Redeveloping Ireland's Residential Property Price Index (RPPI)

Ottawa Group, 10-12 May 2017, Eltville, Germany

Gregg Patrick[1]

Abstract

House price indices are important statistical indicators for monitoring developments in residential property markets. This is particularly true in the case of Ireland, which in the recent economic crisis experienced one of the most pronounced boom-bust property cycles amongst developed countries. To improve measurement and coverage of the housing market, Ireland redeveloped its Residential Property Price Index (RPPI). This new index was launched in September 2016[2]. The original index was based on mortgage transaction data whereas the new index is based on matched administrative data sources. The practical and technical issues encountered in this redevelopment are discussed, including data matching and data progressivity. The methodological and conceptual challenges involved in producing high-frequency and disaggregated house price indices from relatively sparse data are also considered. Results are compared and contrasted for Ireland's New and Original RPPI's.

Keywords: Residential property price index, data matching, administrative data, Ireland

1. Introduction

1.1 The economic crisis

During the recent economic crisis, Ireland experienced the greatest contraction in house prices of any OECD country. From the first quarter of 2008 to the first quarter of 2013, house prices in Ireland fell 52.6%[3]. The next deepest fall in house prices was experienced in Latvia, where prices declined 50.0% between the first quarter of 2008 and the first quarter of 2010. The next greatest falls in price were in Estonia and Greece, with house price declines of 45.0% and 42.5% respectively since their 2008 highpoints. By way of contrast, in the United States, where the crisis first manifested, house prices fell 17.6% from 2008 to 2012 (see figure 1).

The effect of the collapsing property market in Ireland was profound. In 2008, there were 255,000 persons employed in the construction industry. By 2013, this number had dropped to just 96,000 (source: CSO). The national unemployment rate rose from 5.0% at the start of 2008 to 15.1% by the end of 2010 (source: CSO). By 2012 an estimated 314,000 homeowners were in negative equity. The rate of mortgages in arrears rose to highpoint of 12.9% in 2013 (source: Central Bank of Ireland). Collapsing houses sales, and the attendant collapse in tax receipts from stamp duty, had a major effect on government finances. Government finances, which were in surplus in 2007, attained an unprecedented deficit of 32.4% in 2010 (source: CSO).

---

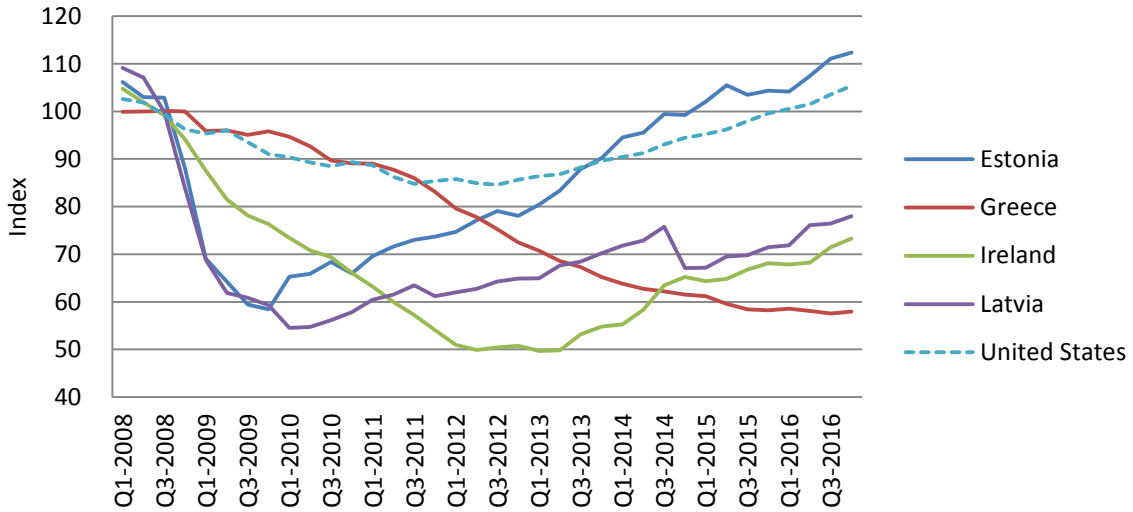
[1] Residential Property Section, Central Statistics Office

[2] https://goo.gl/SX0I8k

[3] Measured on monthly basis, residential property prices in Ireland actually fell 54.4% between April 2007 and March 2013.

The deepening crisis had a calamitous effect on Ireland's banks, which saw their credit ratings slashed as their asset balance sheets were undermined by the collapsing property prices. The downgraded credit ratings created a liquidity crisis in the Irish banking system, which forced Ireland to formally request assistance from the European Union's European Financial Stability Fund (EFSF) as early as 2010. In November 2010 Ireland received a combined "bailout" from the EFSF and International Monetary Fund (IMF) of €65.7 billion to help restore government finances.

Figure 1: Residential Property Price Index (2008=100)



Source: OECD

1.2 The need for house price indices

One of the key lessons from this experience was the need for greater regulatory oversight of the Irish banking system, in particular in relation to the banking sector's exposure to property. There was a need for greater emphasis on a macro-prudential approach to regulation, with macro stress testing and feedback effects being given more prominence.

An important statistical tool for monitoring a nation's exposure to developments in the residential property market is an official house price index. However, at the outset of the recent economic crisis Ireland lacked such an index. An unofficial house price index did exist. This index was developed by one of the main lending institutions, *permanent tsb*, in conjunction with the Economic and Social Research Institute (ESRI). This *permanent tsb/ESRI* index was a quarterly price index based on *permanent tsb* mortgage drawdowns. The index series commenced in 1996. However, the index was becoming increasing difficult to sustain as the economic crisis progressed and the volume of mortgage drawdowns from the bank decreased dramatically.

Whilst the national requirements are very important, in fact it was developments at EU level which provided the driving force for compiling an official house price index in Ireland. The initial impetus came from the Macroeconomic Imbalance Procedure (MIP), launched by the EU in 2011 in response to the worsening economic crisis. As part of the MIP process, an Alert Mechanism Report (ARM) is generated annually. The ARM is based around a scoreboard of fourteen statistical indicators, one of which is house price inflation. An annual rate of house prices inflation greater than 6% (or less than -

6%) triggers a scoreboard alert. Thus, the MIP enshrined metrics on house price inflation at the heart of the EU's system of economic governance.

The incorporation of house price inflation within the MIP accelerated the development of a legislative framework for house price statistics. Within the European Statistical System (ESS) it was long acknowledged that the acquisition and ownership costs associated with owner-occupied housing were very significant gaps in the coverage of the Harmonised Index of Consumer Prices (HICP). These acquisition and ownerships costs could be very significant and their omission was a factor limiting the comparability of the HICP across the EU Member States. The requirement to further harmonise the HICP, and the mandate to investigate the issues affecting harmonisation (including owner-occupied housing), were first laid down in *Council Regulation 2494/95*[4]. After detailed investigation of the owner-occupier costs issue, including various pilot studies (participated in part by Ireland) the Commission enacted *Regulation 93/2013*.

It was anticipation of *Regulation 93/2013* than spurred and shaped Ireland's endeavours to develop an official house price index. Consequently, in 2011, the CSO formally launched its first RPPI (henceforth known as the Original RPPI). This Original RPPI was a monthly house price index based on mortgage transaction data supplied by all the main financial lending institutions. The index was back-dated to January 2005. The Original RPPI was Ireland's first official house price index[5].

2. The Irish housing market

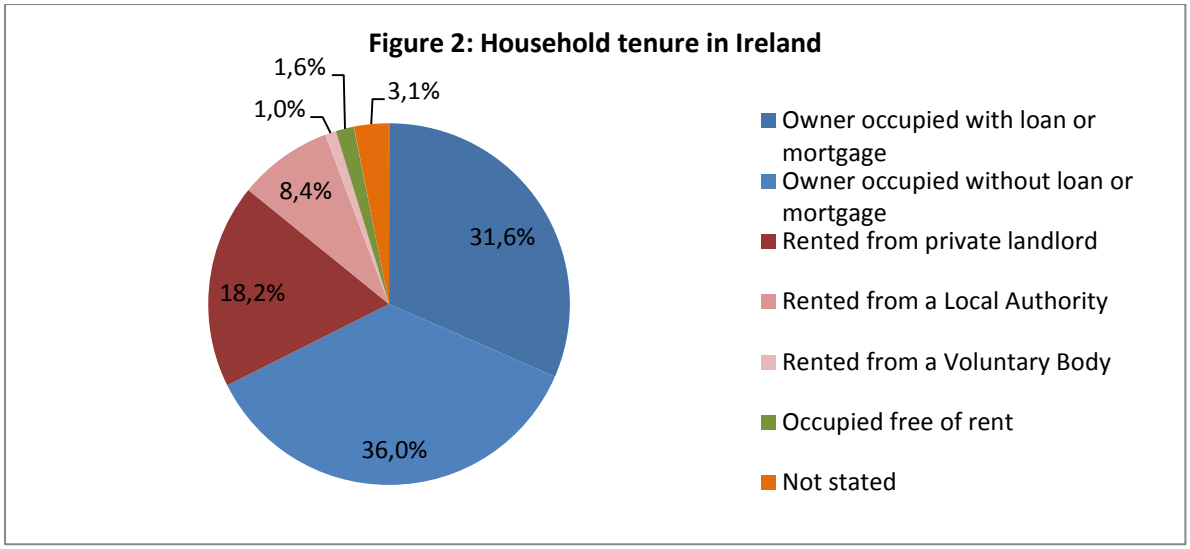
2.1 Stock and tenure of dwellings

Before going further into the regulatory framework of the RPPI, it is useful to have a quick overview of Ireland's residential property market. The stock of houses and apartments in Ireland comprises approximately 2 million dwellings[6] (source: CSO). Of those dwellings occupied by households, 67.6% are owner-occupied (31.6% with a mortgage, 36.0% without a mortgage). A further 26.6% of households rent their dwelling (18.2% from a private landlord, 8.4% from a Local Authority and 1.0% from a voluntary housing body). Some 1.6% of households occupy another's dwelling rent free. The tenure is unknown for the remaining 3.1% of cases (see figure 2).

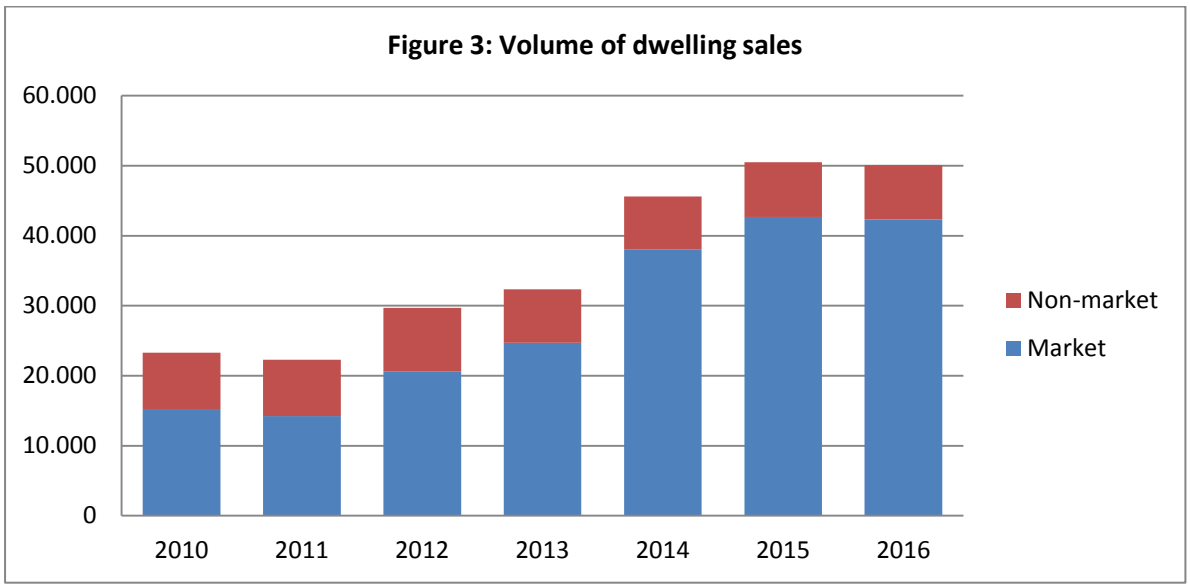
2.2 Sales of dwellings

Regarding house sales in Ireland, reliable information is only available from 2010 onwards. These data show that in 2010 only 23,000 dwellings were transacted in Ireland (source: CSO). By 2015 the number of dwellings transacted had increased to 51,000 (see figure 3). The latest data for 2016 show that 50,000 dwellings were transacted. However, it should be noted that not all of these transactions are market sales. In 2016 there were 8,000 non-market residential dwelling transactions, principally family transfers. This left just 42,000 dwellings sold on the open market, a market turnover rate of 2.1% of the dwelling stock.

[4] Council Regulation (EC) No 2494/95 of 23 October 1995 concerning harmonized indices of consumer prices.
[5] For more information on the launch of the Original RPPI see *Constructing a National House Price Index for Ireland* (N O'Hanlon): http://www.tara.tcd.ie/bitstream/handle/2262/62349/o'hanlon%20pdf.pdf;sequence=1.
[6] According to preliminary results from the 2016 census of population, there were 2,022,895 houses and apartments in Ireland on 1 April 2016 (source CSO).



Source: CSO



Source: CSO

3. Legal framework

3.1 Legal obligations

Commission *Regulation 93/2013* is the legal basis for the RPPI. The full title of the regulation is as follows:

> *Commission Regulation (EU) No 93/2013, of 1 February 2013, laying down detailed rules for the implementation on Council Regulation (EC) No 2494/95 concerning harmonised indices of consumer price indices, as regards establishing owner-occupied house price indices.*

Regulation 93/2013 legally obliges all Member States to compile on a quarterly basis two separate but related sets of price indices;

1. Owner-occupier house price indices that measure the changes in the transaction prices of dwellings new to the household sector and other goods and services that households acquire in their role as owner-occupiers.

2. House price indices that measure the changes in transaction prices of dwellings purchased by households.

The first of these sets of indices, the owner-occupier house price indices, are a relatively specialised set of price indices that are outside the scope of this paper. The second set of indices, the house price indices, provides the specific regulatory basis for an official measure of house price inflation.

3.2 House price indices

Under the second heading, the Regulation identifies three separate house price indices that Member States are obliged to compile;

H.1. Purchases of dwellings

H.1.1. Purchases of new dwellings

H.1.2. Purchases of existing dwellings

3.3 Scope

H.1.1. covers 'turnkey' new dwellings only. These are dwellings that are structurally complete when purchased. Self-builds (one-off housing commissioned by a household where the land is typically purchased in advance) are specifically excluded from this sub-index. *H.1.2.* covers all existing (i.e. second-hand) dwellings that are also structurally complete. Index *H.1.* is an aggregate index of *H.1.1.* and *H.1.2.* These indices are only required at national level from each Member State. These indices are required quarterly (but there is an option to provide them monthly).

It should be noted that *Regulation 93/2013* specifically limits the scope of the house price indices to dwellings *purchased by households* only. Dwellings purchased by institutional investors, for example, are out of scope and should not be included in the official house price indices. The technical manual accompanying the Regulation also implies that only arms-length or market-based transactions are to be considered. Non-market transactions (such as dwelling transfers between family members, which typically change hands for free or at a discounted price) are out of scope.

3.4 Quality adjustment

As the house price indices are compiled under the HICP regulatory framework, they are necessarily governed by the standard statistical requirements and treatments applicable to the HICP. These include the requirement to control for quality, effectively to ensure that house price indices reflect like-for-like comparisons over time.

The HICP achieves a like-for-like comparison by pricing the exact same basket of goods over two consecutive periods. Unfortunately, it is not directly possible to price the same 'basket' of dwellings over two consecutive periods. Every dwelling is unique by virtue of its locational characteristics alone and it is very rare for the same dwelling to be resold in two consecutive periods. This imposes

the need to apply statistical quality adjustment techniques, to separate constant quality house price inflation from house price changes due simply to the differing mix of dwelling types being sold over time.

Regulation 93/2013 is not prescriptive as to how quality adjustment is to be achieved. The technical manual, accompanying the regulation, presents a range of options. The *Handbook on Residential Property Prices Indices* also provides detailed options for the compilation of official house price indices[7]. It is up to each Member State to decide which methodology is most suitable for their particular circumstances.

4. The Original RPPI

4.1 The House Price Statistical System (HPSS)

Initial attempts to develop an official constant quality house price index for Ireland were focused with the Department of Environment, Heritage and Local Government (DoEHLG)[8]. Section 13 of the *Housing (Miscellaneous Provisions) Act 2002* granted the DoEHLG the authority to collect mortgage transaction data from the financial lending institutions for developing a House Price Statistical System (HPSS). A constant quality house price index was to form one component of the HPSS[9]. The DoEHLG developed an experimental house price index that was mix-adjusted using a fine stratification approach. However, the DoEHLG was dissatisfied with the results and formally asked the CSO to take on the responsibility of developing the house price index in 2007.

4.2. Data sources

The CSO began work on developing a house price index in 2008. The CSO commenced by investigating the potential data sources available. Stamp Duty returns were the obvious starting point. All residential property transactions must be registered for Stamp Duty. The solicitor for the buyer must complete a particular tax form, form SDR1, and submit it to Revenue within 44 days of the transaction date, along with the Stamp Duty fees (or be liable for a late payment fine). Stamp Duty returns are the ideal data source in terms of coverage, as they cover every transaction in the state. However, apart from the transaction price, a new or existing designation and the address, they contain no details on the characteristics of the dwellings being sold (characteristics such as the dwelling type, the floor area, the dwelling age, etc.)[10]. Characteristics information is essential for

[7] The development of the RPPI Handbook was co-ordinated by Statistical Office of the European Union (Eurostat) under the joint responsibility of six organizations - International Labour Organization (ILO), International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), United Nations Economic Commission for Europe (UNECE), World Bank and Eurostat.

[8] Since renamed the Department of Housing, Planning, Community and Local Government.

[9] Other components included average house price statistics. These statistics, which are not quality adjusted, are simply the average prices paid for houses every quarter. These average house price statistics, along with other housing-related statistics compiled by the DoEHLG, are hosted on the CSO website: http://www.cso.ie/px/pxeirestat/pssn/doehlg/Database/DoEHLG/Housing%20Statistics/Housing%20Statistics.asp

[10] Although for new dwellings only, the Stamp Duty returns do capture whether the floor area of the dwelling is less than 38m^2, between 38m^2 and 125m^2 or greater than 125m^2.

quality control. Because of the lack of dwelling characteristics information, the Stamp Duty data were deemed unsuitable for constructing a mix-adjusted house price index, at that point in time.

Next attention was re-focussed on mortgage transaction data collected for the HPSS. These data had already been collected with a view to compiling a mix-adjusted house price index. Therefore, the mortgage drawdown data included not only the purchase price and buyer information, but also specific information on the characteristics of the properties themselves, including the floor area, number of rooms, number of bedrooms, plot size and dwelling type (detached house, semi-detached house, terraced house, etc.). In the absence of a viable alternative, this mortgage drawdown data became the basis of the official house price index.

4.3 The rolling year hedonic regression method

As a fine stratification approach had not proved successful, several alternative hedonic methodological approaches were tested for creating a constant quality house price index[11]. Of these, the rolling year hedonic regression method was found to be the most suitable[12]. This used the floor area, number of bedrooms and dwelling type property characteristics to control for constant quality. The type of buyer (first-time buyer or former owner-occupier) and county codes (as well as post codes in Dublin) were also used to control for quality[13]. The rolling year hedonic regression method was applied twice to the data source; firstly to identify outliers (dwellings with exceptional transaction prices) and secondly to produce a house price index after the outliers had been excluded. The method was sufficiently flexible to permit the compilation of monthly (as opposed to quarterly) house price indices.

4.4 Price models

Four separate price models were developed. These were a price model for Dublin houses, for Dublin apartments, for houses outside of Dublin and for apartments outside of Dublin. The rationale for the separate models for houses and apartments was that these were considered to appeal to very distinct market segments, consequently with different price relationships to the key dwelling characteristics (such as floor area and geographical location). In addition, having separate price models for Dublin enabled the exploitation of Dublin post codes in the models. This particular breakdown was also that used in the unofficial *permanent tsb/ESRI* house price index, facilitating comparability with that series.

4.5 Indices

[11] The term *hedonic* comes from hedonic demand theory. This theory postulates that the price of a particular good, a dwelling in this case, is simply the aggregate price of a bundle of individual characteristics, internal and external. Hedonic regression is a method of decomposing the aggregate price into separate prices based on these individual characteristics. The time period in which a dwelling is sold can be viewed as an external characteristic. By identifying the changes in price which occur solely due to temporal changes (i.e. when all other relevant characteristics are accounted for) a constant quality price index can be constructed.

[12] See section 5.11 of the *Residential Property Price Index Handbook* for a general description of the time dummy hedonic method.

[13] The type of buyer is neither a dwelling nor a locational characteristic. However, it was believed to serve as a proxy indicator for dwelling quality, on the assumption that first-time buyers generally cannot afford the same standard of fittings and furnishings etc. as a prior owner-occupier.

The four house price models each generated their own corresponding elementary price index. Various aggregate indices were created by combining the elementary price indices, using weights derived directly from the mortgage transaction data. The full range of price indices produced is given below in table 1. These indices were first published in March 2011 (with the exception of the *National (excluding Dublin) apartments* elementary index, which was considered too volatile to publish), backdated to January 2005. These indices were collectively named the Residential Property Price Index.

Table 1: Original RPPI Price Indices

Code	Name	Type
I1	National Index	Aggregate
I1.1	Dublin residential dwellings	Aggregate
I1.1.1	Dublin houses	Elementary
I1.1.2	Dublin apartments	Elementary
I1.2	National (excluding Dublin) residential dwellings	Aggregate
I1.2.1	National (excluding Dublin) houses	Elementary
I1.2.2	National (excluding Dublin) apartments	Elementary
I1.A	National houses	Aggregate
I1.B	National apartments	Aggregate

5. Rationale for a New RPPI

5.1 Deficiencies of the original RPPI

Whilst the launch of the Original RPPI was a major step forward for the Irish Statistical System in 2011, even at this early stage a number of deficiencies were evident. In terms of importance, these can be ranked as follows; compliance issues, the exclusion of cash buyers, data quality issues, lack of micro-location and possible inclusion of non-market transactions and self-builds. A further concern is an inherent time-lag issue. Each of these deficiencies will be discussed in turn.

5.1.1 Legal compliance.

The Original RPPI complied reasonably well with *Regulation 93/2013* requirements to compile a quality-adjusted national house price index covering all dwelling transactions (specifically *H.1. Purchases of dwellings*). However, as we have seen, the regulation also requires the compilation of two sub-indices, *H.1.1. Purchases of new dwellings* and *H.1.2. Purchases of existing dwellings*. There was a particular problem with compiling a new dwelling price index. Namely, the collapse in the construction industry in Ireland during the economic crisis and a general decline in credit availability meant that the volume of new dwellings mortgages available to the CSO was not sufficient to compile a robust new dwelling price index, particularly for the years 2011 to 2013 (see table 2). Thus

no breakdown was available for new and existing dwellings[14]. This meant that the Original RPPI was not compliant with *Regulation 93/2013*.

Table 2: Mortgage drawdown data available to the CSO

Year	New	Existing	Total
2010	5,761	8,895	14,656
2011	2,466	6,436	8,902
2012	2,626	9,237	11,863
2013	2,096	9,367	11,463
2014	3,136	14,802	17,938
2015	3,930	17,703	21,633
2016	4,924	17,851	22,775

5.1.2 Cash-buyers

The fact that cash-buyers were necessarily excluded was a criticism of the Original RPPI cited frequently by the national media. Of course, this was not a deliberate exclusion but a constraint imposed by data availability. There are two separate issues of concern. The first and most serious concern was that cash buyers may, in certain market conditions, be able to exploit their financial flexibility to secure a better deal than mortgage buyers, effectively paying less for the same properties. If this is the case then the exclusion of cash buyers could potentially bias the Original RPPI. A second concern was that the exclusion of cash buyers simply meant a smaller pool of data available for the computation of the index. As the data pool diminishes, the volatility of the price index increases, and this creates difficulties trying to identify turning points in the residential property market.

5.1.3 Data quality

From the outset, quality issues were also evident with the mortgage drawdown data. These quality issues took two distinct forms. Firstly, there appeared to be under-reporting of the mortgage drawdowns. Comparisons made with drawdown statistics from the Banking & Payments Federation of Ireland (BPFI) suggested that the number of mortgages reported by the financial lending institutions to the CSO represented only 80-90% of the total (see table 3 below). This under-reporting reduced the pool of transaction data further with concomitant implications for the robustness of the price index. A second problem, equally serious, was that often key variables in the mortgage transaction data were left blank. This was particularly a problem with the reported floor area of the dwellings. In some cases, this partial non-response could be addressed by imputation from other property characteristics (albeit this introduced another potential source of error). But where several characteristics were missing this proved unviable. The net effect was to reduce further the pool of 'usable' transaction data (see table 3 again). This further compounded the difficulties in creating a new dwelling price index.

5.1.4. Micro-location

[14] It was very feasible to compile an existing dwelling index. But in practise, given the low volume of new transactions, was argued that the overall price index already served as a very good proxy for an existing dwelling price index.

Lack of micro-location indicators in the mortgage drawdown data was also an issue of concern. It is oft cited that the three most important factors in house purchases are 'location, location, location'! The mortgage drawdown data lacked the addresses of the properties purchased. The only geographical information supplied was a county code (and a post code in the case of Dublin properties). Obviously, the same residential dwelling will command a very different price depending on where exactly it is located within a given county (and, to a lesser extent, within a given post-code, in the case of Dublin). International experience with house price modelling suggests that micro-location, along with floor area and plot size, tend to be the three most important price determining variables[15]. Thus lack of information on the precise neighbourhoods the properties were being sold in was a very significant omission which again potentially affected the accuracy of the Original RPPI.

Table 3: BPFI versus CSO mortgage transaction data

Year	BPFI	CSO	
		Total	Usable
2010	18,313	14,656	12,892
2011	11,050	8,902	7,173
2012	14,160	11,863	9,039
2013	13,472	11,463	6,008
2014	20,155	17,938	13,224
2015	23,836	21,633	14,894
2016	24,891	22,775	15,200

5.1.5 Self-builds

The inclusion of self-builds in the mortgage transaction data was a further source of concern. Self-builds are specifically outside the scope of a house price index, as defined by the technical manual to *Regulation 93/2013*[16]. However, there is no obvious way to identify self-builds them in the mortgage drawdown data. In practise, many self-builds in Ireland are at least partly funded by a mortgage. Therefore, there was a risk that self-builds were 'contaminating' to some extent the Original RPPI. Self-builds are new dwellings and the risk of this contamination made the accuracy of any prospective new dwelling index based on mortgage transaction data even more doubtful.

5.1.6 The time lag effect

A final issue is the time lag effect, or specifically a perceived delay between actual changes in the residential property market and these changes being reflected in the Original RPPI. Again, there are two separate issues at play here. The first is the issue of when exactly a property sale is defined to have occurred. For many people, a sale occurs when a deposit has been paid and the property

[15] Although preliminary studies on plot size seemed to show that it was not a very important price determinant in the case of Ireland. In fact, as plot size had minimal significance it was excluded from the original RPPI price models. Perhaps this is because there is wide variation in the range of plot sizes sold in Ireland. For example, a relatively small and old dwelling in rural Ireland may be sold with a relatively large plot size (typically of low quality land).

[16] Self-builds are not strictly open-market transactions. Although the site may well be purchased on the open market, the option of constructing on that site is unique to the owner.

moves to 'sale agreed' status. However, *Regulation 93/2013* specifies that the house price indices compiled by Member States should be a *transaction* price index, i.e. the total price paid for the dwelling at the point where the ownership is actually transferred. This can often be a few months after the sale is agreed. In the interim, the residential property market may have changed significantly, introducing a perceived lag in the house index response. This is an inevitable limitation of any transaction-based house price index and nothing can be done about this[17].

A secondary time-lag issue is the lag introduced by data smoothing. The Original RPPI indices are very volatile in their native form. This volatility was an inevitable side effect of price modelling sparse data in a relatively heterogeneous housing market such as Ireland. To address this problem, it was necessary to apply a smoothing technique to the Original RPPI, to dampen volatility. The technique employed was a rolling three-month average. Thus, for example, the published index for December is actually an average of computed indices for October, November and December. This has the unfortunate side-effect of introducing a one-month time lag (as in the example above the published December index is actually centred in November). This was a concern in terms of being able to rapidly identify turning points in the market.

5.2 Intensification of concerns

Whilst all these issues were known deficiencies and causes of concern when the Original RPPI was launched in 2011, they were not considered of such magnitude to undermine seriously the quality of the index. However, in the following two years in particular, these concerns were brought into sharper focus as the overall economic situation continued to deteriorate. As the crisis-induced credit freeze in Ireland began to make itself felt, the number of mortgage drawdowns diminished sharply (see table 3 again). At the same time the proportion of cash-buyers increased, with some independent commentators estimating that cash-buyers were accounting for half of the market during this period. This inevitably heightened concerns of bias in the Original RPPI. Also, the quality of the mortgage data decreased noticeably, particularly in 2013, when nearly 48% of the mortgage drawdown data was unusable due to missing key information. This led to anxieties for the future viability of the Original RPPI.

5.3 Review of Original RPPI

These concerns prompted a review of the Original RPPI methodology in 2013. As part of this review, the CSO decided to investigate the possibility of creating a house price index from a separate data source, to independently benchmark the quality of the original RPPI.

6. Interim developments in data sources

[17] A 'sale agreed' house price index, even if it were practical to implement, also is problematic. For example, some sales agreed may fall through (or the price may be renegotiated). Thus a 'sale agreed' house price index is not necessarily representative of the actual house sales which occur. Interestingly, there exist in Ireland several unofficial house price indices based on asking price. But this is also problematic, as the stock of dwellings offered for sale may not necessarily be representative of the dwellings actually sold and the final sale agreed price may differ.

6.1 Changing data environment

In the period between 2011 and 2013 there were a number of significant developments in the availability of administrative data sources. These developments were to have a direct bearing on benchmarking the Original RPPI. These developments completely transformed the data environment insofar as the compilation of residential property statistics was concerned.

6.2 eStamping

To begin with, the availability and quality of Stamp Duty data underwent a major step change. In January 2010 the Revenue Commissioners launched their *eStamping* system. *eStamping* allowed solicitors conveying property to register for Stamp Duty online for the first time (as opposed to submitting a paper form). This not only rendered the whole Stamp Duty procedure much more efficient in general, it greatly reduced the time taken to capture the Stamp Duty information, increasing the timeliness of the data. The online submission process also imposed certain rigours and constraints on the information providers, which improved the quality of the data. In June 2013 the CSO signed an updated Memorandum of Understanding with the Revenue Commissioners which facilitated and regularised the sharing of *eStamping* data for statistical purposes.

6.3 Building Energy Ratings

In December 2002 the EU enacted the *Energy Performance of Buildings Directive*[18]. This directive was transposed into Irish law in 2006[19]. This required that all new dwellings built from planning permits granted in 2007 onwards must have a Building Energy Rating (BER) certificate. It also required that from 2009 onwards, all existing dwellings being sold must have a BER certificate[20]. The BER certification process was managed by the Sustainable Energy Authority of Ireland (SEAI). In the process of assessing the energy efficiency of residential dwellings, a wealth of characteristics information was captured electronically (including the dwelling type and the floor area). In 2013 the CSO signed a Memorandum of Understanding with the SEAI, granting it access to the BER data on a quarterly basis for statistical uses. Upon investigation the BER data proved to be a very high quality data source. For example, there were no issues with missing variables, etc.

6.4 The GeoDirectory

In the meantime, there was another significant development which was to have an important bearing on the creation of a new RPPI. Work on the GeoDirectory, a collaboration between An Post (the Irish postal agency) and Ordnance Survey Ireland (OSi) to attach X-Y coordinates to every commercial and residential address in the country, was progressing. As early as 2009 the CSO began testing the GeoDirectory and integrating it into its Census of Population operations. By 2013 the GeoDirectory database was at a mature stage and available to the CSO for statistical use.

An important feature of the GeoDirectory was that it contained *Small Area* codes. *Small Areas* are geographical zones mapped by the National Institute for Regional and Spatial Analysis (NIRSA), at the behest of OSi and in collaboration with the CSO. *Small Areas* were devised in order to meet user

[18] Directive 2002/91/EC of the European Parliament on the Energy Performance of Buildings (EPBD).
[19] EC Energy Performance of Buildings Regulation 2006 (S.I. No. 666 of 2006).
[20] With some minor exemptions for historic buildings, etc.

demands for more homogenous, small-scale geographical units. Approximately 19,000 *Small Areas* were created, each containing 75-150 households, nesting within existing Electoral Divisions. *Small Areas* were used in Census 2011 to produce Small Area Population Statistics (SAPS) for the first time.

6.5 Deprivation index

A related development, which exploited the availability of SAPS from Census 2011, was the *Pobal*[21] *HP Deprivation Index*, first published in August 2012. The deprivation index is a measure of 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* household as captured in the census. As the original provider of the data, the deprivation index was available to the CSO for statistical uses.

6.6 The basis for a New RPPI

From these interim administrative developments came the genesis of an idea for an alternative RPPI (henceforth known as the New RPPI). The Stamp Duty data provided transaction prices, buyer information and full coverage of the residential market. The BER data and its mandatory requirement ensured building characteristic data were theoretically available for every residential dwelling transaction. The GeoDirectory provided additional information, in particular the *Small Area* which served as a micro-locational code. Once the *Small Area* code was available then micro-location characteristics would be available from SAPS. In particular, the quality of the micro-location neighbourhood was directly reflected in the *Pobal HP Deprivation Index*[22]. If all these various data sources were combined and modelled, the resulting price index could serve as an independent benchmark of the Original RPPI, or even replace it if it proved superior. The challenge was to match and link together all these various data sources at the transaction level.

7. Administrative data matching

7.1 The challenges of data matching

The key challenge in linking the Stamp Duty data to the BER data and the GeoDirectory was the lack of a common unique identifier. There was no single reference code shared by these three datasets[23]. Therefore, the only option was to attempt to match these datasets together by the property

[21] Pobal is not-for-profit organisation that acts as an intermediary for programs funded by the Irish government and the EU.

[22] Strictly speaking, the Pobal HP Deprivation Index is not a direct measure of neighbourhood quality in terms of its physical attractiveness as a place to live, but a measure of the relative affluence or means of its inhabitants. However, the most desirable neighbourhoods (well maintained and service, relatively free form anti-social behaviour, with the greatest number of nearby amenities, etc.) command a price premium. Therefore, they are disproportionately inhabited by the well-off. So the Deprivation Index does serve as an indirect indicator of the quality of a neighbourhood.

[23] Matching the GeoDirectory to both SAPS and the Pobal HP Deprivation index was a mere formality, as these three datasets all shared Small Area codes.

addresses. However, matching by address is computationally very difficult, given the many variations that can occur in address spellings. Alternating house names with house numbers, using inconsistent abbreviations, Irish language versions, misspellings, etc. all add huge complexity to the problem. Invariably, if computational address matching is to be attempted on a large scale, it has to be done based on fuzzy character string matching. A key question from the outset was could sufficiently reliable matches be obtained between the Stamp Duty, BER and GeoDirectory addresses to compile a viable house price index?

7.2 Address string matching

A variety of different methods exist to electronically match character strings. In the context of matching Stamp Duty address strings to BER and GeoDirectory address strings, an algorithm that measured Jaro-Winkler Distance (JWD) was found most suitable. JWD is a measure of how similar or dissimilar two character strings are; with a score of 0 implying no similarity whatsoever and a score of 1 implying an exact match. JWD was found to be particularly useful in the context of matching Stamp Duty and BER address strings as it allowed extra weight to be given to the first four characters of each string (as most addresses begin with a house number and it is important that these numbers should match exactly). JWD was also used to match the Stamp Duty addresses to the GeoDirectory. Although in this case, lesser weight was given to matching the first four characters, as it was sufficient to match to Small Area only (in this context, matching to the address of the house next door was nearly always good enough).

The address string matching is a multi-step process. Firstly, addresses have to be cleaned on the respective datasets (characters such as commas, apostrophes, full stops, etc. are stripped out, words such as 'road' and 'avenue' are converted to 'rd' and 'av', etc. to facilitate the matching). Next the data in the respective datasets are blocked by county (and post code in the case of Dublin). Address string matching is computationally very resource-intensive and blocking minimises the data matching burden by limiting potential matches to the same block. Next the cleaned Stamp Duty address string is compared with every single cleaned address string from the corresponding blocks of both the BER dataset and the GeoDirectory[24]. JWD is calculated for every pairing. The highest scoring match is automatically presumed to be a correct match, provided the JWD is 0.88 or higher[25].

Typically, for any given month, between 40-50% of all Stamp Duty returns are automatically matched to both a BER record and a GeoDirectory Small Area using this approach. Those that fail to match automatically are searched for manually by CSO staff using a purpose-built data matching interface. This interface lists the top twenty candidate matches ranked in terms of descending JWD. CSO staff then judge which of these candidate pairings, if any, is the correct match.

[24] If there are 100 Stamp Duty addresses in a particular county (block) in a particular reference month and 10,000 BER certificates in for houses in that same block, then there will be 100 x 100,000 = 1 million address pairings, each with their own JWD score. The enormous number of pairings in address string matching make the process computationally resource intensive.

[25] In a sample of 1,000 automatic matches between Stamp Duty returns and BER records subjected to visual inspection, only 7 (0.7%) false positives were found, i.e. cases where the matches were observed to be incorrect. However, the actual false positive rate is likely to be higher than this. In Ireland many rural addresses are non-unique. Thus, where no house number exists, a perfect address string match does not guarantee that a Stamp Duty return has been correctly matched with its corresponding BER record.

In the case of the BER matching, if a BER address is not matched automatically, and if staff cannot manually select a match from the top twenty candidate addresses with confidence, then nothing further can be done. No information is available on the physical characteristics of the dwelling and so the transaction cannot be used in the price model. In the case of the GeoDirectory matching, an alternative tool is available to supplement the matching process. Interactive Small Area Population Statistic Maps (SAPMAPs) are available from Census 2011 results. As a last resort, staff could find the Stamp Duty address street or locality in the SAPMAP application (which also depicts Small Area boundaries). Therefore, it was always possible to match a Stamp Duty address to a particular Small Area.

This manual element of the address string matching proved very time consuming. Starting in 2013, it took a team of three staff over three years to process the backlog of stamp duty returns back to January 2010. By 2016 some 177,000 Stamp Duty returns had been processed in total.

7.3 *Data matching results*

Annual overall match rates between Stamp Duty returns and BER records are given in table 4 below. These match rates combine both the automatic and manual matching phases. Note that 100% of the Stamp Duty returns were successfully matched to a small area (thanks to the SAPMAP application). If we compare the number of successful matches with the number of usable mortgage returns in table 2 above, we can see that even after jumping the hurdle of data matching, we emerge with substantially more matched Stamp Duty returns than usable mortgage drawdowns for each year (except 2010). The question posed at the outset of this project was whether a sufficient rate of data matching could be achieved to compile a house price index. The results of the data matching exercise proved that this was indeed possible.

Table 4: Address matching results, 2010-2016

Year	Stamp Duty Returns	BER Pairings	
		no.	percent
2010	14,830	11,133	75%
2011	13,639	10,105	74%
2012	19,431	15,278	79%
2013	23,163	18,429	80%
2014	32,917	23,177	70%
2015	36,952	27,515	74%
2016	36,290	27,043	75%

8. The treatment of Stamp Duty Data

8.1 The need to treat the Stamp Duty data

Stamp Duty applies to a very broad range of property transactions, not just residential property. The Stamp Duty dataset provided by Revenue Commissioners comprises all Stamp Duty transactions in the state. Therefore, it is necessary to find a mechanism for filtering out all non-relevant Stamp Duty transactions. Furthermore, Stamp Duty returns must be submitted irrespective if the property is sold at market value or not. So there is a need to identify and eliminate non-market transactions. Non-

Household transactions also had to be identified and eliminated. There can also be quality issues in relation to the transaction price recorded for new dwellings, arising from the imposition of Value Added Tax (VAT). Each of these issues needs to be addressed before Stamp Duty data can be used in a house price model.

8.2 *Identifying residential dwelling transactions*

A number of different criteria have be satisfied before an individual Stamp Duty return is considered a unique residential dwelling transaction within scope of the house price index. These criteria are based solely on the information returned on the Stamp Duty return (for a full list of the data fields on a Stamp Duty return please refer to Revenue form SDR1). These criteria are as follows:

1. The property is specifically identified as residential (as opposed to non-residential or mixed use, other available options) – *Item 3.1 of form SDR1*

2. The Stamp Duty return is an original document (as opposed to a substitute for a lost instrument) – *Item 3.2 of SDR1*

3. The property is located in the Republic of Ireland – *Item 12.1 of form SDR1*

4. The type of contract is identified as either a
    a. 'Contract for sale'
    b. 'Contract for a completed new house/apartment'
    c. 'None'

    If the contract is listed as 'Contract of Site with Associated Building Agreement' or a 'Unitary Contract (i.e. Combined Site Contract and Building Agreement)' then the transaction is excluded on the assumption that it is a self-build – *Item 12.2.1 of form SDR1*

5. The type of property is defined as either
    a. New Dwelling House/Apartment
    b. Second-Hand Dwelling House/Apartment

    Curtilage is excluded – *Item 12.2.2 of form SDR1*

8.3 Market transactions

The Stamp Duty data contains an indicator for non-market transactions (item 12.5.1 of form SDR1). However, this indicator was found to be not fully reliable, with many cases of €0 transactions not flagged as non-market, etc. Instead, the following criteria are used to identify non-market transactions:

1. If any relationship, either familial or business, existed between the vendor and the purchaser – *Item 11.2 of form SDR1*

2. If the dwelling was transacted for less than €25,000 – *Item 12.8.1 of form SDR1*

3. If the purchaser availed of any of the following tax reliefs – *Item 15.1 of form SDR1*;

a. Certain Transfers following the Dissolution of a Marriage (S97 SDCA 1999)
b. Housing Authorities & Affordable Homes Partnership (S106B SDCA 1999)
c. Transfer between Spouses (S96 SDCA 1999)

4. Revenue had indicated that the transaction was part of an exchange of properties.

8.4 Non-household transactions

The Stamp Duty data did not contain a specific indicator for identifying non-household transactions. However, the data contain the names and addresses of the purchasers. The names had being captured from form SDR1 in a single character string (Item 10.). However, during processing, Revenue separated the first name from the surname and provided these in separate fields. In cases where the purchaser was a private company, a Local Authority or some other form of institution, Revenue placed this information in the surname field and left the first name field blank. Therefore, any Stamp Duty returns that had a purchaser first name and a surname were automatically considered household transactions. Any Stamp Duty returns that had a blank purchaser first name field were assumed to be non-household transactions.

8.5 The treatment of VAT

All new dwellings sold in Ireland are liable for VAT[26]. Under the EU rules VAT, as a consumption tax, is to be included in the purchase price. However, Stamp Duty is charged on the transaction price exclusive of VAT. For new dwellings, the solicitors acting on behalf of the purchasers are asked on form SDR1 to provide the transaction price of the dwelling with the VAT already deducted (the amount for consideration for Stamp Duty). In some cases, it appeared that the full price (inclusive of VAT) was erroneously provided. To address this problem an assumption was made that most dwellings sold in Ireland have transaction prices in multiples of €1,000. For all new dwellings, if the declared amount for consideration for Stamp Duty was a multiple of €1,000 and the expected transactions price, when adjusted for VAT, was not a multiple of €1,000, then the declared amount for consideration was taken as the full price. Otherwise, VAT was calculated in the normal manner and added to the amount for consideration to derive the full transaction price of the new dwelling.

9. Quality adjustment

9.1 The rolling time dummy hedonic method

As discussed earlier, house price indices must be quality adjusted if they are to measure house price inflation under the HICP regulatory framework. There are a variety of statistical techniques and methods to achieve a constant quality house price index. The method found most successful for the Original RPPI was the rolling year hedonic regression model. One of the strengths of this method is that it pools data over several consecutive periods, which improves the statistical quality of the model outputs[27]. This makes the method particularly suitable for residential property markets such

[26] Currently the VAT rate is 13.5%.

[27] The alternatives to the rolling time dummy approach are either to run a single regression on covering all available time periods or to run the regression on just two consecutive time periods (e.g. as in the double-

as Ireland, where data sparseness is an issue. The Original RPPI used a 12 month rolling time dummy method. Some alternative methodologies were explored for the New RPPI, including the double-imputation method. Ultimately, the 12 month rolling time dummy method was found to perform best and this method was retained for the New RPPI.

The rolling year hedonic regression model employs the following standard log-linear equation:

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

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

In the 12 month rolling time dummy variant, there are eleven time dummies, D_2 to D_{12} (the first month being taken as the reference month). Solving the regression gives estimates for the corresponding eleven time dummy coefficients, δ_2 to δ_{12}. Time dummy coefficient δ_t represents how the natural logarithm of price changes in month t, relative to the reference (first) month, independent of the vector of explanatory variables x_{it} (i.e. the constant quality logarithm of price change over time). For an example of the statistical outputs from a sample regression run, please see the appendix.

In practise, as the New RPPI commences in January 2010, the first regression is run over all dwellings transacted in the year 2010. For this first regression, δ_2 to δ_{12} are the price model estimates of how the logarithm of price changes relative to January 2010, when all other factors are held constant. The antilogs of δ_2 to δ_{12} therefore estimate constant quality price inflation relative to January 2010. A second regression is then run over the 12 months from February 2010 to January 2011. A third regression is run over the 12 months from March 2010 to February 2011, etc. For these second and subsequent regressions, we are only interested in how the price changes in the latest month. This is given by the expression;

---

imputation approach). Running a single regression has the disadvantage that when a new period of data becomes available, and the regression is run again, the whole index is revised, not just the latest month. This is a very significant disadvantage for an official house price index. This approach also assumes that implicit prices for particular dwelling characteristics and locations do not change over time, which is unrealistic for a long time series. Alternatively, running a regression on just two consecutive time periods allows buyer preferences to change month to month. However, this necessarily means a smaller pool of data in the regression model and lest robust estimates for the implicit prices of dwelling characteristics and locations. This can lead to a more volatile index. The rolling time dummy method strikes a balance between these two approaches, increasing the pool of transaction data in the regression month, allowing for some variation in buyer preference for month to month, yet avoiding the need to revise the whole price index each month. The rolling time dummy window can be any length, e.g. 24 months, 36 months, etc. However, a 12 month rolling time dummy windows is usually preferred.

$$I_{12} = \frac{e^{\delta_{12}}}{e^{\delta_{11}}} . I_{11}$$

where

I_{12} is the index for the latest month

I_{11} is the index in the month prior to the latest month.

9.2 Stratification

The 12 month rolling time dummy hedonic method described above provides an estimate of the aggregate price changes over time, at constant quality, of all the dwellings in the pool of data to which it is applied. To get separate quality-adjusted price changes for separate market segments, stratification is required. Stratification, in this context, is just dividing the pool of dwelling transactions into various sub-pools, based on various characteristics, and running separate regressions over each. This has the advantage of allowing different estimated price changes for the different market segments[28]. However, it has the attendant disadvantage of running each regression on a much diminished the pool of data with implications for the stability of the results.

The greater number of matched Stamp Duty returns compared to usable mortgage drawdowns allowed a much more extensive degree of stratification for the New RPPI than for the Original RPPI. There were sufficient transactions of houses available within County Dublin to permit separate strata for each of the four Dublin administrative districts. And there were sufficient transactions of houses outside of Dublin to permit separate strata for each NUTS3 region. The final stratification scheme was thus expanded to 13 regression models and 13 corresponding elementary indices. This stratification scheme is depicted in table 5 below, along with the aggregate indices which result.

9.3 Variable selection

A key consideration in defining any hedonic regression model is selecting the appropriate explanatory variables for the model. In a house price index, the explanatory variables describe the physical attributes and locations of the dwellings (the x_{it} vector of characteristics in our log-linear regression formula). Only variables that have a statistically significant impact on the dependent variable, house price in this case, should be included[29]. In the case of the New RPPI, the multiplicity of linked data sources provided a plethora of variables showing a statistically significant relationship to price. For example, whether the dwelling was new or existing, the building energy efficiency, the type of central heating system, the number of stories, whether the dwelling was in an urban or rural Small Area, in a coastal or non-coastal Small Area, distance to the nearest city centre, etc. In practise, to keep the price model simple, just four explanatory variables were used, two variables describing the physical characteristics of the building and two locational variables. These were the

[28] It also has the advantage of allowing different β coefficients to apply to the different market segments, which should theoretically imply a more nuanced overall approach.

[29] Other issues, such as the appropriate transformations of the variables and collinearity concerns are also considerations.

total floor area of the dwelling (m^2), the dwelling type (i.e. detached, semi-detached, terraced, etc.), the Eircode routing key and the Pobal HP Deprivation Index (see table 6 below).

Table 5: New RPPI Price Indices

Code	Name	Type	α,β[1]
I1	National Index	Aggregate	0.5
I1.1	Dublin residential dwellings	Aggregate	0.5
I1.1.1	Dublin houses	Aggregate	0.5
I1.1.1.1	Dublin City houses	Elementary	0.3
I1.1.1.2	Dún Laoghaire-Rathdown houses	Elementary	0.3
I1.1.1.3	Fingal houses	Elementary	0.3
I1.1.1.4	South Dublin houses	Elementary	0.3
I1.1.2	Dublin apartments	Elementary	0.3
I1.2	National (excluding Dublin) residential dwellings	Aggregate	0.5
I1.2.1	National (excluding Dublin) houses	Aggregate	0.5
I.1.2.1.1	Border houses	Elementary	0.2
I.1.2.1.2	Midland houses	Elementary	0.2
I.1.2.1.3	West houses	Elementary	0.2
I.1.2.1.4	Mid-East houses	Elementary	0.2
I.1.2.1.5	Mid-West houses	Elementary	0.2
I.1.2.1.6	South-East houses	Elementary	0.2
I.1.2.1.7	South-West houses	Elementary	0.2
I1.2.2	National (excluding Dublin) apartments	Elementary	0.3
I1.A	National houses	Aggregate	0.5
I1.B	National apartments	Aggregate	0.3

1: α and β values are discussed later in the section *Data Smoothening*

The Eircode routing key is the postal sorting district. There are 139 such districts in Ireland. Eircode routing keys offered much improved geographical granularity over the county breakdown in the Original RPPI (there being just 26 counties in Ireland). *The Pobal HP Deprivation Index* has already been discussed. The advantage of using the Deprivation index is that it allows for further locational differentiation *within* the Eircode routing key areas. Therefore, the combination of Eircode routing keys and the Deprivation Index was a powerful means of capturing the effect of location on house prices, satisfactorily addressing this particular weakness in the Original RPPI.

Table 6: New RPPI Explanatory Variables

Source	Variable	Categories
BER Data	Floor area (m^2)	Continuous variable
	Dwelling type	Detached, Semi-Detached, Mid-Terraced, End-Terraced, Apartment
GeoDirectory	Eircode Routing Key	A41, A42, A45, A63, A67, etc.
Pobal	Pobal HP Deprivation Index	Continuous variable

9.4 Outlier detection

Before running the hedonic regression to estimate the time dummy coefficients δ_{12} and δ_{11}, outliers are first identified and removed from the pool of transaction data. Outliers are in fact identified from a preliminary hedonic regression run, using the exact same price model. In this preliminary regression run Cooks Distance is computed for each transaction. Cooks Distance is a measure of the leverage each transaction has on the overall regression fit (i.e. how influential a particular transaction is in the determination of the 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 characteristics). Any transactions where the Cooks Distance exceeds (*4/n*), the conventional cut-off (where *n* is the number of transactions in the data pool), is considered an outlier. These outliers are then excluded from the final regression run. Table 7 below shows the percentage of transactions identified as outliers for each stratum for each of the 12 regression runs in 2016[30]. These range from a minimum of 1.7% for the stratum 'South Dublin houses' in Jun 2016 to a maximum of 10.2% for the stratum 'National (excluding Dublin) apartments' for October 2016. The median percentage across all strata for 2016 was 5.5%.

Table 7: Outlier Frequency 2016

Code	Name	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
I1.1.1.1	Dublin City houses	2.1%	4.6%	4.9%	5.6%	3.9%	6.5%	3.1%	3.8%	2.1%	5.2%	3.1%	3.8%
I1.1.1.2	Dún Laoghaire-Rathdown houses	5.7%	5.7%	3.3%	9.3%	6.5%	5.8%	8.3%	2.7%	3.6%	2.9%	3.1%	10.1%
I1.1.1.3	Fingal houses	4.4%	7.5%	4.8%	5.9%	7.0%	3.1%	4.7%	4.3%	2.1%	3.5%	8.3%	8.2%
I1.1.1.4	South Dublin houses	3.2%	5.9%	6.6%	2.0%	2.4%	1.7%	4.4%	3.1%	5.6%	3.3%	5.4%	4.9%
I1.1.2	Dublin apartments	2.5%	4.2%	4.6%	3.4%	6.5%	2.2%	8.6%	3.9%	5.9%	6.7%	3.0%	5.1%
I.1.2.1.1	Border houses	10.0%	3.7%	6.8%	9.7%	4.1%	7.1%	2.7%	7.1%	5.6%	5.0%	3.2%	2.3%
I.1.2.1.2	Midland houses	3.4%	6.3%	5.3%	7.3%	4.4%	8.0%	7.3%	7.2%	3.2%	4.6%	6.2%	6.5%
I.1.2.1.3	West houses	6.1%	4.8%	2.4%	8.1%	4.4%	6.3%	4.6%	5.5%	8.0%	7.8%	7.8%	6.8%
I.1.2.1.4	Mid-East houses	5.4%	2.2%	6.4%	6.6%	6.1%	6.2%	6.6%	3.9%	4.7%	5.0%	5.0%	4.2%
I.1.2.1.5	Mid-West houses	5.2%	6.1%	7.6%	4.3%	9.7%	7.9%	5.9%	6.0%	4.3%	6.2%	5.1%	7.6%
I.1.2.1.6	South-East houses	6.0%	3.5%	3.5%	6.9%	4.1%	8.0%	5.1%	9.2%	5.1%	7.8%	6.4%	2.5%
I.1.2.1.7	South-West houses	9.6%	6.2%	4.0%	7.2%	8.1%	7.3%	6.6%	5.5%	5.8%	6.0%	5.2%	5.1%
I1.2.2	National (excluding Dublin) apartments	5.6%	8.2%	2.4%	9.8%	7.5%	8.3%	6.8%	4.9%	5.1%	10.2%	7.4%	4.2%

[30] Note: The Jan 2016 results are based on the February 2015 to January 2016 rolling time window, etc.

10. Weighting

10.1 The weighting scheme

To derive aggregate indices from elementary indices a weighting scheme is required. The weights define the relative importance of each elementary index in the compilation of the aggregate. The weights are in turn derived from the relative expenditures under each elementary heading. Only market-based transaction expenditures are used in the weighting scheme. The RPPI is a *Laspeyres-type* price index. This requires that the weights are derived in a base period and are held constant in the following reference periods. In practise, the weights are annually updated, based on expenditures in the previous year. As there are 13 elementary indices, the national RPPI for any given month is given by the formula;

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

where

$RPPI_{y,m}$ is the national aggregate index in year *y* and month *m*

$I_{n,y,m}$ is elementary index *n* in year *y* and month *m*

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

The aggregate sub-national indices are calculated in an analogous manner. The expenditure weights are calculated on all reported market-based dwelling transactions, both matched (to a BER certificate) and unmatched.

10.2 Dwelling type imputation

The system of weighting requires separate expenditures for house and apartments. Houses and apartments are clearly distinguishable from each other in the matched Stamp Duty/BER data, by the BER variable *Dwelling type*. But what about the approximately one in four Stamp Duty returns for which a reliable BER match cannot be found? In these specific cases it is necessary to impute the basic dwelling type before the weights can be derived. This imputation follows a four-step deterministic approach as follows:

1. If the address string contains one of the words ‘apartment’ , ‘apt.’, ‘flat’, ‘block’, ‘floor’ or ‘condominium’ the dwelling is automatically assumed to be an apartment. If the address string contains the word ‘cottage’ the dwelling is automatically assumed to be a house.

2. If the address string contains a key word associated with a list of known large apartment-type complexes in a particular county (e.g. the ‘Gasworks’, ‘City West Plaza’, ‘Isoldes Tower’, ‘Longboat Quay’ in Co. Dublin, etc.) then it is automatically assumed to be an apartment[31].

[31] The list of known apartment complexes was originally derived from the BER data. The address strings of apartments in the BER data were scrutinised. Where the same keywords appeared 3 or more times these were added to the list.

3. If, according to the latest census data, 95% or more of residential dwelling types in the Small Area were apartments, then the dwelling is automatically 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 applied, where the category of instrument is listed as 'Long-Term Lease greater than 100 years', then the dwelling is automatically assumed to be an apartment. Otherwise, if the category of instrument is listed as 'Conveyance/Transfer of Property' then the dwelling is automatically assumed to be a house.

Table 8 below shows the relative importance of each of these steps in imputed the dwelling type for the years 2010 to 2014. Invariably, most dwelling types were imputed in step 3 on the basis of their small area characteristics (a relatively large number of rural small areas in particular tend to contain houses only). The proportion of apartments imputed is also given along with the 'known' proportion of apartment transactions from the matched data. Generally speaking, the proportion of apartments imputed is in line with expectations.

Table 8: Dwelling type imputation

Year	Unmatched Stamp Duty Returns[1]	Imputation Step Assignments				Apartment Imputation	Matched Apartment Frequency[2]
		Stage 1	Stage 2	Stage 3	Stage 4		
2010	3,697	10.9%	3.4%	64.5%	21.2%	12.5%	9.8%
2011	3,550	10.1%	2.4%	65.9%	21.6%	10.1%	9.6%
2012	4,196	10.1%	2.9%	61.9%	25.1%	10.1%	10.8%
2013	4,787	11.8%	4.6%	57.3%	26.2%	12.4%	14.1%
2014	9,883	12.6%	5.8%	56.8%	24.8%	15.9%	14.0%
2015	9,511	12.7%	5.8%	54.1%	27.5%	16.2%	14.8%
2016	9,246	12.4%	5.4%	53.9%	28.3%	16.0%	15.7%

1: Market-based residential dwellings purchased by households for which a reliable BER match could not be found.

2: The proportion of market-based residential dwellings purchased by households known to be apartments from the matched data.

10.3 The New RPPI weights

The annual elementary index weights used in the new RPPI for the years 2010 to 2016 are given below in table 9. The weights for the years 2010 and 2011 are the same as both are based on 2010 expenditure data (in the absence of reliable Stamp Duty data for 2009). Generally speaking, the weights remain broadly consistent over the period concerned. The index for *Dublin City houses* consistently has the highest weighting, ranging from 15.3% to 18.2% of the total expenditure. *Dún Laoghaire-Rathdown houses* consistently have the second highest weightings. The smallest weightings are observed for *National (excluding Dublin) apartments*, for the years 2010 to 2013 and for *Midland houses* for 2014 to 2016.

Table 9: Weighting scheme for the new RPPI

Code	Name	2010-2011	2012	2013	2014	2015	2016
I1.1.1.1	Dublin City houses	15.7%	17.5%	18.2%	17.6%	16.7%	15.3%
I1.1.1.2	Dún Laoghaire-Rathdown houses	12.0%	12.8%	14.4%	14.3%	12.0%	10.6%
I1.1.1.3	Fingal houses	8.4%	7.2%	7.9%	8.4%	8.0%	8.1%
I1.1.1.4	South Dublin houses	7.5%	6.0%	5.8%	6.2%	6.2%	5.9%
I1.1.2	Dublin apartments	6.7%	6.1%	6.7%	8.7%	9.5%	9.8%
I.1.2.1.1	Border houses	5.2%	5.4%	5.0%	4.7%	5.0%	5.4%
I.1.2.1.2	Midland houses	2.9%	2.8%	2.6%	2.3%	2.7%	2.8%
I.1.2.1.3	West houses	5.6%	5.8%	5.4%	5.5%	5.3%	5.9%
I.1.2.1.4	Mid-East houses	11.1%	10.5%	10.0%	10.4%	11.6%	11.3%
I.1.2.1.5	Mid-West houses	4.9%	5.0%	4.6%	3.9%	3.8%	4.3%
I.1.2.1.6	South-East houses	6.4%	6.8%	6.0%	5.4%	5.7%	6.0%
I.1.2.1.7	South-West houses	11.4%	11.9%	11.4%	9.9%	10.3%	11.1%
I1.2.2	National (exc. Dublin) apartments	2.1%	2.2%	2.0%	2.6%	3.3%	3.6%

10.4 Comparison of New and Original RPPI weights

It is interesting to compare the weights for the New RPPI, based on the Stamp Duty returns, with those for the Original RPPI, based on mortgage transaction data. The New RPPI is more disaggregated than the Original RPPI. But these New RPPI weights can readily be aggregated together into the categories of the Original RPPI. Table 10 below provides the comparable weights for the New and Original RPPI's for the years 2010 to 2016. This comparison shows that the Original RPPI weights tended to overvalue expenditures on houses outside of Dublin. Conversely, the Original RPPI weights tended to undervalue all other expenditures, particularly for apartments. It is not clear exactly why this is the case at this stage (this is a matter for further investigation)[32]. But it is clear that these weight differences affect the overall *National Index* compilation. For example, Dublin prices fell further than the price indices for outside of Dublin in the trough period of 2011-2013. As the Dublin price indices are now given additional weight in the new RPPI, this applies further downward pressure on the new RPPI national index for this period.

[32] For apartments, the case can be made that (1) they are generally cheaper than houses and (2) they are more likely to be purchased by households as a rental investment. Both factors imply that apartments are less likely to be purchased with a mortgage than houses, which would explain their lower weighting in the existing RPPI (being based on mortgage transaction data). However, this logic breaks down when applied to Dublin houses. Dublin houses are generally more expensive than houses outside of Dublin, which would suggest that they should be over-represented in the mortgage transaction data if anything. However, a second factor at play may be self-builds. Self-builds are included in the mortgage data but not in the Stamp Duty transactions (self-builds are only eligible for stamp duty for site purchase). Self-builds are primarily a rural phenomenon. The presence of self-builds in the mortgage data are probably therefore inflating the weights for the National (excluding Dublin) houses index. The interplay between these two factors most likely accounts for the differences in weights between the original and new RPPI.

Table 10: Comparison of weighting scheme for the existing and new RPPIs

Name	2010	2011	2012	2013	2014	2015	2016
Original RPPI							
Dublin houses	32.2%	35.4%	39.3%	46.4%	42.9%	40.3%	41.7%
Dublin apartments	5.3%	3.1%	2.9%	3.9%	4.6%	5.8%	5.7%
National (excluding Dublin) houses	61.0%	60.7%	57.0%	48.7%	51.4%	52.3%	51.1%
National (excluding Dublin) apartments	1.6%	0.8%	0.8%	1.0%	1.0%	1.6%	1.5%
New RPPI							
Dublin houses	43.6%	43.6%	43.5%	46.3%	46.5%	42.9%	39.9%
Dublin apartments	6.7%	6.7%	6.1%	6.7%	8.7%	9.5%	9.8%
National (excluding Dublin) houses	47.5%	47.5%	48.2%	45.0%	42.1%	44.4%	46.8%
National (excluding Dublin) apartments	2.1%	2.1%	2.2%	2.0%	2.6%	3.3%	3.6%

11. Data smoothing

11.1 The need for smoothing

As discussed earlier, as the number of transactions used in the price models decreases, the volatility or statistical 'noise' of the resulting price index increases. This statistical noise can make it very difficult to identify turning points in the residential market in a timely manner. This noise creates a conundrum in price index development. On the one hand, it is desirable to restrict price indices so that there is a sufficiently large volume of transactions available to minimise the noise to acceptable levels. On the other hand there is a demand from users for price indices at the lowest possible levels of disaggregation. A balance needs to be struck between these conflicting objectives. Data smoothing is a technique which helps address this situation.

11.2 Three month rolling averages

The data smoothing technique used in the Original RPPI was the three month rolling average technique. This technique has the advantage of simplicity. But by the same token it is a relatively crude method of data smoothing. And, as we have discussed already, it has the particular disadvantage of effectively introducing a one-month time lag in the price index. Even after applying a three-month rolling average, the original *National (excluding Dublin) Apartments* price index remained too volatile to publish. Therefore, an alternative data smoothing technique was sought for the New RPPI.

11.3 Double exponential smoothing

The particular smoothing technique adopted for the New RPPI is the *Holt-Winters double exponential* method. In this method, the smoothing is achieved as follows: For the first observation in a data 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 advantage of double exponential smoothing is that it gives a wide degree of discretion over the degree of smoothing to be applied, through choice of α and β values. It allows the tailoring of the smoothing to the needs of a specific price index, for example. However, another important advantage is the double smoothing method specifically factors the trend into the data smoothing process. Thus, dependent on the choice of β value, the smoothing of a particular data point can be made more reflective of the long term trend. This is important because incorporating the trend compensates for the time lag effect incidental to some other smoothing techniques (such as the three month rolling average). With the appropriate choice of α and β a smoothed price index with no implicit time lag can be approximated.

The disadvantage of any data smoothening technique, including double exponential smoothing, is that responsiveness to real change is dampened. If the data is over-smoothed then there is a risk of rendering the price index too sluggish. Therefore, a trade-off must be made between reducing volatility whilst retaining a sufficient level of responsiveness.

The α and β values used in the various New RPPI indices were given in table 5. These values were chosen based on empirical testing[33]. For the aggregate indices (covering more transactions and thus

[33] Mathematical techniques can be applied to find the optimum α and β values. However, these optima would differ across all the various indices and theoretically should be recalculated every month as the index is updated. To avoid this level of complexity, a decision was made to set α and β equal to each other (to give equal weighting to the data points and the trend), to limit them to one decimal place and to keep them constant for each index.

with less inherent volatility) the α and β values are larger, as less smoothing is required[34]. These α and β values are considered to provide a reasonable compromise between reducing volatility and maintaining responsiveness. Without imposing these constraints, indices could not be published for some of the less well populated strata, particularly *National (excluding Dublin) Apartments* and some of the regional series for houses outside of Dublin.

11.4 Data smoothing comparisons

Figure 4 below shows the effect of smoothing on the New RPPI *National Index*. The blue series is the price index prior to any data smoothing (the 'raw' index). The red series is the index smoothed by the 3 month rolling average technique. This is certainly less volatile than the raw series but it can be seen that when the raw index is either falling or rising, the red series is shifted slightly to the left. This is the 'lag effect' in action and this effectively leads to a one month delay in identifying significant turning points in the market. The green series is the index smoothed by *the Holt-Winters double exponential smoothing* technique. Again, this green series is noticeably less volatile than the raw index. The green series does show some evidence of a lag effect. However, this effect is much less pronounced than in the green series, almost negligible for the most part, so the responsiveness of the index is not significantly compromised.

Figure 4: New RPPI 'Raw' and Smoothed - *National Index*

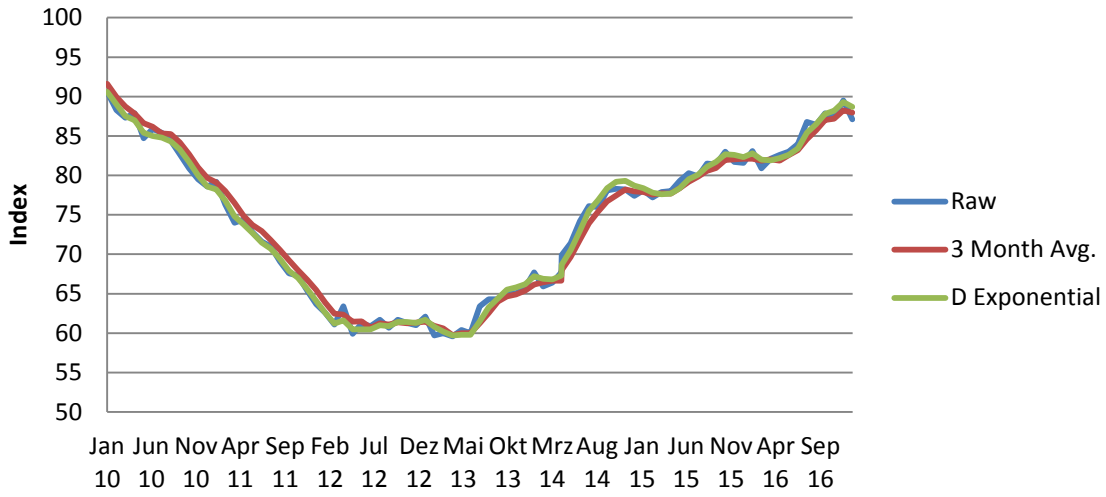
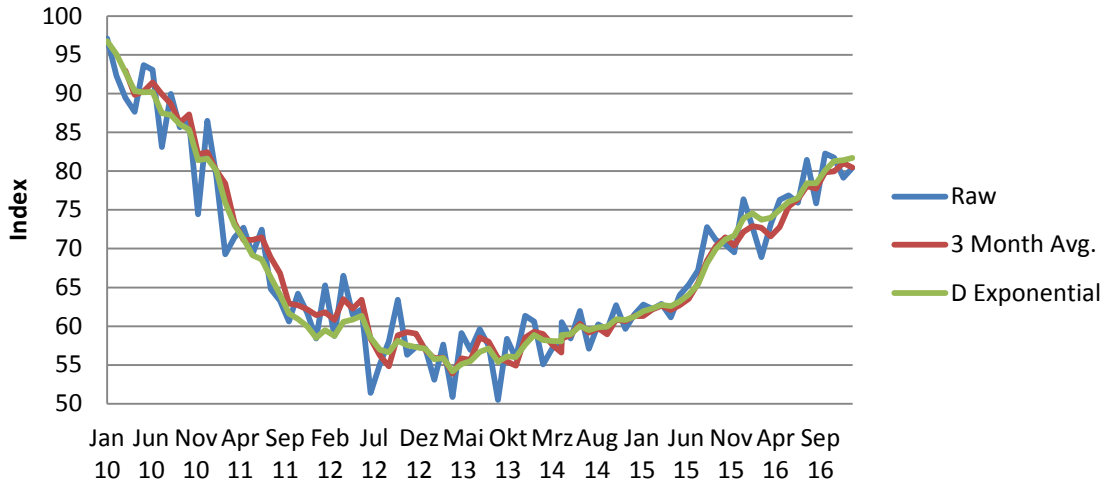


Figure 5 below shows the effect of data smoothing on the New RPPI *Midland Houses* price index. This is a particularly volatile price index given the relatively small number of transactions in this region over the period concerned (there were just 728 Midland houses purchased by households on the open market in all of 2010, rising to a figure of 2,185 in 2016). As can be seen, the raw series oscillates considerably from month to month. In the circumstances, it is impossible to draw any conclusions on the latest trend from comparing the current month to the previous month. Publishing the latest raw monthly percentage change, or even 12 monthly percentage change, could prove very

[34] Both the application of data smoothing techniques and the fact that different α and β values differ for the various New RPPI indices result in a situation where the aggregate price changes do not necessarily equal the sum of the weighted price changes of its component elementary indices in some particular months. This is inconvenient, but is considered a price worth paying for publishing the New RPPI price indices at such a disaggregated level.

misleading. The red series, smoothed by the three month rolling average technique, reduces the volatility, albeit again with some evidence of a lag effect. The green series, smoothed by the double exponential technique, dampens volatility further, without a noticeable concomitant lag effect. It produces a series that is arguably more reflective of the overall trend on a month-by-month basis, and therefore more suitable for publication.

Figure 5: New RPPI 'Raw' and Smoothed - *Midland Houses*



12. Data progressivity

12.1 What is data progressivity?

Typically statistics are compiled from static datasets. A survey is conducted and data is collected, processed, edited and held in a fixed final dataset referring to a specific reference period or point in time. All results are usually compiled from that fixed final dataset and, once published, are unchanging. However, this is not necessarily the case when compiling statistics from administrative data streams. Some administrative datasets are fixed, final datasets referring to a specific period in time. However, other administrative datasets are 'live' administrative datasets, constantly being updated with new information. Of course, a snapshot can be taken of a particular moment in time and statistics can be produced on that basis. But what if subsequent additions to the administrative dataset can lead to substantially different results? Should new results be recalculated as new information comes in? If so, how frequently should these results be updated? This is the data progressivity issue and it can pose a real challenge for the compilation of official statistics.

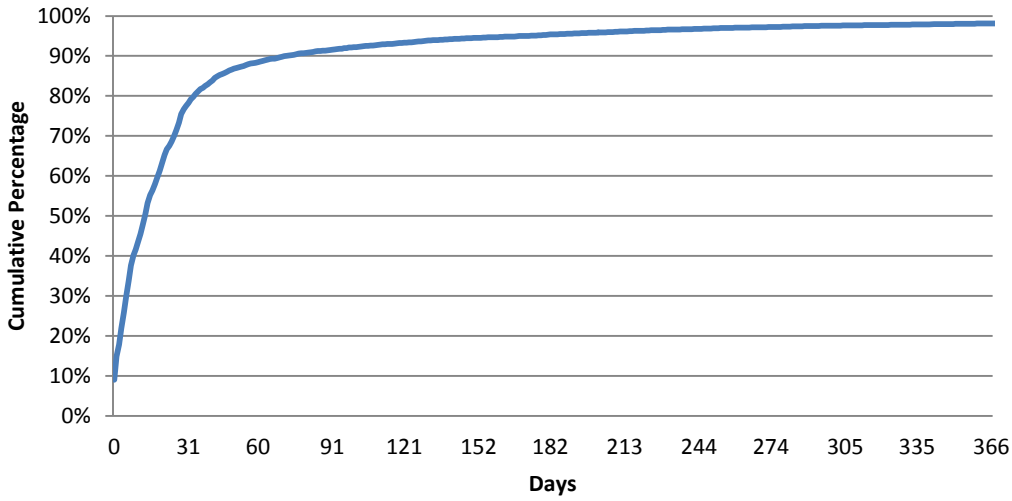
12.2 Stamp Duty data progressivity

The Stamp Duty data collected by Revenue are progressive data[35]. Under current Revenue rules, the solicitor for the purchaser in a property transaction is obliged to submit a Stamp Duty return to

[35] Technically, both the BER data and the GeoDirectory are progressive data also, in that they are 'live' data subject to accumulation and change respectively (the SAPS data and the Pobal HP Deprivation index derived from these are not). In practise these do not pose progressivity problems for the new RPPI. BER certification is necessary before a dwelling is offered for sale, usually several months before the transaction date. Delayed

Revenue Commissioners within 44 days of the transaction being executed. In practise, in some cases, the form may be submitted weeks or months after this deadline (or even years later on occasion). Every month the Revenue Commissioners provide the CSO with a dataset referring to the previous month's transactions. However, this is a dataset of all transactions *filed* with Revenue in the previous month. It is not, and cannot be (given the 44 day deadline) a dataset of all transactions *executed* in the previous month. In fact most of the transactions filed in a given reference month will have been executed in prior months. Figure 6 below shows the accumulated rate of return for Stamp Duty returns filed for household market-based residential dwelling transactions executed in 2010. Some 85% of the returns were filed within the 44 day deadline. However, even after a full year had elapsed, approximately 2% of the returns still had not yet been filed. Several returns were filed a full five years after the date of execution[36].

Figure 6: Filing rate of Stamp Duty returns for dwelling sales executed in 2010



12.3 Provisional and final results

The Stamp Duty progressivity poses a dilemma for the compilation of the New RPPI. On the one hand, the progressivity issue would normally imply that the compilation of the index for a particular reference month should be delayed until a sufficient accumulation of returns for that month to ensure that the index is fully stabilised, and will not be materially affected by subsequent late returns. On the other hand, this would seriously delay the publication of the New RPPI. And any such delay erodes the usefulness of the RPPI as a timely indicator of changing market trends.

To resolve this dilemma, a staged approach is taken to the publication of the New RPPI. Provisional results are initially compiled and published based on the first tranche of Stamp Duty data received for the reference month. The following month, when a new tranche of data is received, these results

---

BER certification does not seem to be any issue, at least as far as the New RPPI is concerned. Whilst the GeoDirectory may change the number of buildings in a particular Small Area over time, the Small Area itself does not change, so this has no real implications for the new RPPI.

[36] Allowance should be made for the fact that eStamping was only introduced in 2010 and that some settling down period may have been necessary. This rate of compliance may have improved in subsequent years (although it is not possible to make an equivalent comparison at this point in time).

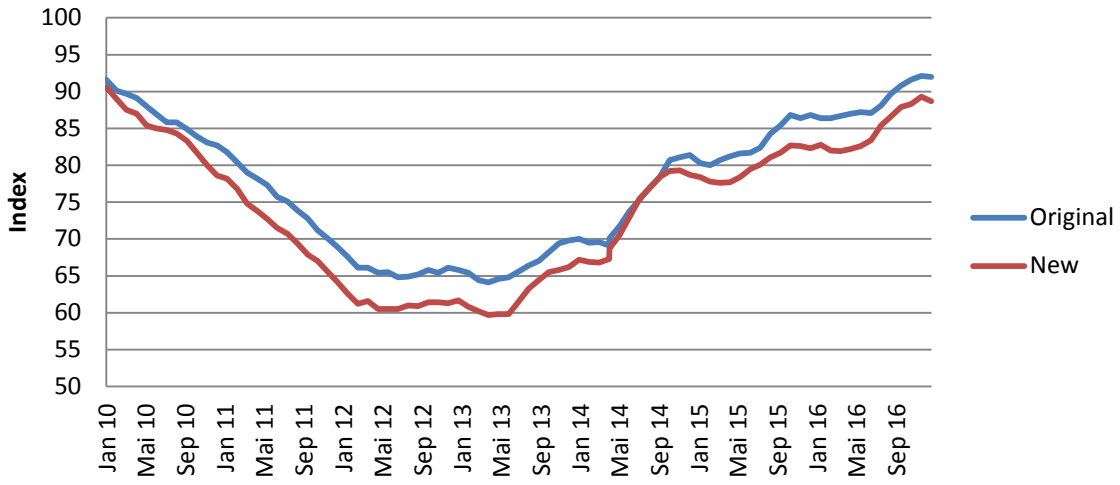
are updated. It is only in the following month again, when yet another tranche of data is received, that the results are finalised. At this stage typically some 90% of the Stamp Duty returns for the reference month have been filed and any subsequent returns are unlikely to materially affect the price trend.

13. Differences between the Original and New RPPI

13.1 Differences at national level

Inevitably, given the different data sources, price models and techniques employed, there are differences between the New RPPI and the Original RPPI. Figure 7 below shows the differences between the New and Original *National Index* for the period 2010 to 2016[37]. Here quite substantial differences emerge. The first obvious difference is that the New RPPI shows a markedly greater decline than the original RPPI, particularly in the period 2012-2013, the trough of the recent housing market cycle. Both indices reached their nadir in March 2013. The Original RPPI dipped as low as 64.1 in this month. However, the New RPPI dipped lower again, reaching 59.7, 4.4 index points lower. Assuming that the New RPPI is the more correct figure for this month, this implies that residential property prices fell further than originally thought. According to these revisions, the overall peak-to-trough decline was 54.4%, not 50.8%, as originally reported.

Figure 7: Comparison of New and Original RPPI's - *National Index*



A second notable feature of the comparison is that although the New and Original *National Index* converged again after the trough for much of 2014, they began to diverge again from September 2014 onwards, with the New index being again significantly lower. This divergence grew towards the end of 2015 in particular. In fact, in December 2016, the latest month for which final results are available, the New *National Index* is 88.7 whereas the Original *National Index* would have been 92.0.

[37] As the Stamp Duty returns are only available from January 2010 onward, this is the first month that the New RPPI commences. For the sake of consistency, the New RPPI indices for January 2010 were set to those calculated for the Original RPPI in January 2010.

This suggests that the Original RPPI was significantly over-estimating the extent of the recovery in residential property prices when compared to their peak levels in 2007.

Apart from the divergence issue, both the New and original RPPI compare very favourably. Both show the same double-dip type of trough. Both show that the market trajectory changed significantly around September 2014. In many respects, this is remarkable given that the two indices are compiled from entirely separate data sources. This supports the hypothesis that the Original RPPI, based on the mortgage transaction data, was reasonably sound and captured quite well the trends and turning points in residential property prices. Only in the degree to which residential property prices fell in the trough period, and rose subsequently from the end of 2014 onwards, did real differences emerge. And these differences, whilst certainly significant, are not dramatic by any means.

13.2 Accounting for the differences

Accounting precisely for the differences between the New and Original RPPI is no easy task given the multiplicity of factors involved. The New RPPI has different data sources, a different scope, different price models, different weights, and a different smoothing technique. At this stage, without further detailed investigation, we cannot be definitive on the precise contribution or relative importance of the various factors involved. However, we can throw some light on the issue by comparing the New and Original RPPI at a more disaggregated level.

Figure 8 below compares the New and Original RPPI sub-index for the *Dublin Residential Dwellings*. Whilst this comparison shows significant divergence between this New and Original sub-index in the early part of the series, from early 2010 to mid-2012, from this point onwards the two sub-indices show a truly remarkable degree of similarity. This strongly suggests that the main factors promoting divergence between the New and Original *National Index* are principally operative in the segment of the residential property market outside of Dublin.

Figure 8: Comparison of New and Original RPPI's - *Dublin Residential Dwellings*

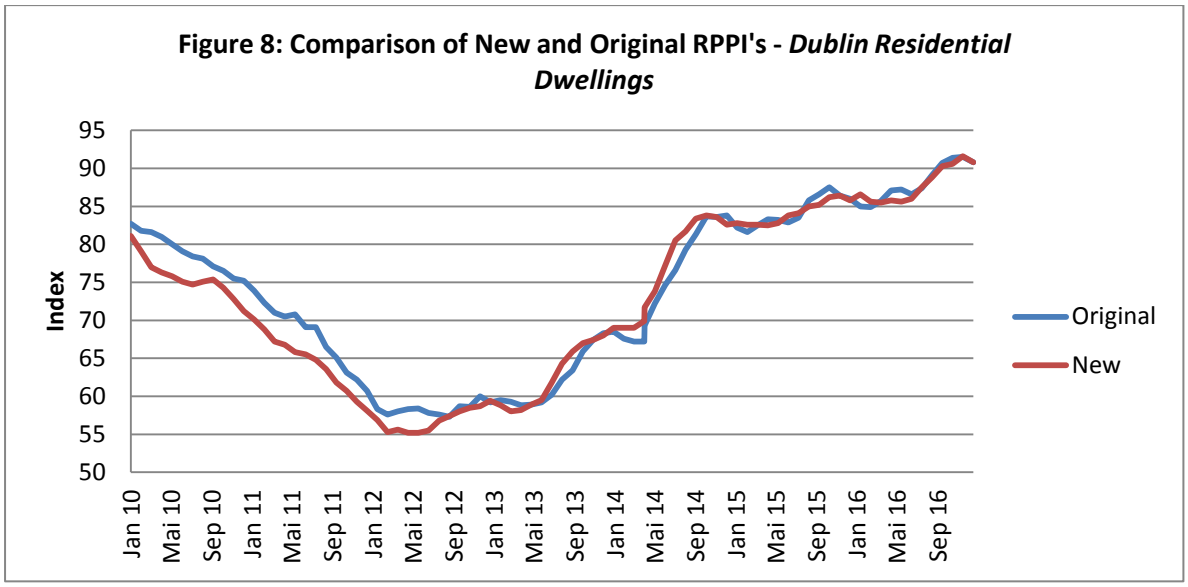
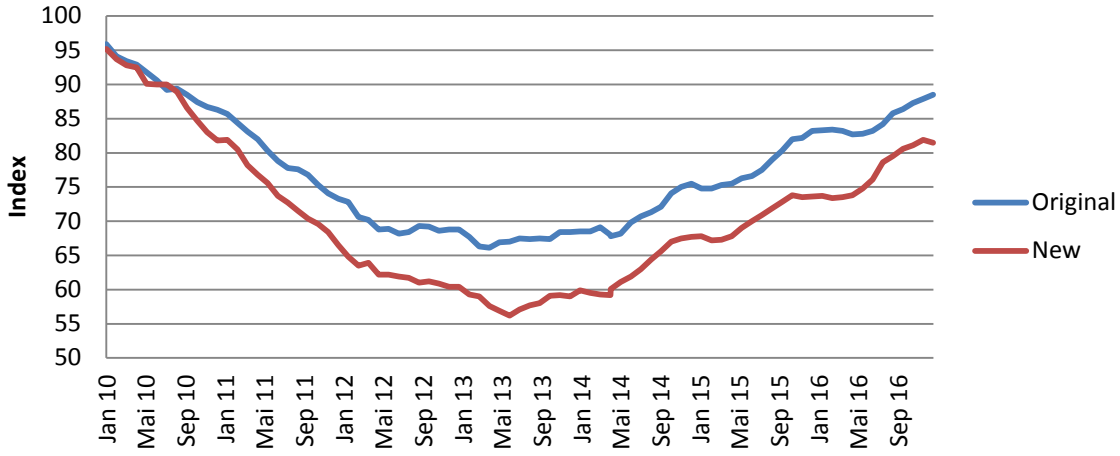


Figure 9 below compares the New and Original RPPI sub-index for *National (excluding Dublin) Residential Dwellings*. Here a very strong divergence can be seen between the two sub-indices. In May and June 2013, where the divergence is greatest, the New *National (excluding Dublin)*

Residential Dwellings sub-index is more than 10 points lower than the Original sub-index. Again this suggests that the factors driving the differences between the New and Original National Index are primarily phenomena at play outside of Dublin. Furthermore, given the relatively small weightings for apartments outside of Dublin (see table 10), we can be confident that these are principally issues related to the purchase of houses.

Figure 9: Comparison of New and Original RPPI's - *National (excluding Dublin) Residential Dwellings*



One factor that is possibly affecting the New *National (excluding Dublin) Residential Dwellings* is the improved capture of locational price effects in the New price model. We have already discussed that concerns over locational differentiation were one of the factors that prompted the development of the New RPPI. This was principally a concern outside of Dublin, where a county code was the only locational differentiation available (recall that within Dublin there existing a post code system that allowed for considerable differentiation between different areas of the city). The new Eircode routing keys combined with the Pobal HP Deprivation Index allows a much greater degree of locational differentiation outside Dublin than was hitherto possible. This would not necessarily affect the *National (excluding Dublin) Houses* sub-index trend if the aggregate quality or attractiveness of the locations where houses were purchased remained consistent throughout the 2010 to 2016 period. However, if there were a drift towards purchasing houses in less desirable areas over this period, then this would apply downward pressure on the sub-index[38]. This is one avenue for further investigation.

A second avenue for investigation is the influence of self-builds. We know that some self-builds at least were included in the Original RPPI and that these are principally a rural phenomenon (i.e. outside Dublin). Any price effect from excluding self-builds, as the New RPPI does, is therefore going to principally affect the *National (excluding Dublin) Houses* sub-index (apartments are never self-built). Therefore, if the price trend of self builds differed from the price trend of regular house purchases, or if the proportion of self-builds in the mortgage data changed over time, this could have a significant impact on the Original *National (excluding Dublin) Houses* sub-index. Unfortunately, there is very limited information on self-builds in Ireland at the moment but it this

[38] If houses are being purchased in more desirable locations then the quality of the dwelling is increasing. A properly functioning constant quality price index compensates for this by deflating the prices paid.

situation can be addressed, then the effect of self-builds on the Original sub-index could be quantified[39].

A third consideration is the impact of including cash-buyers in the New RPPI. Cash-buyers generally purchase smaller or cheaper properties and generally purchase proportionately more buy-to-lets[40]. In principal, this should not matter in a quality adjusted price index such as the RPPI, which takes into account dwelling size and location. However, it may be that cash-buyers pay less than mortgage buyers for the exact same dwelling. Preliminary evidence suggests that this is the case in Ireland at least, particularly under distressed market conditions[41]. If this is true then either a variation in the proportion of cash-buyers or a variation in the implicit discount they receive would affect the trajectory of the New RPPI. We already know that the proportion of cash-buyers in the market increased from 2010 onwards. This factor alone could also have contributed to downward pressure on the New *National (excluding Dublin) Residential Dwellings* sub-index. Again, this is an issue that requires more detailed investigation.

Other factors are also likely to have some bearing on the differences between the New and the Original RPPI price trends. For example the Original RPPI price model included differentiated between new and existing dwellings and different types of buyers. The omission of these variables from the New RPPI price model could conceivably affect the price trend over time. The lessor volumes of transactions available for the Original RPPI may also have impacted the trend. Data quality issues may have had an impact. We have already seen the different weighting schemes applying to the New and Original RPPI. In particular, the Dublin geographical region has greater weighting in the New RPPI. This revised weighting scheme affects the *National Index* at least. We have also seen that different approaches to smoothing result in small but significant differences. Ultimately, all of these various factors need to be fully investigated in turn to arrive at a fully comprehensive account for the differences between the New and Original RPPI's.

14. Conclusion

14.1 Addressing concerns for the Original RPPI

[39] The only definitive information in Ireland on self-builds is the number of planning permits granted. In 2010, for example, there were 5,582 planning permits granted for self-builds. The total number of dwellings purchased by households on the open market in 2010 was 14,830. So self-builds are very significant in this context.

[40] For example, in 2016 the median mortgage purchase price was €250,000, according to the mortgage data available to the CSO. However, the median purchase price recorded in the Stamp Duty returns, which include both cash and mortgage transactions, was just €190,000. It can be seen from table 10 that the weights of apartments increase significantly when cash buyers are included in the RPPI. This implies cash-buyers favour proportionately more apartments (which are also very suitable for renting) than more expensive houses.

[41] As an experiment, the CSO attempted to link mortgage data to the Stamp Duty returns based on county and purchase price. Where links could be established a dummy variable for mortgages was included in the New RPPI price models. This estimated price coefficient for this dummy variable indicated that mortgage buyers paid up to 10% more for equivalent houses outside of Dublin in the 2012-2013 trough period. The premium for mortgage buyers within Dublin within this period was much less, at about 2%. However, matching the mortgage data to the Stamp Duty returns was a very inexact process and further work along these lines is needed before any final conclusions can be drawn.

The New RPPI was conceived in a period of considerable economic crisis, when Ireland had just experienced one of the worst collapses of residential property prices among developed countries in recent years. The new index was developed amid concerns that, during the immediate aftermath of the property crash, the CSO's Original RPPI was becoming increasingly less fit for purpose. The fact that the Original RPPI was based on mortgages, which were declining in market share, was beginning to erode confidence in the index amongst data users. There were also other concerns with the Original RPPI, such as how well it captured locational price effects and the inclusion of self-builds. A key feature of the New RPPI was that it was to be based on Stamp Duty returns, which covered the totality of the residential property market and excluded self-builds.

14.2 Launch of the New RPPI

Work began in earnest on the New RPPI in early 2013. The key challenge was linking the Stamp Duty returns at the transaction level to other administrative datasets, to capture the physical characteristics and precise geographical location of the properties (essential information for compiling a quality adjusted house price index). This challenge was met by employing innovative fuzzy address string matching techniques. It took three years to complete this matching process. But, by the summer of 2016 the backlog of Stamp Duty returns had been cleared and a New RPPI series had been created, commencing in January 2010.

The New RPPI had initially been conceived as a means of benchmarking the Original RPPI, to confirm whether the Original RPPI was in fact fit for purpose. However, it was quickly evident that the New RPPI was superior to the Original RPPI. Not only did it satisfactorily address coverage issues, exclude self-builds, better capture locational price effects and employ higher quality data, but the greater volume of transactions used, combined with more advanced data smoothing techniques, meant that the New RPPI could be disaggregated to a much greater extent. In all, twenty separate price indices could be published for the New RPPI (whereas only eight separate indices could be published for the Original RPPI). Therefore, the CSO decided to adopt the New RPPI as Ireland's official measure of house price inflation.

The New RPPI was formally launched in September 2016. Its publication was accompanied by an extensive range of additional real estate indicators (volume, value and average price statistics)[42]. Both the new indices and the new additional indicators were very well received by users. The launch of the New RPPI was widely regarded as a watershed moment in terms of improving both the quality and quantity of official statistics on the Irish residential property market. The New RPPI greatly strengthens the statistical oversight of Irish residential property market and leaves the EU, the state, the financial lending institutions and all other interested parties better informed of developments in the sector. Hopefully, the New RPPI will play a role in mitigating the worst excesses of residential property boom and bust cycles going forward.

14.3 Further developments

However, work continues on developing the New RPPI. In particular, work is proceeding on producing separate price indices for new and existing dwellings, to fully comply with *Regulation*

[42] It was not possible to publish comparable indicators for the Original RPPI, given that the mortgage transaction data only covered a fraction of the residential property market.

93/2013. Work on this project, is far advanced, aided by the greater volume of new dwelling transactions for modelling and the new data smoothing technique[43]. An innovative web-scraping project has begun to source at least some of the dwelling sales information online from real estate agent websites. The CSO is working towards securing land plot size information, which would add further refinement to the price model and which may facilitate other conceptual approaches[44]. The range of additional indicators published has been expanded and will continue to expand throughout the course of 2017. Thus, the New RPPI remains very much a live project and one which will continue to grow and evolve in the years to come to best serve the needs of its many users.

[43] The CSO delivered the first provisional new and existing dwelling prices indices to Eurostat in December 2016.

[44] One alternative conceptual approach is the Builder's Model approach (Diewert, de Haan and Hendriks, 2015). This approach was trialled with some success for the New RPPI, although it did not prove as effective as the log-linear approach, possibly as only proxy information was available for plot size. If reliable plot size information was available, this approach would be worth trialling again.

Appendix

Sample regression results

Dublin City Houses – 12 Month Regression Dec 2016

The REG Procedure
Dependent Variable: log of price

Number of Observations Read	2672
Number of Observations Used	2672

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	561.34199	22.45368	570.75	<.0001
Error	2646	104.09496	0.03934		
Corrected Total	2671	665.43695			

Root MSE	0.19834	R-Square	0.8436
Dependent Mean	12.75567	Adj R-Sq	0.8421
Coeff Var	1.55495		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	12.60627	0.02348	536.96	<.0001	0
tfa	Total floor area	1	0.00593	0.00011466	51.70	<.0001	1.41994
end_terrace	End-terraced house	1	-0.11007	0.01194	-9.22	<.0001	1.38460
mid_terrace	Mid-terraced house	1	-0.13862	0.00952	-14.55	<.0001	1.53800
HP2011ABS	HP Deprivation Index	1	0.01482	0.00047828	30.98	<.0001	1.99186
D01	D01: Dublin 1	1	-0.58615	0.07737	-7.58	<.0001	1.06240
D02	D02: Dublin 2	1	-0.13339	0.05709	-2.34	0.0195	1.07177
D03	D03: Dublin 3	1	-0.27753	0.01871	-14.83	<.0001	2.13210
D05	D05: Dublin 5	1	-0.34751	0.02014	-17.26	<.0001	2.46935
D06	D06: Dublin 6	1	-0.06646	0.02014	-3.30	0.0010	1.76692
D6W	D6W: Dublin 6W	1	-0.15348	0.02756	-5.57	<.0001	1.31655
D07	D07: Dublin 7	1	-0.34702	0.01854	-18.72	<.0001	2.65625
D08	D08: Dublin 8	1	-0.37831	0.02060	-18.36	<.0001	1.95980
D09	D09: Dublin 9	1	-0.34040	0.01848	-18.42	<.0001	2.47775
Other_Area	Other Routing Key	1	-0.49544	0.01829	-27.08	<.0001	4.72574
mon02	Time Dummy 2	1	-0.00981	0.01850	-0.53	0.5962	1.33222
mon03	Time Dummy 3	1	0.01478	0.01825	0.81	0.4179	1.34007
mon04	Time Dummy 4	1	0.02028	0.01826	1.11	0.2668	1.34192
mon05	Time Dummy 5	1	0.01258	0.01741	0.72	0.4700	1.39211
mon06	Time Dummy 6	1	0.01625	0.01771	0.92	0.3589	1.37352
mon07	Time Dummy 7	1	0.06280	0.01671	3.76	0.0002	1.43207
mon08	Time Dummy 8	1	0.07052	0.01646	4.29	<.0001	1.44725
mon09	Time Dummy 9	1	0.08275	0.01563	5.29	<.0001	1.52237
mon10	Time Dummy 10	1	0.06293	0.01668	3.77	0.0002	1.43315
mon11	Time Dummy 11	1	0.08690	0.01667	5.21	<.0001	1.43842
mon12	Time Dummy 12	1	0.03893	0.01655	2.35	0.0188	1.44672

Note: Reference categories are as follows;

- Dwelling Type: Detached/Semi-Detached House
- Postal District: D04 (Dublin 4)
- Time Dummy: Time Dummy 1 (Jan 2016)

This regression has been run after outliers have been excluded.