	0.164	0.01	16.9	<.0001	50.1
DEPFLOORA	*(..)*[(b)]	0.001	0.00	6.4	<.0001	6.2
SQROOTDEPFLOORA	*(..)*[(b)]	0.009	0.00	4.2	<.0001	7.7
DWELLTRANSAGE	*(...)*[(c)]	-0.010	0.00	-30.9	<.0001	5.9
SQRDWELLTRANSA	*(...)*[(c)]	0.0001	0.00	15.4	<.0001	4.4
DCSYSTEM	*(+)*	0.154	0.01	22.5	<.0001	1.2
DSWIMM	*(+)*	0.155	0.01	13.4	<.0001	1.5
DPARKING	*(+)*	0.067	0.01	12.8	<.0001	2.1
DPRIVCOND	*(+)*	0.089	0.01	6.2	<.0001	1.4
DNEG	*(-)*	-0.029	0.01	-4.1	<.0001	1.2
DDISTRCAP	*(+)*	0.200	0.01	39.5	<.0001	1.4
DREGION1	*(-)*	-0.515	0.01	-82.0	<.0001	1.4
DREGION2	*(-)*	-0.285	0.01	-53.4	<.0001	1.4
DREGION3	*(-)*	-0.384	0.01	-66.7	<.0001	1.3
DREGION5	*(-)*	-0.178	0.01	-13.3	<.0001	1.1
DREGION6	*(-)*	-0.047	0.01	-5.7	<.0001	1.3
DREGION7	*(-)*	-0.077	0.02	-4.8	<.0001	1.2
DSEA	*(+)*	0.100	0.00	22.7	<.0001	1.3
DSCENIC2A	*(+)*	0.082	0.01	9.1	<.0001	1.1
DSCENIC3A	*(++)*[(d)]	0.277	0.03	10.3	<.0001	1.0
DINTERIORLOC3	*(+)*	0.064	0.01	8.7	<.0001	1.1
DBADLOC	*(-)*	-0.131	0.06	-2.3	0.022	1.0
DEXCPLOC	*(+)*	0.403	0.01	60.0	<.0001	1.5
DQRT6	*(+, - or 0)*[(e)]	0.007	0.00	1.9	0.053	1.0

Notes:

n = 27,195
Adj. R-Sqr = 0.67

[(a), (b), (c)] No a priory sign is expected for the individual regressors as the impact of the area and age variables cannot be isolated from their square root and square transformations.

[(d)] Expected a coefficient value higher than the one estimated for DSCENIC2A.

[(e)] The exponential of this parameter reflects the quarter-to-previous quarter price change.

As the table shows, the signs of the coefficients are in accordance with expected results. Moreover, with the exception of GRFLOORA and the time dummy (DQRT6) variables, all regressors included in the model are significant at the 5% significance level[31].

The latter result translates the fact that the regression does not capture significant ("pure") price change from one quarter to the other and, as a consequence, the time variable is statistically rejected[32]. Moreover, it is interesting to note that, while the GRFLOORA variable has no individual significance, GRFLOORA and its square root transformation (SQROOTGRFLOORA) are jointly significant at the 5% significance level[33]. Regression outputs show that regression coefficients are fairly stable in all regressions.

Using the information available in table 3, it is possible to estimate the evolution from the first to the second quarter of 2010 by:

$$\exp\left(\hat{\theta}\right) = \exp\left(0.007\right) = 1.007 \qquad (3.1)$$

The result in (3.1) says that the prices of existing apartments have increased 0.7% from the first to the second quarter of 2010.

The next two charts provide a comparison between the unadjusted price index (solid blue line) and the quality adjusted HPI (dashed red line) for the existing apartments stratum. For comparison purposes, the results obtained through the use of the LAD estimator are also displayed (black solid line with a square as a marker)[34]. The comparison is made over the whole period of the series for index number values and year-on-year change rates.

[31] In order to investigate if the significance of chosen variables was influenced by the number of observations used in the regressions, 30 samples of 2,800 observations (roughly 10% of the total) were drawn at random and used in regression analysis. By and large, obtained results suggest that the significance of chosen variables does not change when the dimension of the sample drops from 28,000 to 2,800 observations.

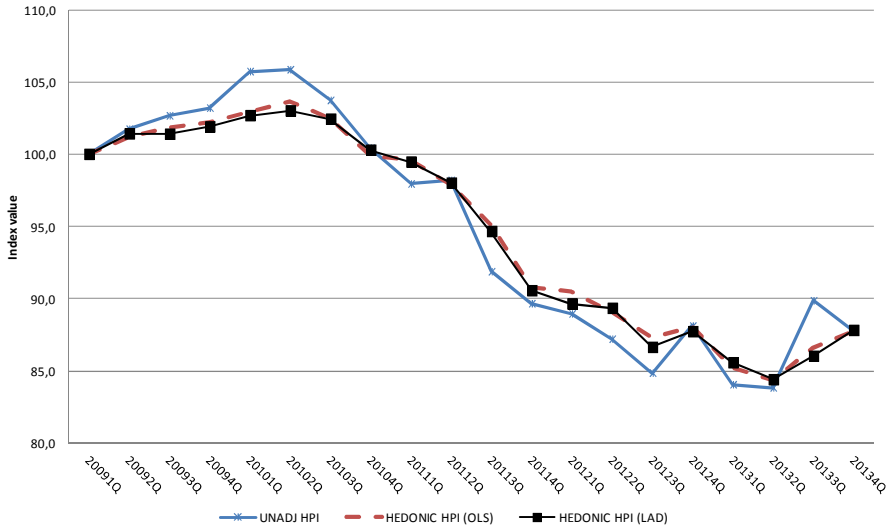
[32] Although the time variable is often rejected when only two periods are considered, its importance was stressed in tests that included more than two quarters in the regression and that rejected the joint null hypothesis that all of the time variables were equal to zero (i.e. that prices were only to be explained by quality change).

[33] The joint significance of these two variables was never rejected in adjacent time dummy regressions.

[34] Both the unadjusted and quality adjusted HPIs are based on the same dataset (i.e., the dataset that is obtained after the application of the criteria mentioned in section 2.5). The unadjusted index is based on the geometric mean formula. The LAD uses all data available (i.e., it does not applies the criteria mentioned in section 2.5.).

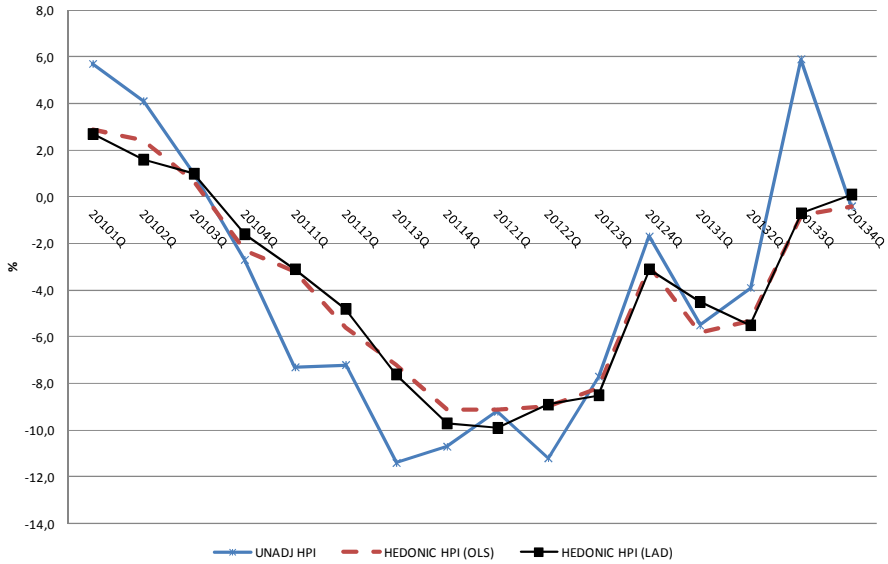
Chats 2 and 3: Unadjusted and hedonically adjusted HPIs for *Existing Apartments*:

Chart 2: Index numbers
(100 = 1Q2009)



	Unadjusted HPI	Hedonic HPI	Hedonic (LAD)
Mean	94.8	94.8	94.7
Stdev	7.82	6.96	6.97

Chart 3: Year-on-year changes
(1Q2010 - 4Q2013)



	Unadjusted HPI	Hedonic HPI	Hedonic (LAD)
Mean	-3.89%	-3.94%	-3.91%
Stdev	5.88 p.p.	4.09 p.p.	4.19 p.p.

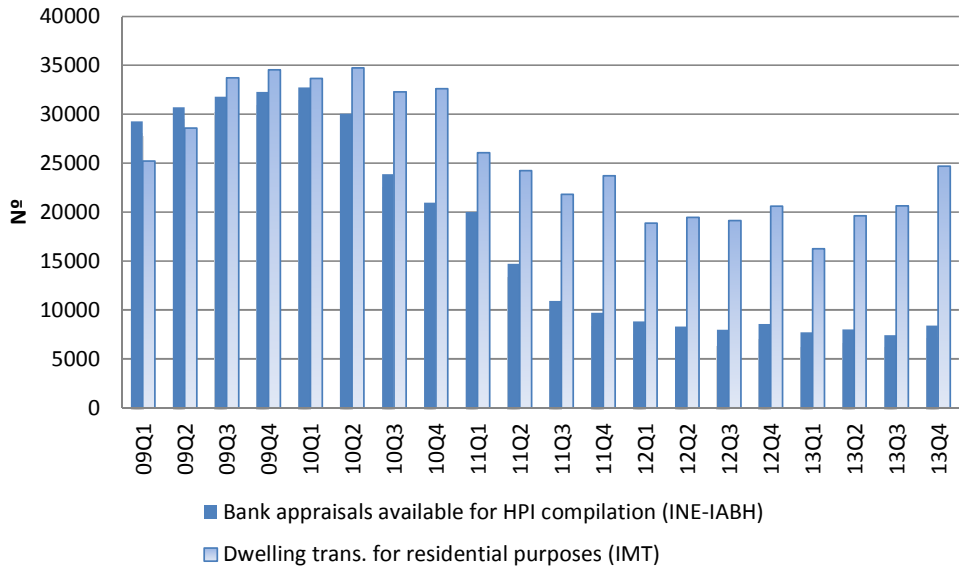
As the information provided in the previous page shows, the unadjusted HPI presents a higher volatility than the hedonically adjusted HPIs. This should come as no surprise, as part of the volatility of the unadjusted price index is to be attributable to changes in the quality mix of transacted dwellings. As the charts illustrate, the behaviour of the quality adjusted price index is similar to its LAD counterpart, which suggests that, at least for the period covered, the OLS estimator is not affected by the presence of outlying data points.

3. Presentation and comparison of results

3.1. Number of transacted dwellings

In order to shed some light on the overall coherence of the data that is used in the compilation of the appraisals- and transactions-based HPIs, it is worthwhile to make a comparison between the number of bank appraisals and the number of observations available in the IMT/IMI dataset.

Chart 4.: Number of bank appraisals and dwelling transactions

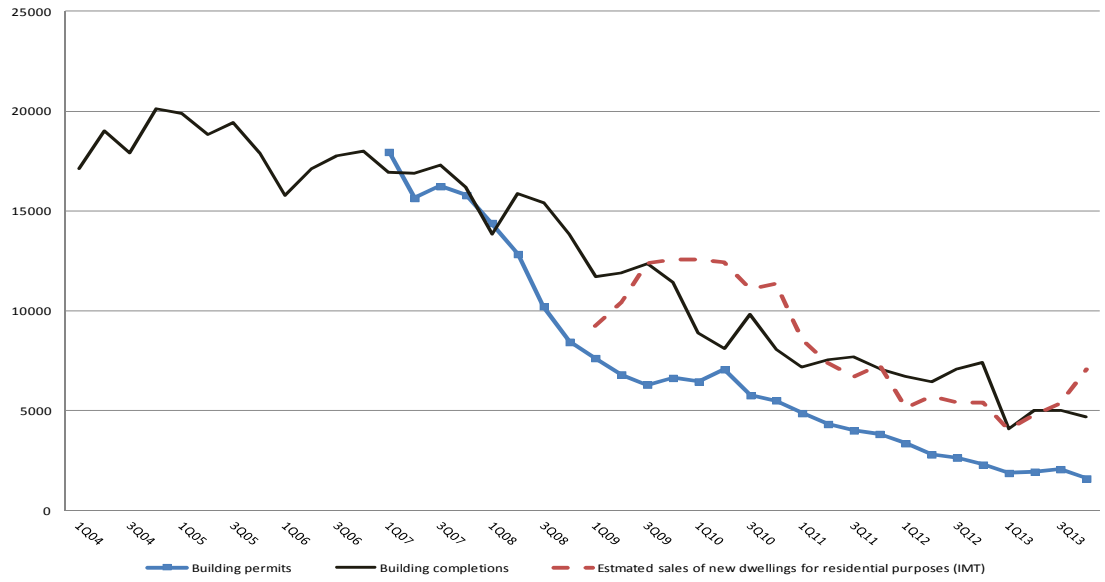


As it is possible to see from the chart displayed above, the number of bank appraisals only outscores those of transactions in the first two quarters of 2009 (when mortgage credit was more abundant).

In the 2Q2010 – 3Q2012 period, bank appraisals drop considerably and, with the exception of two quarters, at a faster rate than transactions. From 4Q2011 onwards, the number of bank appraisals accounts for less than half of the number of recorded dwelling transactions (in the last quarter of the series, the latter represents around one third of the former). The chart illustrates an increase in the counts of bank appraisals and dwelling transactions (this is much more evident in the latter than in the former) data, thus suggesting that the housing market is recovering in 2013.

Another way of analysing the coherence of available data is to make a comparison between the number of building permits (and the number of building completions) and the number of new dwellings purchased, which have been estimated using the information contained in the IMT/IMI dataset. The next chart presents the three series for the 1Q2004-4Q2013 period.

Chart 5.: Building permits, completions and estimated sales of new dwellings



The decrease tendencies depicted by the building permits and completions counts suggest that the housing market had been adjusting the quantity of new dwellings on offer since 2007. In addition, and as it can be seen from the above chart, the three series are consistent and coherent with one another.

Overall, it can be concluded that, based on available information, the datasets that are used in the compilation of the appraisals- and transactions-based HPIs are coherent and in accordance with other statistics on the Portuguese housing market.

3.2. House price indexes

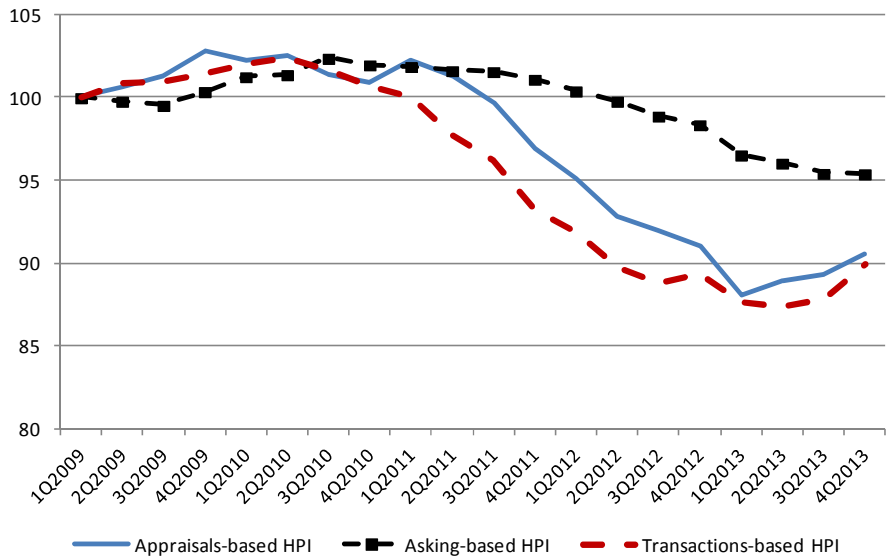
This section compares the results provided by the appraisals-based, transactions-based and the index based on asking prices. The latter index, which is compiled by a private company, is based from the sellers' asking prices of dwellings that are advertised on a real estate portal available on the Internet [35]. There is no information available on the geographical coverage of the portal and on how close advertised asking prices are in relation to final transaction prices. The index is compiled monthly and follows a distinct methodology than the one that is used for the HPI. More information on the methodology of this index is found in CI (2006).

The next page presents the information on these three indexes.

[35] www.lardocelar.com.

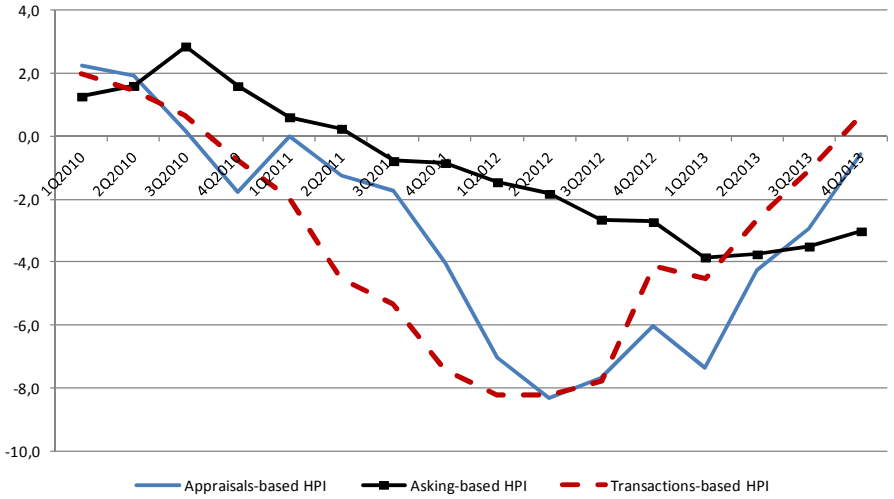
Chants 6 and 7: Asking-based, Appraisals-based and Transactions-based HPI:

Chart 6: Index numbers
(100 = 1Q2009)



	Transactions	Appraisals	Asking prices
Mean	95.47	96.98	99.67
Stdev	5.84	5.37	2.24

Chart 7: Year-on-year changes
(1Q2010 - 4Q2013)



	Transactions	Appraisals	Asking prices
Mean	-3.23%	-3.04%	-1.01%
Stdev	3.23 p.p.	3.46 p.p.	2.24 p.p.

There are a few issues that are worthwhile to point out from the comparison of the three indexes.

The first is that, despite the sharp drop in the number of bank appraisals (see section 3.1), appraisals-based HPI seem to mimic the evolution of the transactions-based indicator reasonably well. Indeed, it is interesting to note that both the appraisals and transactions-based indexes signal the same turning point in 2Q2012 and that, in addition, the correlation between the former and the latter indicators is stronger than the one found between the transactions-based index and the HPI based on asking prices.

The second refers to the fact that the asking prices HPI seems to lag behind both appraisals and transactions-based HPIs. For the 1Q2010-4Q2013 period, the contemporary correlation of the asking-based price index and the appraisals- and transactions-based HPIs are 0.69 and 0.43, respectively. These figures increase to 0.86 and 0.65 when HPI year-on-year rates of change of quarter t are compared with the asking-based price index rates of t+1. These results suggest the existence of a lag in the asking-based HPI.

In addition, the asking-based HPI shows, when compared with the remaining HPIs, a different behaviour in the intensity of the price change in growth and reduction periods. In particular, it is possible to see that the index based on asking prices observes lower price decreases in periods where the market is contracting.

These results should perhaps come as no surprise, as the index that is based on asking prices is best representative of the beginning of the buying and selling process, where prices tend to be, especially in contraction periods, advertised at higher levels than actual transaction prices.

Conclusions and final remarks

The need for the production of new house price indicators has provided the opportunity to compare indicators based on prices information collected at different stages of the buying and selling housing process.

The chosen development strategy, which was based on the exploration of administrative data rather than on the design of new and costly surveys, has proven effective in providing indicators that are able, on the one hand, to meet European Union legal requirements on house price indexes and, on the other, simultaneously provide a more complete picture of the real estate market in Portugal and satisfy the need for higher level policy indicators for macroeconomic surveillance.

Obtained results suggest that, for the period under analysis (i.e., 1Q2009-4Q2013), the HPI based on asking prices lags behind the appraisals and transactions-based HPIs. Another important point to note is the idea that, despite the sharp decrease in bank appraisals, the appraisals-based HPI seems to reflect, in a reasonable way, its transactions-based counterpart. However, more research would be needed to be carried out in order to understand better the differences shown among the different HPIs.

Overall, the work that has been developed reinforces the idea that the HPI based on fiscal administrative data should represent a qualitative improvement in relation to already produced appraisals-based HPI. This stems not only from the use of transaction values (instead of bank appraisals) and from the quality of fiscal data (e.g., it covers the complete universe of transactions), but also from the application of more appropriate methods for the treatment of the changing quality of transacted dwellings.

Annex1: Transforming original data into variables for regression analysis

This annex provides two examples of the data transformations that were done in order to incorporate IMT and IMT data into regression analysis.

A simple example of these transformations is given by the variable DLIFT, which assumes the value 1 when the dwelling has an elevator and zero otherwise. This dummy variable is the combination of the following two variable attributes, which are available in the data:

- Maj10: existence of an elevator in a dwelling with less than four floors; and
- Min19: inexistence of an elevator when the dwelling has more than three floors.

The existence of this discrimination by number of floors has to do with changes in the building code in Portugal (by law, at present, it is not possible to build a dwelling without elevator if the dwelling being built has more than three floors; however, in the past this was possible).

The combination of these attributes gave origin to DLIFT. This is simply done applying the following algorithm:

- IF Maj10 = “Yes” THEN DLIFT = 1; ELSE IF Maj10 = “No” THEN DLIFT = “0”;
- IF Min19 = “Yes” THEN DLIFT = 0; ELSE IF Min19 = “No” THEN DLIFT = “1”.

Another example is given by DBADLOC and DEXCPLOC, taken from the location coefficient variable (*Coeficiente de Localização*), which is used in the appraisal of the value of dwellings. This coefficient measures the quality of public goods (roads, proximity to schools, hospitals, public transports, etc) and varies from 0.35 to 3. As the formula used in the appraisal of dwelling values is of a multiplicative form, the value 1 refers to a “neutral” situation, where location is neither bad nor good. In order to grasp extremely bad and good locations, it was chosen to build the above-mentioned variables using the following basic algorithm:

- IF “location coefficient” <= 0.4 THEN DBADLOC = 1; ELSE DBADLOC = 0; and
- IF “location coefficient” >= 2 THEN DEXCPLOC = 1; ELSE DEXCPLOC = 0.

Annex 2: List of variables used in regression analysis

	Variable name	Description	Source of the variable / Comments
(1) Value of the transaction			
1	LNTRANSVAL	Natural logarithm of the Transaction value	IMT; Regressions' dependent variable
(2) Area/Dimension			
2	GRFLOORA	Gross floor area	IMI
3	SQROOTGRFLOORA	Square root of the Gross floor area	Transformation of (2)
4	DEPFLOORA	Dependent floor area	IMI
5	SQROOTDEPFLOORA	Square root of the Dependent floor area	Transformation of (4)
6	PLOTAREA	Plot area (free land)	IMI
7	SQROOTPLOTAREA	Square root of the Plot area	Transformation of (6)
(3) Age/Time			
8	DWELLTRANSAGE	Estimated dwelling age at date of transaction	-
9	SQRDWELLTRANSA	Square of the dwelling age at date of transaction	Transformation of (6)
10	DBADCONSTATE	Dummy variable assuming the value 1 whenever "bad conservation state of the dwelling" is signalled; 0 otherwise	IMI
11	DQRT*i*	Dummy variable identifying the *ith* quarter in which the transaction was carried out (*i* = 1, …, 19)	IMT; 1 = 1Q2009, ..., 20= 4Q2013
(4) Location			
12	DREGION*i*	Dummy variable identifying the *ith* region (i=1, …, 7) Norte, Área Metr. Porto, Centro, Área Metr. Lx, Alentejo, Algarve, Ilhas)	Stats Portugal
13	DDISTRCAP	Dummy variable assuming the value 1 if the transaction is located in a capital of district; 0 otherwise	Stats Portugal
14	DSEA	Dummy variable assuming the value 1 when the parish in which the dwelling is located has front coast 0 otherwise	Stats Portugal

	Variable name	Description	Source of the variable / Comments
15	DEXCPLOC	Dummy variable assuming the value 1 when the market value of location / quality of public goods/amenities is good; 0 otherwise.	From IMI's parameter *Coeficiente de Localização*
16	DBADLOC	Dummy variable assuming the value 1 when the market value of location / the quality of public goods/amenities is; 0 otherwise.	From IMI's parameter *Coeficiente de Localização*
17	DINTERIORLOC3	Dummy variable assuming the value 1 when the quality of the localization of the appraised dwelling is better than other dwellings of the same building	From IMI's parameter *Localização e operacionalidade relativas*
18	DSCENIC*i*	Dummy variable identifying the *ith* category of the quality of the visual prominence of the location to which the dwelling has access (*i* = 2A , 3A)	From IMI's parameter *Localização excepcional*
(5) Quality/comfort of the dwelling			
19	DPARKING	Dummy variable assuming the value 1 when the dwelling includes parking facilities; 0 otherwise.	From IMI's parameter *Garagem individual* and *Garagem coletiva*
20	DLIFT	Dummy variable assuming the value 1 where the dwelling has a lift; 0 otherwise.	From IMI's parameters *Elevadores em edifícios de menos de 4 pisos* and *Inexistência de elevador em edifícios com mais de 3 pisos*
21	DSWIMM	Dummy variable assuming the value 1 when the dwelling has access to swimming pools; 0 otherwise.	From IMI's parameter *Piscina individual* and *Piscina coletiva*
22	DPRIVCOND	Dummy variable assuming the value 1 when the dwelling is in a private condominium; 0 otherwise.	From IMI's parameter *Localização em condominio fechado*
23	DCSYSTEM	Dummy variable assuming the value 1 when the dwelling includes central heating and/or air-conditioning systems; 0 otherwise.	From IMI's parameter *Sistema central de climatização*
24	DNEG	Dummy variable assuming the value 1 whenever there is at least one negative coefficient of quality and comfort in the fiscal appraisal of the dwelling	IMI
25	DCONSTRQUALITY	Dummy variable assuming the value 1 whenever "the quality of construction of the dwelling" is signalled; 0 otherwise	IMI
(6) Type of transaction			
26	DBARTERTRANS	Dummy variable assuming the value 1 when the transaction is tagged as a *permuta*, 0 otherwise	IMT

Annex 3: Used specifications in regression analysis

	HPI strata/specification			
Explanatory variable [(*)]	***Existing Apartments***	***Existing Houses***	***New Apartments***	***New Houses***
DBARTERTRANS	x	x	x	x
GRFLOORA	x	x	x	x
SQROOTGRFLOORA	x	x	x	x
DEPFLOORA	x		x	
SQROOTDEPFLOORA	x		x	
PLOTAREA		x		x
SQROOTPLOTAREA		x		x
DWELLTRANSAGE	x	x	x	x
SQRDWELLTRANSA	x	x		
DCSYSTEM	x		x	
DBADCONSTATE		x		
DCONSTRQUALITY		x		x
DLIFT			x	
DSWIMM	x	x	x	x
DPARKING	x	x	x	x
DPRIVCOND	x	x	x	x
DNEG	x	x	x	x
DDISTRCAP	x	x	x	x
DREGION1	x	x	x	x
DREGION2	x	x	x	x
DREGION3	x	x	x	x
DREGION5	x	x	x	x
DREGION6	x	x	x	x
DREGION7	x	x	x	x
DSEA	x	x	x	x
DSCENIC2A	x	x	x	x
DSCENIC3A	x	x	x	x
DINTERIORLOC3	x		x	
DBADLOC	x	x	x	x
DEXCPLOC	X	x	x	x

Notes:
Dependent variable = LNTRANSVAL
[(*)] included variables are signalled by an "x".

References

- Berndt. E. R. (1991). *The practice of econometrics: Classic and contemporary*. Addison-Wesley Publishing Company. Reading (MA)

- Butler. R. (1982). The Specification of Hedonic Indexes for Urban Housing. *Land Economics*. 58(1). 96-108

- Confidencial Imobiliário (2006). *Índice Confidencial Imobiliário: Metodologia (Base 2005)*. available (in Portuguese only) on http://confidencialimobiliario.com/sites/default/files/Metodologia_Indice_CI.pdf

- Chin, T. L. and Chau, K. W. (2003). A Critical Review of Literature on the Hedonic Price Model, *International Journal for Housing and Its Applications*, 27(2), 145-65

- Cropper. M.. Deck. L. and McConnell. K. (1988). On the Choice of Functional Form for Hedonic Price Functions. *Review of Economics and Statistics*. 70(4). 668-75

- Direcção de Serviços de Avaliações (2011). *Manual de Avaliação de Prédios Urbanos*. Versão 5.0. Direcção Geral dos Impostos. Lisboa

- European Commission (2011). *Scoreboard for the Surveillance of macroeconomic Imbalances: Envisaged Initial Design*. Commission Staff Working Report. Available on http://epp.eurostat.ec.europa.eu/portal/page/portal/macroeconomic_imbalance_procedure/methodology

- Eurostat (2011). *Residential Property Price Indices Handbook*. Available from http://epp.eurostat.ec.europa.eu

- Goodman and Thibodeau (1995). Age-Related Heteroskedasticity in hedonic House Price Equations. *Journal of Housing Research*. 6(1). 25-42

- Gouriéroux. C. and Laferrère. A. (2009). Managing Hedonic housing price indexes: the French experience. *Journal of Housing Economics*. 18(3). 206–13

- Halvorsen. R. and Palmquist. R. (1980). The Interpretation of Dummy Variables in Semilogarithmic Equations. *American Economic Review*. 70(3). 474-5

- INE (2009). *Housing Price Index: Methodology*. Available on http://www.ine.es/en/daco/daco42/ipv/metodologia_en.pdf

- INSEE (2013). Note méthodologique: Indice de prix des logements. Available on http://www.insee.fr/fr/indicateurs/ind112/Logement-neufs_m.pdf

- ISTAT (2012). *House Price Index: Methodological note*. Available through the following page http://www.istat.it/en/archive/100092

- Kennedy. P. E. (1981). Estimation with Correctly Interpreted Dummy Variables in Semilogarithmitic Equations. *American Economic Review*. 71(4). 801

- Kiel. K. And Zabel. J. (2008). Location. location. location: The 3L Approach to house price determination. *Journal of Housing Economics*. 17. 175–90

- Koev. E. (2003). *Combining Classification and Hedonic Quality Adjustment in Constructing a House Price Index.* Available on http://www.stat.fi/og2004/koevp.pdf

- Musgrave. J. C. (1969). The Measurement of Price Changes in Construction. *Journal of the American Statistical Association*. 64(327). 771-86

- Ramsey, J.B. (1969). Tests for Specification Errors in Classical Linear Least Squares Regression Analysis. *Journal of the Royal Statistical Society*. B 31. 350-71

- Rosen. S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation and Pure Competition. *Journal of Political Economy*.82. 34-55

- Triplett. J. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*. OECD

- Wooldridge. J. M. (2003). *Introductory Econometrics: A Modern Approach (2nd Ed.)*