An Phríomh-Oifig Staidrimh
Central Statistics Office

Technical Paper

RPPI Technical Paper

Introduction

Residential Property Price index (RPPI) is designed to measure the change in the average level of prices paid by households for residential properties sold in Ireland at market prices. Non-market purchases, purchases by non-household buyers and self-builds are specifically excluded. The index is mix-adjusted to allow for the fact that different types of property are sold in different time periods.

For the reference periods from 2005 to 2009 the index was compiled based on the mortgage data received from the main lenders on the market. From 2010 onwards the RPPI is based on Stamp Duty returns supplied by the Revenue Commissioners, and is accompanied by a comprehensive range of additional indicators of the residential property market.

Scope

The RPPI is compiled in accordance with Regulation (EU) 2016/792 of the European Parliament and of the Council of 11 May 2016 on harmonised indices of consumer prices and the house price index, and forms part of Ireland's international obligations to provide harmonised house price indices to the European Statistical System (ESS).

The RPPI covers all market purchases of houses and apartments by households, both cash and mortgage-funded. Non-market transactions, e.g. family transfers, and purchases by non-household buyers, e.g. by companies or institutions, are specifically excluded from the index as per the Regulation. Also excluded are self-builds, where the land is purchased separately, and purchases of partially built dwellings.

Data Sources

Stamp Duty data

The principal data source for the RPPI is stamp duty data provided to the CSO by the Revenue Commissioners. Under the Stamp Duties Consolidation Act (SDCA) 1999, all buyers of residential property have a legal obligation to notify the Revenue Commissioners, so that the transaction can be assessed for stamp duty. Since 2010, this notification takes place on-line, via Revenue's e-Stamping system and is typically carried out by the buyers' solicitors as part of the conveyancing process. Buyers are required by law to submit the transaction details to Revenue within 44 days of the transaction execution date.

The stamp duty data is a very rich data source in terms of providing characteristics of the buyers and sellers of property (see Revenue form SDR1 for details of the information collected). However, it is much more limited in terms of describing dwellings transacted, only providing the address and at times Eircode of the dwelling, whether it is new or existing, and the transaction price for stamp duty assessment purposes.

Building Energy Rating (BER) data

The second key data source is Building Energy Rating (BER) data. Under Statutory Instrument (S.I.) No. 243 of 2012, a BER certificate must be presented with all residential property advertised for sale (with some very

minor exceptions). This means that for all sales there should be a corresponding BER certificate available. As part of the BER assessment process, detailed information on the physical characteristics of the dwellings are collected, including the type of dwelling (detached, semi-detached, etc.) and the floor area of the dwelling. The address of the dwelling is also captured. These data are compiled by the Sustainable Energy Authority of Ireland (SEAI) and made available to the CSO on a quarterly basis.

GeoDirectory

The third data source used in the compilation of the RPPI is the GeoDirectory. This is a dataset containing information on all buildings in the State, created and maintained by GeoDirectory using data from An Post and Ordnance Survey Ireland (OSi). The GeoDirectory contains the Eircode and the postal address of every building in the State, along with other geographical information such as X-Y coordinates and Small Area codes. Small Areas are the smallest geographical classification used in statistics. They are nested within existing Electoral Division boundaries and typically contain 80-120 households. The Small Area codes allow mapping to census Small Area Population Statistics (SAPS) and related statistics.

Deprivation index

A final data source used by the RPPI is the Pobal Haase-Pratschke (HP) deprivation index. The Pobal HP Deprivation Index measures the relative affluence or disadvantage of each Small Area. A scoring is given to the area based on a national average of zero and ranging from approximately -35 (being the most disadvantaged) to +35 (being the most affluent). The index is derived from the demographic profile, social class composition and labour market situation of the population of each Small Area as measured in the latest Census of Population.

Processing the Stamp Duty data

Residential and non-residential

The first step in using Stamp Duty returns as a basis for a house price index is to separate transactions of residential dwellings from other real estate transactions. A wide variety of real estate transactions are liable for stamp duty, including commercial property and agricultural land sales. A stamp duty return is considered a relevant residential dwelling transaction if the property has an address in the State and the following criteria apply (see Table 1).

Table 1: Inclusion of Residential Dwellings based on Revenue form SDR1

Section	Particulars	Inclusions	Exclusions
3.1	Property Description	Residential	Non-Residential
			Mixed-Use
3.2	Instrument	Original return	Substitute return
			Proxy return
12.2.2	Type of Property	New Dwelling House/Apartment	Curtilage of Residential Property
		Second-Hand Dwelling House/Apartment	

Household and non-household purchases

There is no specific data field in the stamp duty data which indicates whether the purchaser is a household or a non-household. However, there are two data fields that capture the name of the purchaser, a first name field and an 'entity' name field (which captures either a surname or a company/institution name). The recording of a first name is taken as indicating a household purchase (i.e. one or more private individuals is involved in the purchase). The absence of a first name is taken as indicating a non-household purchase.

Market and non-market transactions

The stamp duty data contains a data field that nominally indicates whether the transaction is a market or non-market transaction (as well as a field for reporting the estimated market value of the property). However,

the market indicator field is not considered entirely reliable (for example, there are transactions between family members for relatively low prices which are declared market transactions). Instead, a transaction is deemed to be a market one if there is no declared relationship between the buyer and the seller, if the sales price is at least €25,000 and if certain other specific conditions are met (see Table 2).

Table 2: Identifying Non-Market Transactions based on Revenue form SDR1

Section	Particulars	Market	Non-Market
2.1	Instrument	Conveyance/Transfer of Property	Exchange of property
		Long Term Lease (over 100 years)	Short term lease (less than 100 years)
11.2	Relationship	None	Any relationship
12.8.1	Consideration Information	Consideration >= €25,000 and Consideration >= Declared Market Value	Consideration < €25,000
			Consideration < Declared Market Value
15.1	Tax Relief	No Tax Relief/Other Tax Relief	Transfer between spouses/civil partners
			Certain transfers following the dissolution of a marriage/civil partnership
			Transfer by Housing Authorities and affordable homes partnerships
			Certain transfers for cohabitants

Value Added Tax (VAT)

The sales prices used for calculating the index and additional indicators include VAT at 13.5%, where applicable, as required by the Regulation (EU) 2016/792.

Multiple Dwelling Transactions

In some cases, more than one dwelling can be purchased in a single transaction and reported on a single stamp duty entry, e.g. Apartments 1 - 12. In such cases, each of the dwellings is accounted for as a separate unit sold at the average price per unit in the bundle.

Administrative data matching

Unique identifier

The key challenge in linking the stamp duty dataset to both the BER and the GeoDirectory datasets was the lack of a common unique identifier. The only option available was matching by the address of the property, a very difficult and computationally demanding process, given considerable variation that exists in address spellings.

Eircodes were introduced in July 2015, and these are ideally positioned to become a common unique identifier for the properties. Unfortunately, the uptake of the Eircode is yet less than comprehensive, and only 24% of addresses were matched by Eircode between all three datasets in 2019. For the remaining 76% of properties sold linking by address still remains the only way to obtain the information on dwelling type, floor area, deprivation index and exact location.

Address string matching

To match Stamp Duty addresses to BER and GeoDirectory addresses, an algorithm is used that measures *Jaro-Winkler Similarity* (JWS) between them. JWS is a metric of similarity between two character strings, with a score of 0 implying zero commonality and a score of 1 implying an exact correspondence.

To facilitate the matching process, addresses on the respective datasets are standardised, uncommon

characters such as commas, apostrophes, full stops, etc. are stripped, and dwelling street numbers, if available, are separated as a new variable. To make the procedure of matching between two datasets more efficient, both of them are split into strata, and the stamp duty addresses are only compared with GeoDirectory and BER addresses within the same stratum.

Every stamp duty address is compared to every BER or GeoDirectory address *in the same stratum* and the JWS is calculated for each pairing. The highest scoring match is automatically presumed to be a correct one, provided its score is above 0.9. Where no address pairing achieves this threshold, the top 20 highest scoring pairings are checked visually by CSO staff.

For matching to GeoDirectory the strata are defined by county of the address, since the match needs only to be accurate enough to define a Small Area, where the property is located. Where neither automatic, nor manual inspection was able to provide an exact match, a Small Area code is assigned to each property using the CSO's SAPMAP interactive tool or GeoHive maps supported by OSi.

A successful match to the BER data provides information on the floor area and type of the dwelling, which is unique for that dwelling. Therefore, the strata for BER matching are defined by the county of the address *and* by the street number of the property. This means that the stamp duty addresses are only paired with the BER addresses, which are located in the same county *and* have the same street number.

In 2019 just over 74% of the addresses of the properties sold to household buyers at market prices were linked to BER data and assigned precise dwelling type and floor area. This figure is comprised of 24% matched by Eircode, 22% by automatic algorithm and 28% manually by CSO staff.

When the exact match to a BER certificate is not available, in some cases it is still possible to establish the floor area and the type of the dwelling sold, e.g. from the sale advertisements on the estate agents and other property websites. Since these are the only explanatory variables provided by BER data (see below), such transactions are included in the calculations of the index.

Overall match rates between household market stamp duty returns and BER certificates by year of execution are given in table 3 below.

Table 3: Stamp Duty Return to BER Match Rate

Year	Stamp Duty Returns Filed	BER Matches	Floor Area and Type of Dwelling found	% Overall Match Rate
2010	17,930	13,505	156	76.2
2011	14,981	11,330	97	76.3
2012	20,700	16,667	67	80.8
2013	24,505	20,128	70	82.4
2014	34,978	26,618	112	76.1
2015	39,270	31,003	237	78.9
2016	39,696	29,738	401	74.9
2017	43,084	34,778	3,627	80.7
2018	44,325	36,243	6,141	81.8
2019	44,983	35,294	2,306	80.0

All 100% of the properties above were assigned a small area code.

Quality adjustment

Why quality adjust?

Residential properties are heterogeneous, meaning that no two houses or apartments are exactly identical. This poses a challenge when trying to construct a price index, as there is a need to separate pure price change from differences in the quality and mix of the products being bought over time. Typically, this is done by comparing the prices of exactly the same products, time after time. For example, this is the method used in the Consumer Price Index, where a fixed basket of consumer goods is re-priced every month. However, in the case of residential properties, price is determined by many characteristics (location, size, dwelling type etc.) which make direct price comparisons difficult. Furthermore, only a small portion of the total housing stock is sold in any given month. The combination of these factors means that the price comparison process that would typically be used to calculate a price index cannot be used in the case of houses and apartments.

Hedonic regression

The hedonic method is the prevalent statistical process for the measurement of change of residential property price. In this method, transactions over two or more successive periods are pooled and the characteristics which influence price (dwelling type, dwelling size, geographical location and neighbourhood deprivation/affluence) are analysed and their relative contributions to the overall price are estimated. By excluding the price change determined by these characteristics independently, we are left with a pure price change for a consistent set of characteristics from one time period to another, which is used then to construct an elementary property index.

The hedonic method used for the RPPI uses a log-linear functional form. The equation is as follows;

$$\ln(p_{it}) = \boldsymbol{x}_{it} \cdot \boldsymbol{\beta} + \delta_t D_t + \mu_{it}$$

where

p_{it} is the price of a dwelling i in period *t;*

$\boldsymbol{x}_{it}$ is a vector of explanatory variables (size, type of dwelling, etc.) of dwelling i in period *t;*

$\boldsymbol{\beta}$ is a vector of regression coefficients;

δ_t is a time period coefficient;

D_t is a 'time dummy' (value=1, if in time period *t*, otherwise 0);

μ_{it} is an error term.

When the regression is applied to a pool of data covering multiple time periods, the time coefficients δ_t can be derived for each period.

The change in the price between two successive periods can then be found by dividing the antilog of δ_t divided by the antilog of δ_{t-1}, and the value of the price index for period *t* is given by

$$I_t = \frac{e^{\delta_t}}{e^{\delta_{t-1}}} \times I_{t-1}$$

where I_t, I_{t-1} are the index values in the current and preceding periods.

The rolling time dummy approach

So far, we have considered hedonic regression as it applies to a single dataset covering multiple time periods (months in the case of the RPPI). In practice, every month a new set of transactions can be added to the data and the whole regression can be run again. However, there are two problems with this approach. Firstly, this would inevitably lead to revisions for every other month as well, which is not ideal for an official house price index. Secondly, this approach implicitly assumes that buyers' preferences for the various dwelling characteristics does not change over the whole period (the coefficients apply to the whole period), which may not be the case over the longer term. To avoid these disadvantages, a 12 month rolling time dummy regression model is used for the RPPI.

In the 12 month rolling time dummy method, the regression is first run over the first 12 months of data. For month 13, the regression is run over months 2 to 13. For month 14, the regression is run over months 3 to 14, etc. Thus, the regression is always applied to a 12-month window, which is always moving forward. In this manner the regression coefficients are updated every month, allowing buyer preferences to change.

Stratification

To produce separate price indices for subsets of the housing market, stratification is required. Stratification entails dividing the pool of transaction data into the desired subsets and running the hedonic regression method on each subset independently. The choice of the strata is very much limited by the availability of transaction data. Too few transactions lead to very volatile or even unviable price indices.

National RPPI is an aggregation of 13 elementary indices, each calculated based on the transactions in a separate stratum. The elementary indices are combined into 7 aggregate indices by using chain-linked Laspeyres approach. The full scheme of aggregate and elementary price indices is shown below in Table 4 (except for the aggregate for National Houses).

Table 4: RPPI Stratification

Aggregate 1	Aggregate 2	Aggregate 3	Elementary
National Index	Dublin residential dwellings	Dublin houses	Dublin City houses
			Dún Laoghaire-Rathdown houses
			Fingal houses
			South Dublin houses
		National Apartments	Dublin apartments
	National (excluding Dublin) residential dwellings		National (excluding Dublin) apartments
		National (excluding Dublin) houses	Midland houses
			West houses
			Mid-East houses
			Mid-West houses
			South-East houses
			South-West houses
			Border houses

Aggregate indices for 2005 - 2009

For the period 2005 – 2009, RPPI was based on mortgage data provided by the main lending institutions in the country. The volume of the data was insufficient to permit the level of disaggregation achieved with stamp duty data in the later years. Therefore, only 4 sub-indices are available for this period: for apartments and houses in Dublin and the rest of Ireland.

Explanatory variables

Apart from the time dummies, just four explanatory variables (the characteristics) are used in the RPPI hedonic regression. These are tabulated below in Table 5. The total floor area tends to be the most important explanatory variable across the price models, followed by the Eircode routing key. The Eircode routing keys are used to create dummy variables in the price models. The exact Eircode routing keys vary from one price model to another, depending on the geographical location.

Table 5: RPPI model variables

Variable	Source
Total floor area (m^2)	BER
Dwelling type (semi-detached/detached/terraced) *	BER
Eircode routing key	GeoDirectory
Deprivation Index	Pobal HP Deprivation Index

** Dwelling type is not applicable to the apartment price models*

Outlier detection

For each reference month the regression is run twice on the 12 months of pooled data. The first or preliminary run is to identify outliers, transactions with unusual or extreme prices that exert undue influence. In the preliminary run the *Cook's Distance* is computed for each transaction. Cook's Distance measures the leverage of each transaction on the overall regression fit, i.e. how influential the transaction is in the determination of the regression coefficients . Higher leverage is associated with extreme values, i.e. transactions where the dwelling price appears exceptionally high or exceptionally low. If the Cook's Distance exceeds (*4/n*), where *n* is the size of the data pool, the transaction is considered to be an outlier. The outliers are then excluded from the data pool and the regression is run again to estimate time dummy coefficients. Table 6 below shows the percentage of transactions identified as outliers for each price model for the three months in 2019, by way of example.

Table 6: Outlier frequency by price model and reference month

Model	2019		
	October	November	December
Dublin City houses	4.6%	4.6%	4.6%
Dún Laoghaire-Rathdown houses	4.5%	4.4%	4.7%
Fingal houses	4.0%	4.0%	4.0%
South Dublin houses	5.0%	5.2%	5.0%
Dublin apartments	4.9%	5.0%	4.9%
Border houses	5.9%	5.7%	5.7%
Midland houses	5.4%	5.6%	5.5%
West houses	5.8%	5.7%	5.8%
Mid-East houses	4.8%	4.7%	4.8%
Mid-West houses	6.7%	6.8%	6.4%
South-East houses	5.5%	5.3%	5.1%
South-West houses	5.2%	5.1%	5.2%
National (excluding Dublin) apartments	6.4%	6.3%	6.2%

Price model performance

For illustrative purposes, some basic regression characteristics are shown in Table 7 for each price model for the reference month November 2019. Three regression characteristics are provided: R-square, Root mean square error and Error degrees of freedom. These characteristics refer to the 12 months of pool data, from December 2018 to November 2019 after outliers have been removed.

The R-squares for the Dublin price models are generally very high, implying that the four price-determining characteristics used in the models, floor area, dwelling type, Eircode routing key and the deprivation index, plus the time dummies, are sufficient to explain around 80% of the price variation observed in the data. The R-squares are generally not as high in the Border and Midland regions, suggesting that there are other unaccounted factors at play.

Table 7: Model performance characteristics for November 2019

Price Model	R-Square	Root mean square error	Error degrees of freedom
Dublin City houses	0.8469	0.17022	3177
Dún Laoghaire-Rathdown houses	0.7967	0.16229	1592
Fingal houses	0.7857	0.16076	2545
South Dublin houses	0.8481	0.10914	2257
Dublin apartments	0.7633	0.17045	3367
Border houses	0.6532	0.26326	1428
Midland houses	0.7178	0.19832	1283
West houses	0.8478	0.20669	1700
Mid-East houses	0.8218	0.14705	5457
Mid-West houses	0.7770	0.20002	2013
South-East houses	0.7295	0.20604	2300
South-West houses	0.7962	0.19515	3694
National (excluding Dublin) apartments	0.7251	0.26375	2571

Data smoothing

The need for data smoothing

As the number of transactions used in the price model decreases, the volatility or statistical 'noise' of the resulting price index increases. This noise can make it very difficult to identify turning points in the residential market in a timely manner, and represents a dilemma in price index modelling. On the one hand, it is desirable to restrict price indices to sufficiently large volumes of transactions to minimise the noise. On the other hand, there is a competing demand from users for price indices at ever increasing levels of disaggregation. A balance needs to be struck between these conflicting objectives. Data smoothing helps optimise this balance.

Holt-Winters double exponential smoothing

The RPPI uses the Holt-Winters double exponential data smoothing method. In this method, the smoothing is achieved as follows:

For the first observation in the price index series, there is no smoothing, i.e.

$$s_1 = x_1$$

where

s_1 is the 'smoothed' value in period 1

x_1 is the 'raw' value in period 1

For the second observation in the series, there is likewise no smoothing, i.e.

$$s_2 = x_2$$

But the trend b_2 is calculated as follows:

$$b_2 = x_2 - x_1$$

For subsequent periods both the smoothed data value and the trend are calculated as follows:

$$s_t = \alpha x_t + (1 - \alpha)(s_{t-1} + b_{t-1})$$

$$b_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1}$$

where

α is the data smoothing factor;

β is the trend smoothing factor

Both α and β are given values between 0 and 1. Essentially, the closer the values of α and β are to zero, the smoother the data trend.

The α and β values used in the various RPPI indices are given below in Table 8. All values were chosen based on empirical testing.

Table 8: RPPI Double exponential smoothing parameters

Model	Index Type	α	β
Dublin City houses	Elementary	0.300	0.300
Dún Laoghaire-Rathdown houses	Elementary	0.300	0.300
Fingal houses	Elementary	0.400	0.400
South Dublin houses	Elementary	0.400	0.400
Dublin apartments	Elementary	0.300	0.300
Border houses	Elementary	0.300	0.300
Midland houses	Elementary	0.300	0.300
West houses	Elementary	0.300	0.300
Mid-East houses	Elementary	0.400	0.400
Mid-West houses	Elementary	0.400	0.400
South-East houses	Elementary	0.400	0.400
South-West houses	Elementary	0.400	0.400
National (excluding Dublin) apartments	Elementary	0.300	0.300

'Direct' vs 'indirect' smoothing

From January 2020 RPPI is employing the 'indirect' method of smoothing. This means that the Holt-Winters double exponential smoothing is only applied to calculated elementary indices. These are then aggregated using Laspeyres method, as described below, and no additional smoothing is applied to calculated aggregate indices, which, therefore, are only smoothed indirectly. The 'indirect' method has the advantage of preserving the additivity of the index, i.e. the important theoretical relationships between elementary and aggregate indices and the weights, as suggested by Laspeyres formulae. Additional aggregates can be easily calculated on demand and will be consistent with existing.

However, in the years 2005 – 2019 an alternative, 'direct', approach to smoothing was employed by RPPI. In this method the 'raw' aggregates are first calculated from the 'raw' elementary indices, and both elementary and aggregate indices are smoothed directly afterwards. The advantage of this method is that less smoothing may be applied to aggregate indices (larger α and β), as these are based on the larger number of transactions and are less volatile in their 'raw' state.

This advantage comes at a cost, as the additivity of the index is lost. 'Direct' smoothing makes calculation of consistent additional aggregates impossible and may lead to certain statistical inconsistences, such as the aggregate index being out of range of its components.

The α and β values used in the various RPPI indices in 2005 – 2009 are given below in Table 9.

Table 9: RPPI Double exponential smoothing parameters used in 2005 – 2019

Model	Index Type	α	β
National All Dwellings	Aggregate	0.500	0.500
National Houses	Aggregate	0.500	0.500
National Apartments	Aggregate	0.300	0.300
Dublin All Dwellings	Aggregate	0.500	0.500
Dublin Houses	Aggregate	0.500	0.500
Dublin City houses	Elementary	0.300	0.300
Dún Laoghaire-Rathdown houses	Elementary	0.300	0.300
Fingal houses	Elementary	0.300	0.300
South Dublin houses	Elementary	0.300	0.300
Dublin apartments	Elementary	0.400	0.400
National (excluding Dublin) All Dwellings	Aggregate	0.500	0.500
National (excluding Dublin) Houses	Aggregate	0.500	0.500
Border houses	Elementary	0.300	0.300
Midland houses	Elementary	0.300	0.300
West houses	Elementary	0.300	0.300
Mid-East houses	Elementary	0.300	0.300
Mid-West houses	Elementary	0.300	0.300
South-East houses	Elementary	0.300	0.300
South-West houses	Elementary	0.300	0.300
National (excluding Dublin) apartments	Elementary	0.400	0.400

Aggregate indices

The weighting scheme

As required by the Regulation (EU) 2016/792 aggregate indices, including the national RPPI, are calculated using *chain-linked Laspeyres* method.

Firstly, *price relatives* with respect to previous December, $R_{t,n}$ are calculated for each elementary index $I_{t,n}$

$$R_{t,n} = \frac{I_{t,n}}{I_{D,n}}$$

where

$I_{t,n}$ is the value of the elementary index *n* at the period *t* in year *Y*

$I_{D,n}$ is the value of the elementary index *n* in December of year *Y-1*

The price relative for the aggregate indices are computed by weighting the price relatives of their constituent elementary indices. For example, the price relative for the national RPPI is a weighted average of all 13 elementary indices

$$R_{t,IE} = \sum_{n=1}^{13} R_{t,n} \times W_{n,Y-1}$$

where

$R_{t,IE}$ is the aggregate national index

$W_{n,Y-1}$ is the annual weight of elementary index *n* in year *Y-1*

The value of the aggregate index at time *t* is then computed as a product of its price relative above and its value at the precious December

$$I_{t,IE} = R_{t,IE} \times I_{D,IE}$$

The aggregate sub-national indices are calculated in an analogous manner.

RPPI is using relative weights, that are calculated as the ratio of the value of the purchases of residential properties by households at market prices in the corresponding category to the total value in the country during the previous year, *Y-1*. The weights, thus, are held constant throughout the reference year, but are updated every January to reflect the changes in the structure of the market.

It should be noted that it is the *price relatives*, and *not the aggregate indices* themselves, which represent the weighted average of the components, as the reference values of the indices in previous December may not be equal to each other.

Change of Base year

As any other index number, RPPI represents a ratio of an underlying absolute value in the reference period to that at some chosen period, called the base. In case of RPPI the underlying value is the price of a standardised dwelling. As time progresses this approach becomes more and more flawed, as the properties of the product for which the index is computed are becoming widely different in the reference and base periods. For example, it would very difficult to compare a standard house of 1910 with that of 2010, even with quality adjustment procedure described above. Therefore, the current methodological approach for any series of index numbers is to change the base year on a regular basis, so that the properties of products in the reference and base period are not too different.

Regulation (EU) 2016/792 requires the member states to change the base period of the house price index series every 5 years, and this the adopted methodology in Irish RPPI.

As CSO still wishes to provide researchers and general public with the long-term series, it was decided that at every change of base (rebase) the RPPI series at the new base will still start in January 2005, which was the start of the original series. However, the values of the index and each sub-index will be changed by the factor equal to the ratio of the index at old and new base periods.

For example, during the change of base period from January 2005 to annual average 2015 the following changes occur to the value of the national index in December 2017:

- Value of the index in 12/2017 at Base January 2005 = 100: 100.76
- Value of the index in 2015 (average) at Base January 2005 = 100: 80.06
- Value of the index in 12/2017 at Base 2015 = 100: 100.76 x 100 / 80.06 = 125.86

The new value of the national index in December 2017 at base 2015 is therefore 125.86.

Similar recalculations are applied to all elementary and aggregate indices separately, so that all of them equal 100 in the new base period.

RPPI weights

The relative weights used in the RPPI for the reference years 2011 to 2018 are given in Table 9 below. The weights remain broadly consistent over the period concerned, however, variations with time are not insignificant.

Table 10: RPPI relative weights

Model	Reference Year							
	2011	2012	2013	2014	2015	2016	2017	2018
Dublin City houses	13.3%	16.1%	17.1%	16.8%	15.8%	14.5%	15.0%	13.6%
Dún Laoghaire-Rathdown houses	10.0%	11.8%	13.8%	14.1%	12.0%	10.4%	11.7%	11.3%
Fingal houses	8.0%	7.2%	7.9%	8.6%	8.6%	9.3%	9.5%	9.3%
South Dublin houses	6.9%	5.9%	5.8%	6.4%	6.5%	6.0%	5.7%	6.5%
Dublin apartments	7.8%	6.4%	7.0%	8.6%	9.4%	9.5%	9.6%	10.5%
Border houses	4.5%	4.0%	3.6%	3.4%	3.3%	3.6%	3.3%	3.1%
Midland houses	3.2%	2.9%	2.6%	2.3%	2.7%	2.7%	2.6%	2.6%
West houses	5.7%	5.9%	5.4%	5.4%	5.2%	5.7%	5.0%	4.8%
Mid-East houses	14.1%	12.5%	12.2%	12.4%	13.6%	13.7%	13.5%	14.6%
Mid-West houses	6.0%	6.3%	5.2%	4.5%	4.3%	4.9%	4.9%	4.8%
South-East houses	6.0%	5.9%	5.6%	4.8%	5.0%	5.3%	5.1%	4.8%
South-West houses	11.9%	12.4%	11.7%	10.2%	10.4%	11.0%	10.7%	10.0%
National (excluding Dublin) apartments	2.8%	2.7%	2.2%	2.6%	3.2%	3.5%	3.3%	4.0%
Total	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Dwelling type imputation

The system of weights requires separate values for houses and apartments. In approximately three out of four cases where the stamp duty returns are matched to a BER certificate, the dwelling type is known. For the remaining one-in-four cases the dwelling type is imputed by the following step-by-step deterministic process:

1. If the address contains one of the words 'Apartment', 'Apt.', 'Flat', 'Block', 'Floor', 'Unit', 'Mews' or 'Condominium' the dwelling is assumed to be an apartment. If the address contains the word 'Cottage' or 'Terrace' the dwelling is assumed to be a house.

2. Otherwise, if the address contains a key word associated with a pre-defined list of known apartment-type complexes (e.g. the 'Gasworks', 'City West Plaza', 'Isoldes Tower', in Co. Dublin, etc.), then the dwelling is assumed to be an apartment.

3. Otherwise, if according to Census 2011, at least 95% of residential dwelling types in the Small Area were apartments, then the dwelling is assumed to be an apartment. Conversely, if 95% or more of residential dwelling types in the Small Area were houses, then the dwelling is automatically assumed to be a house.

4. If none of the above apply, if the stamp duty instrument is a 'Long-Term Lease greater than 100 years', then the dwelling is assumed to be an apartment. Otherwise, if the category of instrument is listed as 'Conveyance/Transfer of Property' then the dwelling is assumed to be a house.

Price indices for new and existing dwellings

An additional set of separate price indices is produced for new and existing dwellings, as defined by the indicator on the stamp duty form. The methodology used to compile these is very similar to the one used in the main RPPI. Hedonic regression model using rolling time dummy approach, as described on page 5, is used for quality adjustment, and the separate 'raw' price indices are calculated for new and existing dwellings. However, due to comparatively low number of purchases of the new dwellings, it is not possible to produce a reliable monthly index for the new dwellings, nor to produce separate indices for new apartments/ houses, nor to provide the same level of geographical detail as for the existing dwellings. The indices for new and existing dwellings, therefore, are available on quarterly basis and at national level only.

The rolling window used in regression covers 4 successive quarters of the data. The model variables are the same as in the main RPPI, with addition of the dwellings' BER rating to the model for new dwellings only, see Table 11.

Table 11. Regression variables for new and existing dwellings

Variable	Source
Total floor area (m^2)	BER
Dwelling type (semi-detached/detached/terraced/apartment)	BER
Eircode routing key	GeoDirectory
Deprivation Index	Pobal HP Deprivation Index
Building Energy Rating (new dwellings only)	BER

'*Raw*' elementary indices are then '*smoothed*' using Holt-Winters double exponential method, as described on page 9. Both smoothing parameters are set to 0.5 for both new and existing dwellings' models.

After smoothing, the aggregate quarterly national index is calculated using annually chain-linked Laspeyres methodology as described on page 11 above. The relative annual weights used in aggregation are the shares of the values of purchases of new and existing dwellings, respectively, in the total expenditure in the previous calendar year.

One would expect the aggregate quarterly national index computed in this process to be consistent with the monthly RPPI. This does not always happen automatically, as the two aggregates are obtained using separate, although similar, methods.

To achieve consistency, the indices for new and existing dwellings are *calibrated* so that the quarterly value of the aggregate index is equal to the average of the three corresponding monthly values of the national RPPI.

The price indices for new and existing dwellings are calculated from Q1 2010 onwards, as the dwellings status has only become available with the introduction of electronic stamp duty forms, which happened in 2010.

Provisional results

Stamp Duty data progressivity

The legal deadline for submitting a stamp duty return is within 44 days from the date of execution. This means that only a fraction of stamp duty returns is filed in the same month as the date of execution, and is available for calculating the index.

Results of the analysis of residential property transactions by household buyers at market prices, which were executed between 2015 and 2018, are shown in the Table 12 below.

Table 12. Rate of return of SDR1 forms for transactions executed in 2015 – 2018

Delay from execution	Monthly rate	Cumulative rate
0 (same month)	59.1%	59.1%
1	35.5%	94.5%
2	3.6%	98.2%
3	0.6%	98.9%
4	0.3%	99.1%
5	0.2%	99.3%
6	0.2%	99.4%
7 and more	0.6%	100.0%

Provisional and final results

The progressive nature of stamp duty returns poses challenges for the compilation of the monthly RPPI. In principle, the optimum price index is produced by waiting until effectively all returns have been made for a particular reference month. In practice, users require timely information on house price developments and in this context delaying publication for several months is not acceptable.

To resolve this dilemma, the RPPI is published on a provisional basis for three months. Initially, the RPPI is compiled for a particular reference month based solely on the transactions that were both *executed* and *filed* with Revenue in that month. In the following three months, as further transactions executed in the reference month are filed, the RPPI for the reference month is recompiled and revised. At this point it is considered final. Therefore, *the last three months of the RPPI series and the latest quarter of indices for new and existing dwellings are preliminary results subject to revision.*

Additional indicators

To meet demand from users for supplementary information on the residential property market, the RPPI is complemented by a range of additional indicators, such as total volume and value of residential property sales, plus average and median prices of the properties. Note that the volume is calculated as the number of properties sold, not as the number of transactions. The two figures may be different as more than one dwelling may be purchase in a single transaction and reported on a single stamp duty instrument. Neither average, nor median price are quality adjusted, hence, the RPPI represents a better measure of price change than either of the additional indicators.

The additional indicators include all stamp duty returns, not just those matched to a BER certificate, and thus reflect the totality of the property market. Due to this completeness of the data a very fine granulation can be achieved in the publication of results. Additional indicators are available separately for new and existing buildings, houses and apartments, for each county, local authority and Eircode routing key (the first three characters of the Eircode). Specific variable on the SDR1 form allows the compilation of the indicators by type of buyer, i.e. for first-time buyers, former owner-occupiers, non-occupying household buyers and non-household buyers.

Executions and Filings

The compilation of the additional indicators is also affected by the progressive nature of stamp duty data. Accordingly, two sets of indicators are presented, by the date of execution and filing:

- Executions refer to the month the property was transferred.
- Filings refer to the month the stamp duty return was submitted to Revenue.

Execution statistics are the definitive guide to residential property sales. However, they are necessarily incomplete for a period of time. For any given reference month, the figures will change as further returns are filed. To avoid revising the statistics indefinitely, a 12-month cut-off date is applied. After 12 months from the

execution date elapse, the indicators are considered final and further late returns are ignored for statistical purposes.

Filing statistics only represent administrative activity. However, filing statistics are not subject to revision (once a filing date has passed it cannot be revisited). There is necessarily a correlation between filings and executions. Therefore, filings may serve as useful lead indicators for developments in the residential property sector in advance of execution statistics being finalised.