An Phríomh-Oifig Staidrimh
Central Statistics Office

information
notice

RPPI Technical Paper

Launch of new Residential Property Price Index (RPPI)

Introduction

The CSO is launching a new Residential Property Price Index (RPPI) for Ireland. The new RPPI is based on Stamp Duty data supplied by the Revenue Commissioners, covering all residential property transactions in the State. The new RPPI replaces the existing mortgage-based RPPI launched in 2011. The new RPPI is statistically more robust than the existing RPPI and provides for a more detailed geographical breakdown of house price indices. The new index is accompanied by a comprehensive range of additional indicators on the residential property market.

Scope

The new RPPIcovers all market purchases of houses and apartments by households, both cash and mortgage-based transactions. Non-market transactions (e.g. family transfers) and non-household purchases (e.g. purchases by private companies or institutions) are specifically excluded from the index. Also excluded are self-builds (i.e. where the land is purchased separately) and purchases of partially built dwellings.

This coverage (i.e. market purchases by households) is defined by the EU rules for compilation of price indices on Owner-Occupied Housing in the Harmonised Index of Consumer Prices (HICP).

Data Sources

Stamp Duty data

The principal data source for the new 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 mostly 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 of the dwellings, whether they are new or existing and the transaction price for stamp duty assessment purposes.

Building Energy Rating (BER) data

A second key data source for the new RPPI 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 is 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 house, semi-detached house, etc.) and the floor area (m^2) 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

A third data source used in the compilation of the new RPPI is the GeoDirectory. This is a dataset of all buildings in the State, created and maintained by GeoDirectory using data from An Post and Ordnance Survey Ireland (OSi). The GeoDirectory contains the postal address of every building in the State, along with other geographical information such as X-Y coordinates. The GeoDirectory also contains Small Area codes. Small Areas are the smallest geographical classification used in Census of Population statistics. They are nested within existing Electoral Division boundaries and typically contain 75-150 households. The Small Area codes allow mapping to census Small Area Population Statistics (SAPS) and related statistics.

Deprivation index

A final data source used in the new 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 was derived from the demographic profile, social class composition and labour market situation of the population of each Small Area as measured in the 2011 Census of Population.

Filtering 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 identifying residential dwelling transactions from other real estate transactions. A wide variety of real estate transactions are liable for stamp duty, including commercial property transactions and agricultural land sales. A stamp duty return is considered a relevant residential dwelling transaction if the following criteria apply (see Table 1). In addition, to be a relevant transaction, the property must have an address in the State.

Table 1: Identification 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
12.2.1	Type of Contract	Contract for Sale	Contract for Site with Associated Building Agreement
		Contract for Completed New House/Apartment	
		None	Unitary Contract
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 (or arm's length) transaction 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).

Value Added Tax (VAT)

For new dwellings, the sales price reported in the stamp duty data is exclusive of Value Added Tax (VAT). In these cases, the transfer price of the property is increased by 13.5% to take account of VAT.

Multiple Dwelling Transactions

In some cases, more than one dwelling can be purchased in a single transaction. CSO Volume statistics report the volume of dwellings transacted rather than the volume of transactions per se.

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 greater than 100 years	
11.2	Relationship	None	Spouse
			Lineal Descendent
			Related by Marriage
			Joint Owners
			Tenants-in-Common
			Co-Shareholders
			Associated Companies
			Director/Company
			Partner/Partnership
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 partnerships
			Housing Authorities and affordable homes partnerships

Administrative data matching

Lack of a unique identifier

The key challenge in linking the stamp duty dataset to both the BER dataset and the GeoDirectory dataset was the lack of a common unique identifier. There was no single identifying reference number shared by the three datasets (the stamp duty data lacked Eircodes prior to July 2016). There was no alternative to matching these datasets by the

addresses of the properties, a very difficult process to achieve computationally, given the considerable variation that exists in address spellings. On the other hand, matching the GeoDirectory was easily matched at Small Area level to both SAPS and the HP Deprivation index, as these three datasets all contain Small Area codes.

Address string matching

For the matching Stamp Duty addresses to BER and GeoDirectory addresses, an algorithm is used that measures *Jaro-Winkler Distance*. Jaro-Winkler Distance 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. A useful feature of Jaro-Winkler Distance as far as address matching is concerned is that it gives extra weight to the first four characters of each string. Urban addresses usually begin with house numbers and so it is particularly important that these should agree in the case of matching stamp duty addresses to BER addresses (this was less important in the case of matching to the GeoDirectory, as only the Small Area was required).

The address string matching is a multi-step process. Firstly, addresses are 'cleaned' on the respective datasets (uncommon characters such as commas, apostrophes, full stops, etc. are stripped from the addresses etc.). Next the addresses are 'blocked' by county (and post code in the case of Dublin). Then every clean stamp duty address is compared to every single BER and GeoDirectory clean address in the same block. The Jaro-Winkler distance is calculated for each pairing. The highest scoring match is automatically presumed to be a correct match, provided the score is at least 0.88. Where no address pairings achieve the 0.88 threshold the top 20 highest scoring pairings are checked visually by CSO staff. If this fails to confirm a viable match, a Small Area code is assigned using the CSO interactive Small Area Population Statistic Maps (SAPMAP) tool for Census 2011.

Data matching results

Overall match rates between household market stamp duty returns and BER certificates by year are given in table 3 below. All of these returns have been assigned a Small Area code.

Table 3: Stamp Duty Return to BER Match Rate

Year	Stamp Duty Returns Filed	BER Matches	% BER Match Rate
2010	13,298	10,030	75.4
2011	13,923	10,307	74.0
2012	18,668	14,594	78.2
2013	22,777	18,582	81.6
2014	31,985	22,155	69.3
2015	37,840	28,176	74.5

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 - or more simply - a residential property price index.

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

$$\ln(p_{it}) = x_{it}\beta + \delta_t D_t + \mu_{it}$$

where

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

x_{it} is a vector of explanatory variables (size, type of dwelling, etc.) of dwelling *i* in period *t*

β is a vector of explanatory price coefficients

δ_t is a vector of time period coefficients

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 coefficient δ_t can be derived for each period (except the reference period, typically the first period, where $\delta_1 = 1$).

For any two successive time periods, t-1 and t, the antilog of δ_t divided by the antilog of δ_{t-1} provides an estimate of the aggregate quality-adjusted house price change that has occurred (i.e. the change in house prices after changes in the various known explanatory variables have been accounted for). Thus the index for period *t* is given by

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

where

I_t is the index in period *t*

I_{t-1} is the index in period *t-1*

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 β 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 β coefficents are updated every month, allowing buyer preferences to change. As the index result for the latest month is only calculated in reference to the previous month, there is no need for wholesale revisions of the price series.

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 strata is very much limited by the availability of transaction data. Too few transactions lead to very volatile or even unviable price indices.

For the RPPI there are 13 separate strata, each producing their own elementary price indices. These comprise 11 house strata and 2 apartment strata. Four of the house strata relate to the four local authorities in Dublin. The remaining seven house strata comprise a regional breakdown outside of Dublin. Aggregate indices are created from combining the elementary price indices. The full scheme of aggregate and elementary price indices is

shown below in Table 4 (with the exception of the aggregates for National Houses and National Apartments).

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
			Dublin apartments
	National (excluding Dublin) residential dwellings	National (excluding Dublin) houses	Border (exc. Louth) houses
			Midland houses
			West houses
			Mid-East (inc. Louth) houses
			Mid-West (inc. South Tipperary) houses
			South-East (exc. South Tipperary) houses
			South-West houses
			National (excluding Dublin) apartments

Explanatory variables

Apart from the time dummies, just four explanatory variables (the x_{it} 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 elementary index for each reference month, two hedonic regressions are run on the 12 months of pooled data. The first or preliminary regression is run to identify outliers,

transactions with unusual or extreme prices that exert undue influence. In the preliminary regression run Cooks Distance is computed for each transaction. Cooks Distance measures the leverage of each transaction on the overall regression fit (i.e. how influential a particular transaction is in the determination of the β explanatory price coefficients). Higher leverage is associated with extreme values (i.e. transactions where the dwelling price appears exceptionally high or exceptionally low for its particular set of known characteristics). Transactions where the Cooks Distance exceeds (*4/n*), the conventional cut-off (where *n* is the number of transactions in the data pool), are considered outliers. These outliers are then excluded from the data pool and the regression is run again to estimate the δ_t and δ_{t-1} time dummy coefficients. Table 6 below shows the percentage of transactions identified as outliers for each price model for the six successive preliminary regressions run from February 2016 to July 2016, by way of example. In these cases the frequency of outliers varies from 0.0% to 10.2%.

Table 6: Outlier frequency by price model and reference month

Model	Month					
	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16
Dublin City houses	6.3%	6.6%	5.6%	3.9%	6.3%	4.9%
Dún Laoghaire-Rathdown houses	4.3%	3.3%	5.3%	1.1%	3.1%	5.6%
Fingal houses	6.7%	4.8%	4.2%	7.0%	0.0%	5.6%
South Dublin houses	7.1%	7.9%	2.0%	2.4%	1.7%	4.5%
Dublin apartments	4.2%	4.6%	1.4%	3.9%	2.2%	7.5%
Border (exc. Louth) houses	3.7%	10.2%	8.4%	5.0%	5.8%	5.9%
Midland houses	9.4%	6.4%	6.4%	5.3%	6.4%	10.1%
West houses	6.8%	5.6%	7.4%	4.4%	9.1%	10.2%
Mid-East (inc. Louth) houses	2.2%	6.4%	6.6%	4.9%	5.9%	7.0%
Mid-West (inc. South Tipperary) houses	7.6%	8.4%	3.6%	7.6%	9.6%	6.6%
South-East (exc. South Tipperary) houses	5.7%	6.4%	6.9%	5.3%	8.1%	1.5%
South-West houses	6.6%	4.0%	7.2%	8.1%	6.1%	7.3%
National (excluding Dublin) apartments	8.2%	1.8%	7.4%	6.9%	7.9%	7.8%

Price model performance

For illustrative purposes, some basic regression characteristics are provided for each price model for the reference month July 2016. 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 August 2015 to July 2016 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 74-87% of the price variation observed in the data. Outside of Dublin, the R-squares are

generally not as high, particularly in the Border region, suggesting there are other unaccounted factors at play.

It should be noted that these regression characteristics are based on stamp duty returns filed with Revenue up to July 2016 only. As further returns are filed relevant to the period concerned and the regressions are rerun, the error degrees of freedom will increase for this 12 month window, which may improve both R-square and the root mean square error.

Table 7: Model performance characteristics for July 2016 reference month

Price Model	R-Square	Root mean square error	Error degrees of freedom
Dublin City houses	0.83635	0.20388	2454
Dún Laoghaire-Rathdown houses	0.74270	0.19203	1099
Fingal houses	0.78775	0.19887	1497
South Dublin houses	0.87307	0.14042	1177
Dublin apartments	0.77346	0.20539	1990
Border (exc. Louth) houses	0.47500	0.38939	1670
Midland houses	0.62997	0.30778	1254
West houses	0.71518	0.32868	1879
Mid-East (inc. Louth) houses	0.81216	0.21582	3248
Mid-West (inc. South Tipperary) houses	0.61251	0.33416	1880
South-East (exc. South Tipperary) houses	0.63259	0.29111	1994
South-West houses	0.68913	0.30666	3164
National (excluding Dublin) apartments	0.72567	0.31933	1913

Weighting

The weighting scheme

The aggregate RPPI indices are compiled by weighting the monthly price changes of their constituent elementary indices. The weights are based on the total value of the corresponding household market transactions in the previous year. The RPPI is therefore a *Laspeyres-type* price index, as the weights refer to a past period and are held constant throughout the reference year.

As there are 13 elementary indices, the national RPPI for any given month *m* and year *y* is given by the formula;

$$RPPI_{y,m} = \sum_{n=1}^{13} I_{n,y,m} \times W_{n,y-1}$$

where

$RPPI_{y,m}$ is the national aggregate index

$I_{n,y,m}$ is elementary index *n*

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

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

Dwelling type imputation

The system of weighting requires separate expenditures for house 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. The dwelling type is imputed according to the following step-by-step deterministic process:

1. If the address contains one of the words 'apartment' , 'apt.', 'flat', 'block', 'floor' 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 it is automatically 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.

Table 8 below gives the proportion of dwelling types imputed in each of these four stages and the final percentage of unknown dwelling types that are imputed as apartments. The majority of dwelling types are imputed in step 3, based on the Census 2011 Small Area profiles.

Table 8: Dwelling type imputation

Year	Imputation Step Assignments				Apartment Imputation Rate
	Stage 1	Stage 2	Stage 3	Stage 4	
2010	10.9%	3.3%	64.5%	21.3%	12.4%
2011	10.1%	2.3%	65.9%	21.6%	10.1%
2012	10.1%	2.8%	61.9%	25.3%	10.0%
2013	11.8%	4.3%	57.4%	26.5%	12.1%
2014	12.6%	5.1%	57.0%	25.3%	15.2%

RPPI weights

The relative weights used in the RPPI for the reference years 2011-2016 are given in Table 9 below. The weights remain broadly consistent over the period concerned. The index for Dublin City houses price index consistently has the highest weighting. The smallest weightings are for *National (excluding Dublin) apartments*, for the years 2011 to 2013 and for *Midland houses* for 2014 to 2016.

Table 9: RPPI relative weights

Model	Reference Year					
	2011	2012	2013	2014	2015	2016
Dublin City houses	15.7%	17.5%	18.2%	17.6%	16.7%	15.3%
Dún Laoghaire-Rathdown houses	12.0%	12.8%	14.4%	14.3%	12.0%	10.6%
Fingal houses	8.4%	7.2%	7.9%	8.4%	8.0%	8.0%
South Dublin houses	7.5%	6.0%	5.8%	6.2%	6.2%	5.9%
Dublin apartments	6.7%	6.1%	6.7%	8.7%	9.5%	9.8%
Border (exc. Louth) houses	3.6%	3.7%	3.5%	3.3%	3.4%	3.7%
Midland houses	2.9%	2.8%	2.6%	2.3%	2.7%	2.8%
West houses	5.6%	5.8%	5.4%	5.5%	5.3%	5.9%
Mid-East (inc. Louth) houses	12.7%	12.2%	11.6%	11.8%	13.1%	13.0%
Mid-West (inc. South Tipperary) houses	5.7%	5.9%	5.1%	4.5%	4.4%	5.0%
South-East (exc. South Tipperary) houses	5.6%	5.8%	5.4%	4.8%	5.1%	5.3%
South-West houses	11.4%	11.9%	11.4%	9.9%	10.3%	11.1%
National (excluding Dublin) apartments	2.1%	2.2%	2.0%	2.6%	3.3%	3.6%
Total	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Data smoothing

The need for data smoothing

As the number of transactions used in the price models 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. This noise is 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 10. The aggregate indices have larger *α* and *β* parameters (i.e. less smoothing is applied) as these are based on larger volumes of transactions and are inherently less volatile. For each individual index, the *α* and *β* values are given equal weighting. All these values were chosen based on empirical testing.

RPPI 2005-2009

Mortgage transaction data

Only stamp duty returns filed from 2010 onwards are available to the CSO for the purposes of compiling the RPPI (Revenue's eStamping online system launched in January 2010). Therefore, it is only possible to compile a house price index based on stamp duty data from this date forward. Prior to this, the RPPI is based on mortgage transaction data provided by the financial lending institutions. This mortgage-based series dates back to January 2005, when the original RPPI was first launched (for full details on the methodology that applies for this 2005 to 2009 period, see paper Constructing a National House Price Index for Ireland by N O'Hanlon). To maintain continuity

between the original and the new RPPI, the new stamp-duty based RPPI commences in January 2010 at the original mortgage based RPPI indexation level for this month.

Table 10: RPPI Double exponential smoothing parameters

Model	Index Type	Alpha	Beta
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 (exc. Louth) houses	Elementary	0.300	0.300
Midland houses	Elementary	0.300	0.300
West houses	Elementary	0.300	0.300
Mid-East (inc. Louth) houses	Elementary	0.300	0.300
Mid-West (inc. South Tipperary) houses	Elementary	0.300	0.300
South-East (exc. South Tipperary) houses	Elementary	0.300	0.300
South-West houses	Elementary	0.300	0.300
National (excluding Dublin) apartments	Elementary	0.400	0.400

Missing series

For the period 2005-2009, there are no house price indices for the four Dublin administrative areas nor are there regional house price indices outside of Dublin. The volume of mortgage transactions in this period was insufficient to permit this level of disaggregation.

Data smoothing

To maintain consistency with the RPPI from 2010 onwards, the RPPI from 2005-2009 has been smoothed using the *Holt-Winters double exponential smoothing* technique. Previously, the RPPI for this period had been smoothed using a three month rolling average. This results in some very minor differences between the originally published and the current RPPI for the 2005-2009 period.

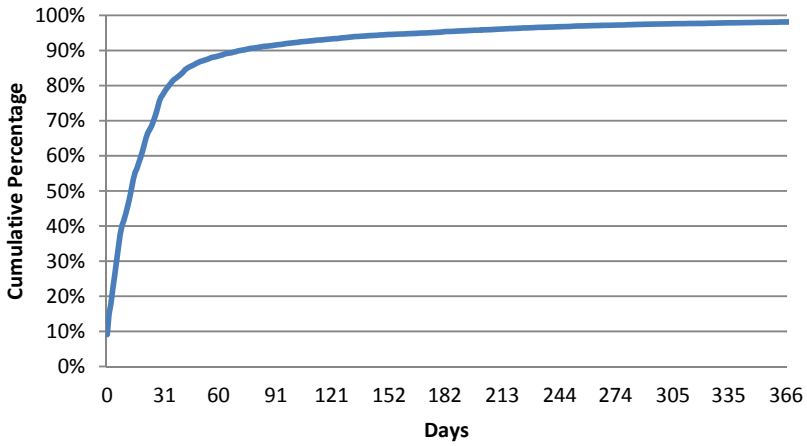
Provisional results

Stamp Duty data progressivity

The legal deadline for submitting a stamp duty return is within 44 days of the date of execution. Some returns are submitted later than this deadline (in a very small number of cases, months or even years late). In practice, therefore, only a fraction of the transactions executed in any given month are available for index compilation the following month.

Figure 1 below shows the cumulative rate of return for stamp duty returns filed by households for transactions executed in 2010. Some 85% of these returns were filed within the 44 day deadline. However, even after a full year had elapsed, 2% of the returns still had not been filed.

Figure 1: Accumulated rate of return for SDR1 forms filed for household market-based residential dwelling transactions executed in 2010



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 two months. Initially, the RPPI is compiled for a particular reference month based solely on the transactions that were both *executed* in that month and *filed* with Revenue in that month. The following month, as further transactions executed in the reference month are filed, the RPPI for the reference month is recompiled and revised. The month after that, as yet more transaction executions for the reference month are filed, the RPPI is revised again. At this point it is considered final. Therefore, *the last two months of the RPPI series are preliminary results subject to revision.*

Rather than delay the RPPI, preliminary results are prepared based on the early returns. These preliminary results are updated the following two months as new transaction data for

the reference month becomes available. Therefore, the RPPI results for the latest two months are provisional.

Additional indicators

Volume, value and average price

To meet demand from users for supplementary information on the residential property market, the RPPI is complemented by an extensive range of additional indicators. Three specific indicators are being provided, the volume, value and average price of residential property transactions. Volume is the number of dwellings transacted (note than more than one dwelling can be purchased in a single transaction). Value is the total value of all dwellings transacted (in millions of euro, inclusive of VAT on new dwellings). Average price is the value divided by the volume (note that average price is not quality adjusted). These indicators are compiled on the basis of all stamp duty returns (not just returns matched to a BER) and thus reflect the totality of residential property transactions.

Breakdowns available

Several breakdowns are available for the additional indicators. Firstly, there is a breakdown between market and non-market transactions (as defined previously). Secondly there is a breakdown between household and non-household purchases (again as defined previously). Household purchases are further broken down into first-time buyer owner-occupiers, former owner-occupiers and non-occupiers. These breakdowns present the full range of residential dwelling transfers. However, it should be noted that only household market-based purchases are reflected in the RPPI. The additional indicators are also available broken down by new and existing dwellings and by county.

Specifically for household market purchases, further breakdowns are available for the additional indicators. These include a breakdown into houses and apartments, although this classification is partially based on imputed information. A more detailed county breakdown is also available, with Dublin split into its four Council areas and Cork, Galway, Limerick and Waterford split into City Council and County Council areas. Volume value and average price statistics are also provided broken down by Eircode routing keys (the first three characters of the Eircode). Descriptions are provided for the Eircode routing keys; these are not official labels but they generally follow pre-existing postal town names.

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, executions and filings:

- 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. For any given reference month, they will increase in subsequent months as further returns are filed. To avoid revising these statistics indefinitely, a 12-month cut-off date is applied. After 12 months elapse, the execution indicators are final (further late returns are ignored for statistical purposes). Prior to this 12 month period, executions statistics are subject to revision.

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.

Enquiries to:

Gregg Patrick on tel: 021 453 5202 or email gregg.patrick@cso.ie
Dan Gallagher on tel: 021 453 5210 or email dan.gallagher@cso.ie
Residential Property
Central Statistics Office
Skehard Road
Cork
T12 X00E

21 September 2016