Constructing a National House Price Index for Ireland

Niall O'Hanlon[1]

Central Statistics Office and Centre for Policy Studies UCC

(*read before the Society, 19 May 2011*)

Abstract: The Central Statistics Office will publish the first results of its national House Price Index on May 13th 2011. This hedonic index is constructed using data on mortgage drawdowns by eight lending institutions under Section 13 of the Housing Act (2002). This paper describes how the development of the index has been largely driven by an impending European legislative requirement to produce indices of the costs of Owner Occupied Housing in the context of the Harmonised Index of Consumer Prices. It discusses the limitations of data on Irish residential property transactions in the context of the national House Price Index and the proposed register of property transactions. Practical considerations covering the treatment of data, the design of the hedonic functions, the rolling year regression model employed and the weighting of sub indices to form a composite national index are described. The index results are examined in the context of some of the other measures of house prices in Ireland. Finally, the paper explores some future challenges for further development of the measurement of residential property prices in Ireland.

1. INTRODUCTION

On May 13th, the CSO published the first set of results for its Residential Property Price Index (RPPI) ahead of an impending EU legislative requirement to provide RPPI data to Eurostat from 2012. The hedonically mix-adjusted index is compiled using mortgage drawdowns data provided by the main mortgage lenders. It is the only transactions based measure of residential property prices in Ireland.

Section 2 of this paper describes the European context in which the index has been created. Section 3 describes the data source and explains why other data on property transactions cannot currently be used. Section 4 details the quality of the mortgage drawdown data and the various data preparation processes. Section 5 covers the quality adjustment procedure. Section 6 introduces the initial results of RPPI and section 7 outlines some future challenges.

2. EUROPEAN REQUIREMENTS FOR HOUSE PRICE STATISTICS

2.1 Owner Occupied Housing in the HICP

While there is an obvious need at a national level for an official measure of residential property price indices, the design of the CSO RPPI has been largely determined by impending legislative requirements at EU level. It is necessary therefore to examine these requirements in order to set the context in which the index has been developed.

Indices of the acquisition and ownership costs of owner occupied housing (OOH) represent the most significant gaps in coverage of the Harmonised Index of Consumer Prices (HICP). This exclusion of OOH costs is the largest

---

[1] I thank Steve MacFeely, CSO and Bill Sjostrom, UCC for very helpful comments. All views expressed, as well as any errors or omissions are those of the author alone.

difference between national indices and the EU wide measure of consumer goods price inflation. Consequently, the HICP may not accurately measure the full spectrum of inflationary pressures in the EU. In order to address this issue the EU initiated a pilot project to measure the costs of OOH in 2001 with 5 countries participating in the first stage. The project is currently in its 4th stage which will run until the end of 2011 with 24 countries participating. Ireland joined the project in 2008 during Stage 3.

The current financial crisis, and the role that house prices have played in it, has given an added importance to the harmonised house price indices in their own right and they are now one of the indicators required for the scoreboard for the surveillance of macro-economic imbalances. The ECB has requested volume and value data on the residential property sector in addition to price indices so it can assess the dynamics of housing market activities. These data will comprise of both the number of residential dwellings sold as well as the transaction values. These transactions data are available for several EU countries. However, they are generated from various sources and may not be comparable. The ECB requires harmonised statistics covering residential property sales at least at a quarterly frequency and within three months after the reporting period. Discussions of how this concept should be implemented will continue in 2011.

A draft Regulation on OOH (which includes the development of a harmonized House Price Index) is currently being discussed at EU level with an expectation that it will be enacted in time for the transmission of national house price indices from the second quarter of 2012. The Regulation will establish OOH price indices with a view to improving the relevance and comparability of the Harmonized Indices of Consumer Prices ('HICPs'). The key conceptual elements of the draft legislation to note are:

- **'owner-occupied housing price index'** means an index that measures 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.

- **'house price index'** means an index that measures the changes in the **transaction prices** of dwellings that households purchase.

- the OOH index follows the **'Net acquisitions'** approach which measures changes in **actual prices** paid by consumers for the acquisition of dwellings that are **new to the household sector**, and changes in other costs related to the ownership, and transfer of ownership, of dwellings. Transactions of dwellings within the sector cancel out, with the exception of transfer costs.

It is expected that the legislation will require coverage of the following categories in respect of OOH indices;

O.1	**Owner-occupiers housing expenditures**
O.1.1	**Acquisitions of dwellings**
O.1.1.1	New dwellings
O.1.1.1.1	Purchases of new dwellings
O.1.1.1.2	Self-build dwellings and major renovations
O.1.1.2	Existing dwellings new to households
O.1.1.3	Other services related to the acquisitions of dwellings
O.1.2	**Ownership of dwellings**
O.1.2.1	Major repairs and maintenance
O.1.2.2	Insurance connected with the dwelling
O.1.2.3	Other services related to ownership of dwellings

The following expenditure categories will be covered in the **house price index**:

H.1.	Purchases of dwellings
H.1.1.	Purchases of new dwellings
H.1.2.	Purchases of existing dwellings

The periodicity of the indices will be at least quarterly. The indices covering the acquisition of residential properties will be transmitted from the 2nd quarter of 2012. Member States will be required to transmit those indices measuring inflation faced by owner occupiers from the 2nd quarter of 2014. This recognises that fact that most of the effort by Member States has been in respect of developing acquisition indices (or House Price Indices). Much work is still required on these additional indices both at the national and the EU level.

At this time it is not certain how the indices will be incorporated into the HICP framework. However, in the draft regulation the European Commission (Eurostat) commits to

> "*within* **four years of the date of entry into force of this Regulation**, *prepare a report on the indices established pursuant to this Regulation and in particular on their degree of compliance with HICP standards and present it to the to the European Statistical System Committee. The report shall also address the suitability of the owner-occupied housing indices for their potential integration into the HICP coverage. Taking into consideration the results of the review, the European Statistical System Committee shall give its view, on whether the integration of the owner-occupied housing into the HICP is warranted.*"

It is likely therefore that the earliest possible entry date into the HICP will be 2017 or 2018.

Eurostat has also set out a specific set of criteria by which OOH indices should be evaluated in addition to its more general definition of quality of statistics; (Eurostat, 2003). These quality criteria in particular have informed the design of the CSO RPPI and are set out under a classification system (see next page) under which estimation methods and processes graded as:

A methods: most appropriate methods;
B methods: methods which can be used in case an A method cannot be applied (i.e. "acceptable" solutions);
C methods: methods which shall not be used.

It is important to note that both the draft legislation and the evaluation criteria explicitly state that the OOH (and by extension the national RPPI indices) should be transactions based. Prices can only be observed where a sale of property takes place as individual properties are infrequently sold and highly heterogeneous in nature. Property is therefore different from most other consumer products, where transactions are sufficiently frequent that we can measure their price without observing the actual transaction. Appraisal and asking price observations are deemed C methods – i.e. they cannot be used as they cannot be considered a robust verifiable measure of price by any standard. Furthermore, indices should be quality-adjusted i.e. they should take account of the change from period to period in the quality/characteristics of properties transacted (location, size, building type etc). Only after this is done can a pure price index be calculated.

Table 1. Tentative criteria for Eurostat evaluation of OOH Indices

Classification Scheme		
Criteria	**A Most appropriate method/solution**	**B Acceptable/interim method/solution**	**C Non-acceptable method/solution**
1. Definition and aim of the index	• OOH index targeting the application of the net acquisitions approach to OOH	• Use of the HPI to measure OOH	• Exclusion of OOH
2. Geographical coverage	• Covering all country	• More than half	• Less than half
3.Weight concept	• Value of transactions	• Size of transacted dwellings • Value of dwellings stock	• Number of transactions • Population
4. Used prices relevant price concept source timing	• Full transaction price • Official register or accredited primary data source (mortgage lenders, Notaries, Real estate agencies, etc) • When first binding contract is signed	• Registered price w/o unregistered part • Secondary source (e.g. an association or a company with an agreement to provide data) • When ownership rights transferred or mortgage approved or preliminary contract signed	• Expert estimate (e.g. bank appraisals) • Secondary source (with no or little influence to the content of data) • Asking price (e.g. advertised price)
5. Treatment of land prices	• Excluded	• Partially excluded (e.g. only in the self-builders case detached houses case)	• Included
6. Adjustment for quality change	• Hedonics (possibly with combined stratification)	• fine stratification • stratification variables found on the basis of statistical analysis	• none ("unit value approach") • coarse stratification
7. Periodicity	• Monthly	• quarterly	• longer interval
8. Base period for the weights for the index for the prices	• three years • the same as the HICP • December of each year	• one year	• one month • defined randomly
9. Publication of results production delay revisions policy metadata	• produced in parallel with HICP results • no revisions • well documented method supported by the empirical studies; metadata available to external and internal users	• two months of delay (m+2) • irregular • documented method; seldom considered and lack of theoretical literature	• more than two months of delay • frequent and irregular • Poorly documented method
10. Reliability	• No bias, high precision • Small bias and good precision	• Bias and precision quantifiable; reasonable	• Bias and precision quantifiable; poor • Not known

3. DATA SOURCES

3.1 Potential Data Sources deemed unsuitable for index compilation

All transfers of residential properties on Ireland require that a Stamp Duty return is made to the Revenue Commissioners and that the change of title or deed is registered with the Property Registration Authority (PRA). Therefore stamp duty and registrations based data cover the entire market. These returns should, at least in theory, include information on the transaction price and exact address of each property transacted. However, they do not collect sufficient detail on the characteristics of individual properties to allow for an appropriate level of mix-adjustment.

The CSO has examined all stamp duty returns to the Revenue Commissioners made between 2000 and 2009. While these data provide useful insights into the size of the residential property market, they do have a number of weaknesses. This preliminary analysis conducted by the CSO highlighted a large portion of returns where the transaction price has not been recorded. It is likely that these returns relate to property transactions where stamp duty is not payable. There also appears to be some data quality issues in respect of duplicate records and records with extreme values. Furthermore, the records are not adjusted to take account of over or under payments in stamp duty that might result from incorrect transaction prices.

From 2010, all stamp duty returns are made online via the Revenue Commissioners website. While this will improve the quality of stamp duty returns it does not require any additional data on the characteristics of the property to be submitted. In time, these data are likely to become central to estimating the size of the market but will remain unusable in respect of price index calculation. Data from this new method of return are not yet available to the CSO.

The "Renewed Programme for Government (2009)" contained a commitment to establish a register of property transaction prices. This register will be a very welcome addition to the available data on residential property transactions but again it is important to note that it will not, at least under the current plans, include detail on the characteristics of those properties transacted as it will be based on Stamp Duty returns. Therefore this register will be of more benefit to users with detailed knowledge of the characteristics of specific properties (such as neighbours, local estate agents) than to others. It will not, for example, be possible to compile a mix-adjusted price index directly from the register. Should the register provide geo-coded data there may be some potential for linking the transaction price with characteristics captured elsewhere (for example by estate agents).

3.2 Data Source used in compilation of the Residential Property Price Index

Irish mortgage lenders are required under Section 13 of the Housing (Miscellaneous Provisions) Act 2002, to submit monthly mortgage returns to the Department of Environment, Heritage and Local Government (DoEHLG), containing data on both mortgage approvals (occurring where a formal letter of mortgage offer has issued) and mortgage drawdowns (where the loan has been drawn down – i.e. the transaction completed). This data requirement was set up primarily to generate a national mix-adjusted house price index but also to produce other relevant statistics that would inform housing policy generally.

The DoEHLG developed a House Price Statistical System (HPSS) to process these data and produce statistical analysis, including a national house price index. It was envisaged that mortgage approvals data would be used as monthly price observations as these represent the earliest formal recording of agreed price. The original proposal for a mix adjusted design followed a fine stratification approach where each month mortgage approvals were stratified into 288 cells.[2] Mortgage drawdowns would be used to internally weight each of the 288 cells.

The DoEHLG was unhappy with the results of its initial attempts at developing indices however. In mid 2007, the DoEHLG asked the CSO to become formally involved in the design and build of a national house price index. In response to this request and mindful of the impending requirement for OOH indices in the context of the HICP, the

---

[2] 8 geographic regions * 3 house size categories (based on number of bedrooms) * 3 house types * 2 buyers status types (first time buyer or otherwise) * 2 age of property (new or previously lived in)

CSO began developing a house price index in 2008. Eight of the mortgage lending institutions making monthly returns to the DoEHLG are also transmitting these data directly to the CSO since March 2010.[3]

The data consist of an individual record for every loan approval and loan drawdown made by the lender in the reference month. The data are anonymised – neither the individual borrower nor the property purchased is identifiable. Each record contains 67 variables of which; 2 relate to the financial institution, 32 to the borrower(s), 18 to the loan details and status and 15 to details of the property to be mortgaged. These variables are set out at Appendix I. Those variables relating to the property purchased, which could be used directly in the compilation of a House Price Index are:

- Transaction type – private purchase or with government subsidy
- Agreed purchase price of the property
- County of location (26 administrative regions)
- City indicator (for 4 cities excluding Dublin)
- Postcode, where relevant (for Dublin only)
- Newly built property indicator
- Year of Build
- Dwelling Type (detached, semi detached, terraced, flat or bungalow)
- Construction type (brick/block, timber frame or pre-cast concrete)
- Floor area
- Plot size (land)
- Number of rooms
- Number of bedrooms
- Use of Property
- Price at drawdown (only 0.04% of records have a different value for this where populated)

It should be noted the national location descriptors are limited to administrative county, city indicator and postcodes where relevant. Ireland does not yet have a national postcode system or harmonised spatial indicator. Postcodes are limited primarily to Dublin City which is divided into 22 postal districts. This limitation on address detail and the absence of a national system of location specific postcodes in Ireland complicates the compilation of a mix adjusted index and places limitations on the potential explanatory power of a hedonic model and the level of "granularity" in published results.

4. DATA QUALITY AND TREATMENT

4.1   Data Coverage

Along with the limitation on address detail, the partial coverage of the residential property market afforded by the mortgage data is a weakness. The CSO has conducted a preliminary assessment of the extent of non-coverage using the data on Stamp Duty returns supplied by the Revenue Commissioners. Given the large portion of returns where the transaction value has not been recorded it is confined to a volume comparison only. The data suggest that in 4 of the 5 years between 2005 and 2009 the mortgaged cover in excess of 75% of market transactions, as measured by Stamp Duty returns (see Table 2 overleaf).

---

[3] AIB, BOI, EBS, Haven, ICS, KBC, NIB and PTSB

Table 2. Coverage of Stamp Duty returns versus mortgage drawdown returns

Year	Stamp Duty Returns where only a residential property is transacted	Mortgage drawdown returns	% coverage
2005	80,685	71,138	88
2006	86,086	66,749	78
2007	66,090	51,852	78
2008	42,682	29,763	70
2009	19,719	18,596	94

4.2 Data Quality

The quality of the monthly mortgage data has been examined, prior to any regression analysis, in respect of duplicate records and values that are missing, outside permitted ranges (in the case of classification variables) and spurious values. This data checking does not cover the entire 67 variables but the subset of variables, as listed above, which could be used in the calculation of a House Price index.

4.3 Duplicate Records

Duplicate records fall into two categories, complete and partial. Complete duplicates refer to where exact matches of records occur. Partial duplicates refer to records which are not exact matches but appearing to refer to the same mortgage applicant and property. This usually occurs where a mortgage drawdown has been split into two parts, one attracting a fixed rate of interest and the other a variable rate. In these cases the full purchase price is reported in both records. Partial duplicates are assumed where several variables (county of location, purchase price, floor area, number of bedrooms, new or second indicator, type of house, year of construction as well as age and gross annual pay of the main applicant) are identical across different records. Each month data is checked for both complete and partial duplicates and any occurrences are excluded from the data prior to index calculation. These checks are confined to the month in question only – there are no checks spanning different periods.

No complete duplicates have been identified. The number of partial duplicates is small, accounting for less than 1% of records on an annualised basis, except in 2005 when the figure was just under 2%.

Table 3. Occurrence of Duplicates

	Records	Duplicates	% Occurrence
2005	71,138	1,379	1.9
2006	66,749	489	0.7
2007	51,852	366	0.7
2008	29,763	207	0.7
2009	18,593	67	0.4
2010	14,661	111	0.8

The addition of the relevant property folio number to each record in the monthly data would allow for a more precise identification of duplicates. The addition of the folio number could yield other benefits, which will be discussed later.

4.4 Data Edits

Two sets of data edits are reported. Firstly, a set of edits are run on all variables that may be used in the construction of a house price index. These edits primarily identify missing values and values that are outside predefined ranges in the case of categorical variables. Not all of these checks might be considered critical. However, the analysis does

provide a very useful insight into the quality of data. Results are presented in Table 4 and they show that the occurrence of basic errors is very high. In particular the quality of "size of plot" and "number of rooms" variables is especially poor. In 2009 some 53% of records failed at least one of the basic edit checks. While this represents a substantial improvement on previous years (where errors were found in excess of 78% in 2005 and 2006) it does point to the persistence of serious data quality issues.

Secondly, a set of edits are run on a number of key variables. Records that fail to pass these edits are deemed unusable for the purposes of index calculation and/or OOH calculation. Erroneous or missing values for these variables cannot be cleaned by imputation or by excluding specific variables from analysis as they are used to classify observations into the appropriate hedonic models. Results for these edits are presented at Table 5. They are unlike some of the edits listed under Table 4 below which can be reasonably dealt with through imputation. The key variables were selected on the basis of an assessment of the overall quality of each variable and their impact on various hedonic models. A balance must be achieved between including relevant characteristics in the analysis and maximising the number of records that can be used. This is especially important in current market conditions where there are very low levels of transactions. Tables 4 and 5 show that the basic quality of data has improved in recent years and this improvement corresponds with CSO engagement with data providers. It may also be influenced by the greatly reduced level of transactions and a more stringent analysis of mortgage application and associated valuations by lenders. Nevertheless, further improvement is required.

Table 5 shows that with the exception of measures of property size the levels of occurrence of these critical edits are relatively low. Furthermore, in 2009 and 2010 the occurrence of property size critical errors have decreased by around 2/3's of previous levels. Consequently, approximately 12% of records would be unfit for analysis using these criteria despite the fact that over 50 % of records in each of these years have at least one error as specified by the "long list".

Table 4. Percentage of All Error Records

	2005	2006	2007	2008	2009	2010
Error records	78.6	77.9	75.6	68.3	61.2	52.9
Wrong month	0.2	0.0	0.0	0.0	0.0	0.0
Price missing	0.6	0.7	1.0	0.6	0.1	0.0
Transaction type missing	0.0	2.1	2.2	1.4	0.0	0.0
County missing	0.5	2.1	2.2	3.0	0.4	0.4
Exact location missing	4.4	3.3	3.0	3.7	2.2	2.7
Dublin postcode error	1.3	1.5	1.1	1.0	1.1	1.0
House status missing	0.9	1.9	1.6	0.7	0.1	0.0
Year of construction missing (Old)	27.8	29.2	20.9	15.6	5.1	4.4
Year of construction missing (New)	6.2	6.7	6.0	6.6	3.0	1.9
Not to be built within 2 years	0.1	0.0	0.0	0.0	0.0	0.0
Second hand but new	0.2	0.1	0.1	0.1	0.1	0.0
House type missing	3.5	5.7	4.0	4.1	2.3	2.1
Construction type missing	24.7	24.5	23.0	23.3	8.5	10.0
Floor area missing	21.4	23.4	22.4	24.7	8.9	7.0
Number of bedrooms missing	3.4	5.4	4.1	5.0	3.5	3.8
Number of rooms missing	47.5	57.8	39.8	26.9	10.3	8.6
More bedrooms than rooms	6.4	8.7	6.5	5.8	6.4	5.9
Plot size missing for detached/bungalows	18.5	13.8	21.1	24.6	25.1	19.6
Plot size missing for all house types	41.7	30.2	41.3	46.2	43.2	35.9
Property usage missing	0.3	0.4	0.3	0.3	0.1	0.0
Buyer status missing	0.1	0.0	0.5	1.2	1.3	0.3

Table 5. Percentage of Critical Error Records

	2005	2006	2007	2008	2009	2010
Error records	25.0	25.9	25.7	27.4	12.3	11.4
Wrong month	0.2	0.0	0.0	0.0	0.0	0.0
Price missing	0.5	0.7	1.0	0.6	0.1	0.0
Transaction type missing	0.0	2.1	2.2	1.4	0.0	0.0
County missing	0.5	2.1	2.2	3.0	0.4	0.4
Exact location missing	4.4	3.3	3.0	3.7	2.2	2.7
Dublin postcode error	1.3	1.5	1.1	1.0	1.1	1.0
House status missing	0.9	1.9	1.6	0.7	0.1	0.0
House type missing	3.5	5.7	4.0	4.1	2.3	2.1
Size errors	20.8	23.8	20.4	22.7	7.6	7.2
Property usage missing	0.3	0.4	0.3	0.3	0.1	0.0
Buyer status missing	0.1	0.0	0.5	1.2	1.3	0.3

4.5 Thresholds

The issue of implausible or extreme values must also be considered. These values arise where specific variables have values that appear highly unlikely, illogical or where they contradict values given for other variables. For example, we can be almost certain that a reported floor area of 700 sq metres for of a two bed roomed apartment is incorrect (what's more we can be confident that the floor area should in fact have been reported as 65 sq metres i.e. 700 sq feet!). Implausible values may arise for a number of reasons such as a recording error made during the mortgage application process, an error during data entry or possibly even as a result of errors in the lenders reporting software. The example of the 700 square metre apartment is a real one – and is not surprising giving the continued use of both imperial and metric measures. Where values exceed thresholds and are therefore considered spurious new values can be imputed for them. The use of thresholds and subsequently imputation is designed to ensure that for each observation the physical characteristics of a property are logical and consistent. Procedures to impute for missing or implausible values will be outlined under section 4.6.

Checks for implausible values are primarily conducted through use of thresholds. They are applied to the measures of property size; bed rooms, floor area, and plot size. Although number of rooms is used in both the establishment and application of thresholds ultimately it is not used in index calculation due to the very high number of missing and implausible values and the fact that size is adequately described by floor area and number of bedrooms.

While the application of thresholds might be considered subjective, they are formulated based on a detailed assessment of data covering the period from January 2005- September 2009. In the case of thresholds for number of bedrooms by property type, if a value for number of bedrooms of a given property type represent less than 1% of records for that property type have then it is considered outside the threshold. This 1% rule is used in setting the thresholds for the number of bedrooms and floor area. It results in the identification of only a small number of values for each of the variables as implausible.

Thresholds for plot size are targeted towards setting a minimum limit of plot size for houses and identifying records where the floor area is given as the plot size for houses. Two rules apply: firstly, that plot size should not be less than a third of the total floor area of the house (allowing for multi-storey); and secondly, that plot size should not exactly equal the floor area of the house.

Age of property must also fall within a range of allowable values. For new build houses the year of construction must be between three years prior to and two years after the year of sale. For second hand houses year of construction must fall between the year of sale and 1700.

The incidence of exceeded threshold values is detailed in the following tables. These show that between 2005 and 2009 over 20% of records contained at least one variable which was outside its relevant threshold. The majority of these resulted from the implausible values for number of rooms.

Table 6. Percentage of records with thresholds exceeded

	Records	Bedrooms	Floor area	Age	Rooms	Plot area
2005	21.7	0.6	1.8	0.7	12.9	6.8
2006	23.7	0.4	1.0	0.9	15.9	6.8
2007	21.3	0.7	2.4	0.9	12.1	6.2
2008	21.8	1.1	3.6	1.1	12.7	4.0
2009	20.4	0.9	3.6	1.1	13.0	2.2
2010	18.4	0.9	3.6	1.2	12.2	0.9

4.6   Imputation for missing, erroneous or implausible values

A series of imputations are run on four size related variables; floor area, number of bedrooms, rooms and plot area. There imputations are applied where values for these variables are either missing or where original variables failed the edits or threshold tests outlined above. Imputed values are calculated from data for the previous calendar year. This facilitates a balance between taking account of changing buyer preferences and deriving imputed robust values. Imputed values for floor area are calculated each at the end of each year based on median values for each combination of house type by number of bedrooms from only those records which have values for each of these variables that are within the threshold limits. The appropriate median value is then applied, as an imputed value to records requiring imputation in the following year. Median values are used rather than mean values as floor area data is positively skewed, perhaps due to some records reporting in squared feet rather than squared metres.

The same median values for floor area are also in used in the imputation for number of bedrooms. For those records where the number of bedrooms is missing the floor area is matched to the median floor area for each number of bedrooms for the relevant house type, giving an imputed value for bedrooms.

Median values for the excess of size of plot over floor area are calculated annually for each combination of house type by geographic regions (at NUTSIII level but with a class for the cities of Cork, Galway, Limerick and Waterford) which have values for each of these variables that are within the threshold limits. These values are then used for the subsequent year. The appropriate median value for the excess of plot area over floor area is added to the relevant floor area, providing an imputed value for plot area for records requiring imputation.

The incidence of imputed values is detailed in the following table. These show that in each year between 50% and 75% of records contained at least one variable which was imputed for. The majority of these resulted from missing values or spurious values for plot size and number of rooms.

Table 7. Occurrence of imputed values %

	Records	Bedrooms	Floor area	Rooms	Plot area
2005	75.3	0.6	3.8	49.8	41.0
2006	74.8	0.5	3.4	65.3	27.4
2007	71.2	0.7	5.8	40.9	39.7
2008	61.5	1.3	7.0	21.3	41.4
2009	59.9	1.3	5.4	17.9	42.8
2010	50.2	1.2	4.1	15.3	35.1

4.7   Outlier detection

The outlier detection test is run as a preliminary hedonic regression. In fact it involves running several individual models mirroring exactly the combination of stratification and regression used for the compilation of index results. These models will de described fully under section 5.3. The models are run once to identify outliers and then again with the outliers excluded to produce results. Cooks Distance, which measures the distance between each observation and the means of the dependent and independent variables and therefore the influence they have on the result of the model, is used to identify outliers. The conventional cut-off of 4/number of observations is applied and the test "passes" only those observations that are normally distributed. As leverage increases the size of residuals tolerated decreases.

It is important to ensure that certain classes of observations are not routinely excluded simply because they consistently have large leverages. This could occur, for example, in a Dublin postcode area with high property values but relatively low levels of transactions. Combining stratification and model approaches may help overcome this problem. In fact such an examination of the leverage of studentized residuals could be used in the selection of an appropriate stratification.

The hedonic technique used in the calculation of the index is the rolling year hedonic regression model which will be explained in more detail under section 5.3. This model involves pooling data from the 12 most recent months. For the first year of data (2005) time-dummy models (12 months of data) are used. For each subsequent month rolling year hedonic regression models are used. This involves adding observations for the newest month and dropping observations for the oldest month thereby maintaining a rolling pool of 12 months of price observations. As the historic data (months 1-11) have already been stripped of outliers and this data is pooled with the observations for the current month the test for the current month is more severe than it would be if it was just performed on the current months price observations without any pooling of data. Once the regression model moves from the annual time dummy to the rolling year the number of outliers identified increases significantly as shown in Table 8. These rates of detection are not unreasonable given the number of characteristics used in the model (each of which could potentially be reported incorrectly), the lack of detailed address information and the general quality of data. However, the spike in the number of outliers in 2007 and 2008 merits further analysis.

Table 8. Percentage of records identified as outliers

	Outliers %
2005	4.4
2006	10.6
2007	15.8
2008	17.4
2009	12.1
2010	11.1

Table 9 below details the percentage of records used in the regression analysis i.e. after duplicates, critical errors and outliers are removed (records with values imputed for variables as previously described are not removed). Data covering 2008 has the lowest proportion of useable records reflecting elevated levels of errors and outliers occurring that year.

Table 9. Percentage of records used in regression analysis

	%
2005	69.5
2006	67.4
2007	62.1
2008	59.5
2009	76.8
2010	78.2

5.   QUALITY ADJUSTMENT

5.1 Why quality adjust?
The challenge in compiling any price index is to separate pure price change from changes in the quality of the products being bought over time. This is done by comparing the prices of exactly the same products, the method typically used in price indices, such as in the Consumer Price Index. However, in the case of residential properties, price is determined by many characteristics (location, size, build type etc), the reason being that no two properties are exactly identical. Furthermore only a small portion of the total housing stock will be sold in any given month. The combination of these factors means that the matching process that would typically be used to calculate a price index cannot be used in the case of houses. Therefore, a form of quality adjustment must be introduced which allows us to estimate pure price change from highly heterogeneous and rarely transacted goods. This form of quality adjustment is somewhat unique to property price indices. Typically in price indices for other goods quality adjustment is used to remove quality improvement as products evolve.

5.2 The rolling year hedonic regression model
The hedonic method is the prevalent statistical process for the measurement of house price change. In this method a number of house characteristics that influence prices are analysed so that we can estimate and exclude the part of the price change that can be attributed to them. This leaves us with an index of pure price change for a consistent set of characteristics - or more simply - a house price index. As noted earlier the hedonic method (possibly combined with stratification) is classified by Eurostat as the most appropriate method of mix-adjustment.

There are a number of different hedonic approaches or methods the more common being the time-dummy (or constrained hedonic), hedonic re-pricing, hedonic imputation and characteristics prices methods. These methods are well documented and described in the literature on house price indices (see for example Hill 2011 and Duffy 2009).

The CSO RPPI uses a variant of the time-dummy method known as the rolling year hedonic regression model. The time-dummy approach is relatively simple to apply. As the estimation of the time coefficient is included in the regression equation it can be estimated directly from it.

In the case of n observation periods n-1 time dummies have to be constructed such that time dummy t_i equals 1 if the observation belongs to period i and 0 otherwise. So the regression equation, using a log-linear model, has the following structure:

$$\ln\left(p_{it}\right) = x_{it}\beta + \delta_t D_t + u_{it} \qquad\qquad (1)$$

where
$p\ it$ is the price of dwelling *i*
x is the vector of explanatory variables (size, type of dwelling, location, …)
$D\ t$ is the time dummy (value=1 if in time period *t* otherwise 0)
u is the unexplained or error term

As there are n periods there will be *(n-1)* time coefficients and the anti-logs of these are consistent estimates of the quality adjusted price index. Therefore the index can then be calculated directly from the equation.

In this method the influence of the x_i's are assumed constant (do not change from period to period). However, the coefficients for the characteristics do not remain constant. The addition of new data for each subsequent period result in changes to these coefficients as the effectiveness of the model improves. This results in continuous revisions to results for previous periods and so is generally unsuited to use in official statistics. The use of the adjacent-period regression is not appropriate using these data as 2 months data does not provide sufficiently robust coefficients. Shimizu *et al.* (2010) introduced a method which employs multiple neighbouring periods such as 12 months rather than just 2.

Diewert (2011: 38) describes the practical application of the method very well:

> "First, one chooses a "suitable" number of periods (equal to or greater than two) where it is thought that the hedonic regression model will yield "reasonable" results; this will be the *window length* (say M periods) for the sequence of regression models which will be estimated. Secondly, an initial regression model is estimated and the appropriate indexes are calculated using data pertaining to the first M periods in the data set. Next, a second regression model is estimated where the data consist of the initial data less the data for period 1 but adding the data for period M+1. Appropriate price indexes are calculated for this new regression model but only the rate of increase of the index going from period M to M+1 is used to update the previous sequence of M index values. This procedure is continued with each successive regression dropping the data of the previous earliest period and adding the data for the next period, with one new update factor being added with each regression".

A window length of 12 months was chosen for the RPPI. The method has the advantage of keeping the coefficients relatively up-to-date (and so takes accounts for changes in price determinants as market conditions change) despite the use of pooled data.

5.3   Selecting the appropriate hedonic models
Stratification is often combined with hedonics to allow for the separate measurement of different segments of the market using their own hedonic models. This is particularly helpful where sub-indices are required (covering different geographic regions or property types). It is also appropriate where available data on different segments of the market may have differing sets of characteristics.

At the outset there were a number of obvious strata by which transactions could initially be grouped and analysed so as to provide a basis for analysis of different hedonic models.

Dublin is divided into 4 local authority areas and Dublin city is further divided into 23 postal districts. Data is coded to these local authority areas and postal districts (where applicable) giving a reasonable level of micro-location. Data for the remainder of the country is divided into 26 counties (North and South Tipperary are classified separately) and 4 cities (Cork, Galway, Limerick and Waterford). Micro location detail is not therefore available for transactions outside Dublin. As a result, Dublin and the rest of Ireland are stratified separately. Houses and apartments have different explanatory characteristics and so they are also stratified separately.

179

Therefore the starting point for the analysis of potential hedonic models was 4 individual strata namely; Dublin Houses, Dublin Apartments, Rest of Ireland Houses and Rest of Ireland Apartments.

Notwithstanding the absence of very detailed micro-location data (especially outside Dublin), the dataset contains 9 characteristics which could be used in a hedonic regression;

- location
- property type
- size (3 different measures; floor area, number of rooms, plot area)
- new or second hand (implications for taxation on the transfer)
- age of property
- construction type
- first time buyer or not (while this is not a characteristic of the property it does provide useful explanatory power in respect of price determinants)[4]

It should be remembered however that data quality is not consistent across the characteristics with some being reported more accurately than others. The quality of number of rooms and plot area are particularly poor.

In order to analyse the appropriateness of different hedonic regressions a standard approach to testing them was applied. Annual time dummy models were created for each strata and run for each year between 2005 and 2010. Results were analysed in respect of the size, direction and significance of coefficients, the stability of these coefficients over the 5 years and the overall fit of each model in terms of its R^2. The models were run in 2 stages, firstly to identify outliers, and secondly to produce results from data stripped of outliers.

It was decided to exclude "number of rooms" from the analysis due to the very high portion of records with an imputed value for it. Furthermore, size could be adequately measured by floor area and number of bedrooms (commonly used as an approximation of property size in Ireland).

Age is commonly used to adjust for depreciation in the value of a house over time. However some older house can attract higher prices due to vintage and/or location effects.[5] Various prototype models for Dublin showed that prices tended to rise as age of house (in decades) increased. The size of these increases was erratic between decades and across data for different years. Furthermore as age cannot be imputed for, a large residual "age not reported" category was present in each annual dataset from 2005-2008.[6] Excluding records where age was missing would have significantly reduced the number of observations available to the regression. Finally the inclusion of age did not significantly alter the resultant index for Dublin. For houses outside Dublin a depreciation effect was evident although the size of the effect was somewhat erratic across data for different years. Again the large residual category for "age not reported" was present in data for the period 2005-2008. As in the case of houses in Dublin the inclusion of age did not significantly alter the resultant index. Age was therefore excluded for both Dublin houses and rest of Ireland houses.

Age was not included in the models for "Dublin Apartments" or "National excluding Dublin Apartments" as most of the apartments transacted were built in recent years.

Plot-size was excluded for houses both inside and outside Dublin. In both cases the coefficient was always very close to 1, meaning that the impact on price of each additional square metre in plot-size was extremely small. In Dublin between 2006 and 2009 each additional square metre added less than 0.01% to the price of a house. In both 2005 and 2010 the effect was negative i.e. each additional square metre reduced very slightly the price of a house (by less than 0.001%). Outside Dublin the impact on price of each additional square metre was to reduce the price of the house by less than 0.001%). This was unexpected as a larger plot size should, ceteris paribus, add value to a house. There are two factors that might explain this unexpected direction of the plot-size coefficient. Firstly, as we

---

[4] See related discussion in Conniffe, D. and D. Duffy (1999)
[5] Li, W., M. Prud'homme and K. Yu (2006), page 8
[6] Following the approach set out by Laferrere (2003)

have already seen, the quality of the variable is not as high as it could be (some 30% to 40% of values are imputed) and the imputation for missing values may be introducing a bias. Secondly, in the case of houses outside Dublin in particular, if plot-size tends to be larger in more rural areas then the model may be confusing the price effect of larger plots with an unrecorded negative rural effect.

Construction type was also excluded from the models as it was not found to have a significant impact on the models. Over 95% of records where construction type was recorded were brick/block builds.

The following tables detail the characteristics used in the regression models:

Table 10. Characteristics for Dublin Apartments and National excluding Dublin Apartments

Characteristic	Description
Location Dublin	26 dummy variables taking the value of 1 if property located in a particular postal district or local authority area. Otherwise 0.
Location National excluding Dublin	29 dummy variables taking the value of 1 if property located in a particular county or city. Otherwise 0.
Number of bedrooms	3 dummy variables (1,2, 3-4 bedrooms) taking the value of one if the property corresponds to number of bedrooms. Otherwise 0.
Floor area	Actual floor area in metres squared
New or old property	Dummy variable taking the value of 1 if a new property. Otherwise 0.
First time buyer	Dummy variable taking the value of 1 if a first time buyer. Otherwise 0.

Table 11. Characteristics for Dublin Houses and National excluding Houses

Characteristic	Description
Location Dublin	26 dummy variables taking the value of 1 if property located in a particular postal district or local authority area. Otherwise 0.
Location National excluding Dublin	29 dummy variables taking the value of 1 if property located in a particular county or city. Otherwise 0.
Property Type	4 dummy variables taking the value of 1 if property is one of detached, bungalow, semi-detached or terrace. Otherwise 0.
Number of bedrooms	4 dummy variables (1-2, 3, 4, 5+ bedrooms) taking the value of one if the property corresponds to number of bedrooms. Otherwise 0.
Floor area	Actual floor area in metres squared
New or old property	Dummy variable taking the value of 1 if a new property. Otherwise O.
First time buyer	Dummy variable taking the value of 1 if a first time buyer. Otherwise O.

The log-linear (or semi-log) model was chosen because an examination of distribution of residuals from various prototype models showed a positively skewed distribution when price was used as the dependent variable. This was corrected by using log of price. Furthermore, results generated by the log-linear form can be interpreted very simply - as the percentage change in the price of the house when they are multiplied by 100.

Finer levels of geographic strata were tested the outside Dublin models. These included NUTSII and NUTSIII based strata, strata based on provinces, a separate stratum for cities (Cork, Limerick, Waterford and Galway) and groupings of high/medium and low value counties and cities. All were inferior to the broader "National excluding Dublin" classification in respect of the volatility of the resultant indices and the explanatory power and stability of the models. This broader classification limits the level of disaggregation possible in respect of published indices. However the desire from users for more detailed indices must be balanced against the requirement to produce robust results.

Some examples of the regressions for March 2011 (based on data pooled from April 2010- March 2011) are presented below. The full list of location variables is not included in the presentation. Only the current month and previous months are shown.

It can be seen that the R^2 for National Excluding Dublin – Houses is lowest at 0.56 reflecting the absence of more detailed address information.

Table 12. Regression for Dublin Apartments March 2011

Variables	Anti-log of Coefficients	Standard error	p-value
Floor Area	1.004	0.0005	<.0001
1 Bed	0.859	0.0182	0.0001
3 or 4 Bed	1.114	0.0147	0.0009
New	0.947	0.0221	<.0001
FTB	0.929	0.0235	0.0448
Dublin 3	1.288	0.0685	0.0002
Dublin 4	1.623	0.0540	<.0001
February	87.897	0.0312	0.0014
March	83.715	0.0315	0.0060

Obs=575

Adjusted R^2=0.6409

Table 13. Regression for Dublin Houses March 2011

Variables	Anti-log of Coefficients	Standard error	p-value
Detached	1.073	0.015	<.0001
Bungalow	0.968	0.021	0.1171
Terrace	0.907	0.009	<.0001
Floor Area	1.006	0.0002	<.0001
1 or 2 Bed	0.960	0.012	0.0011
4 Bed	1.082	0.012	<.0001
5+ Bed	1.043	0.027	0.1228
New	0.868	0.014	<.0001
FTB	0.913	0.009	<.0001
Dublin 3	1.169	0.027	<.0001
Dublin 4	1.553	0.043	<.0001
February	91.118	0.021	<.0001
March	92.796	0.024	0.0021

Obs=2635

Adjusted R^2=0.7963

Table 14. Regression for National Excluding Houses March 2011

Variables	Anti-log of Coefficients	Standard error	p-value
Detached	1.113	0.009	<.0001
Bungalow	0.989	0.009	0.1991
Terrace	0.928	0.011	<.0001
Floor Area	1.002	0.0001	<.0001
1 or 2 Bed	0.895	0.015	<.0001
4 Bed	1.128	0.008	<.0001
5+ Bed	1.205	0.013	<.0001
New	0.950	0.006	<.0001
FTB	0.887	0.006	<.0001
Cork City	1.155	0.016	<.0001
Kerry	0.845	0.015	<.0001
February	89.031	0.015	<.0001
March	87.554	0.016	<.0001

Obs=6833

Adjusted R^2=0.5618

5.4 Weighting of strata to construct national index

Weights are calculated at the beginning of each year based on the value of transactions (expenditure shares) during the previous year as given by the mortgage drawdown data. The use of expenditure weights ensures consistency with the HICP approach. Simply put, the weights for each of the 4 strata are their expenditure shares in the previous year.

The index is an annual chain-linked Laspeyres-type index. It is calculated by updating the previous month's weights by the estimated changes to the average prices.

$$I_{t,0} = \left(\frac{\sum (V_{c-1}) \frac{P_c}{P_{c-1}}}{\sum Vy - 1} \right) \times 100$$

where:

$Vy - 1$ and V_{c-1} are the expenditure values (weights) of an index item heading in the previous year (y-1) or previous (c-1) month;

P_c and P_{c-1} are the average price of the same index item heading in the current (c) and previous (c-1) month;

Σ represents summation over all index item headings

The current cost of the fixed quantity of each stratum is calculated by updating the previous month's cost by the estimated monthly change in its average price. This index is also known as a weighted average of *price relatives*, the weight being the expenditure on the stratum in the previous year.

Ideally, Stamp Duty data would be used to derive the internal weights (aggregating each of the strata) and the external weights (total weight for acquisition of residential properties, residential properties new to the sector etc.) However, data returned via the old ST21 form are not suitable for these purposes due to the data quality issues previously described. It is expected that returns from 2010 onwards collected via the new online reporting system will offer much better potential in respect of weighting.

The distribution of weights across the 4 strata has changed significantly since 2005. Table 15 below shows that the apartments now account for less than 7% of the total value of transactions which is less than half of their weight of just over 16% in 2007. Properties outside Dublin account for over 60% of the total weight each year peaking at just over 68% in 2007 before falling to 62.5% in 2010.

Table 15. Distribution of Weights (%) for Index strata

Year	Dublin Apartments	Dublin Houses	National excluding Dublin Apartments	National excluding Dublin Houses
2005	8.8	26.9	4.6	59.7
2006	9.8	26.1	4.8	59.4
2007	10.6	21.1	5.7	62.6
2008	10.3	22.3	4.6	62.8
2009	7.4	26.6	2.4	63.6
2010	5.3	32.2	1.6	61.0

6. RESULTS

6.1   RPPI Release

The Residential Property Price Index (RPPI) will be published monthly approximately 3-4 weeks after the end of each month, via a paper based release. It can also be viewed interactively via the CSO's database direct.

Results are published separately for 3 of the 4 strata:
- Dublin - Houses
- Dublin – Apartments
- National excluding Dublin - Houses

The index for National excluding Dublin - Apartments is not published due to the very low number of transactions recorded for the strata and the subsequent high level of short-term volatility in the index. The following aggregate indices are also published:

- National – All properties
- National – Houses
- National – Apartments
- Dublin – All properties
- National excluding Dublin – All properties

In order to smooth out short-term volatility in the published series and highlight longer-term trends the published indices are based on a 3 month rolling average, i.e. a simple average of the current month and the previous 2 months.

6.2 Results January 2005-March 2011
Results are the 3 published strata are presented below in Figure 3. The "National - all properties" index and the indices for "National excluding Dublin - all properties" and "Dublin – all properties" are contained in Figure 4. Indices are presented at Appendix 4.

Figure 2.

Detailed Indices Jan05=100

Figure 3.

Aggregate Indices Jan05=100
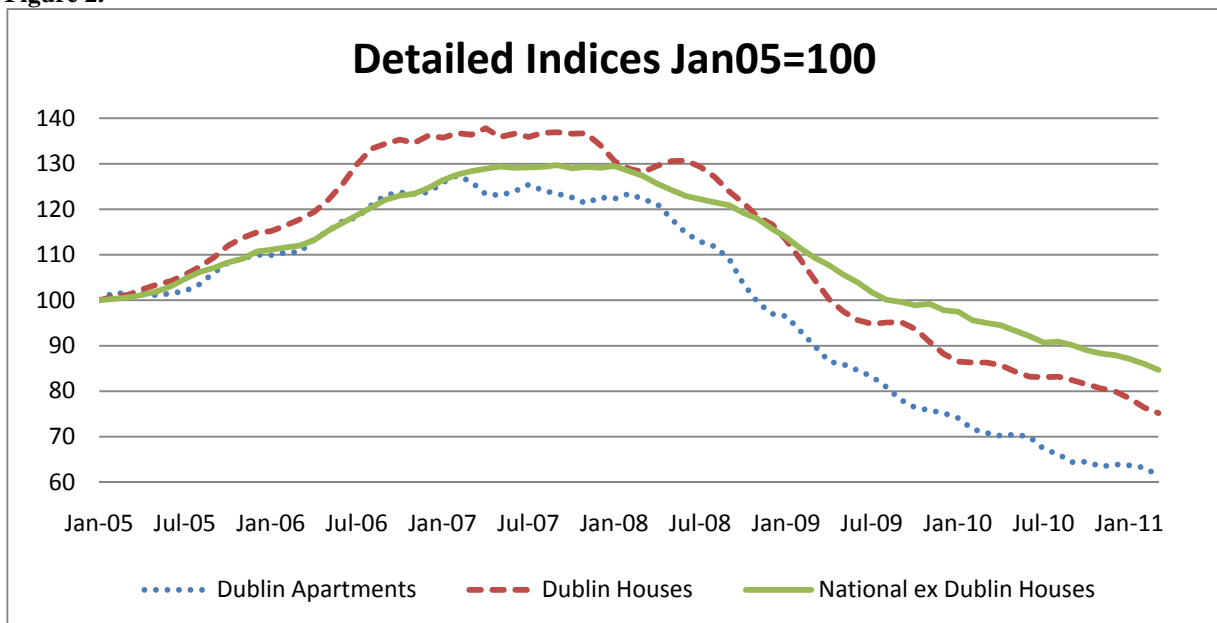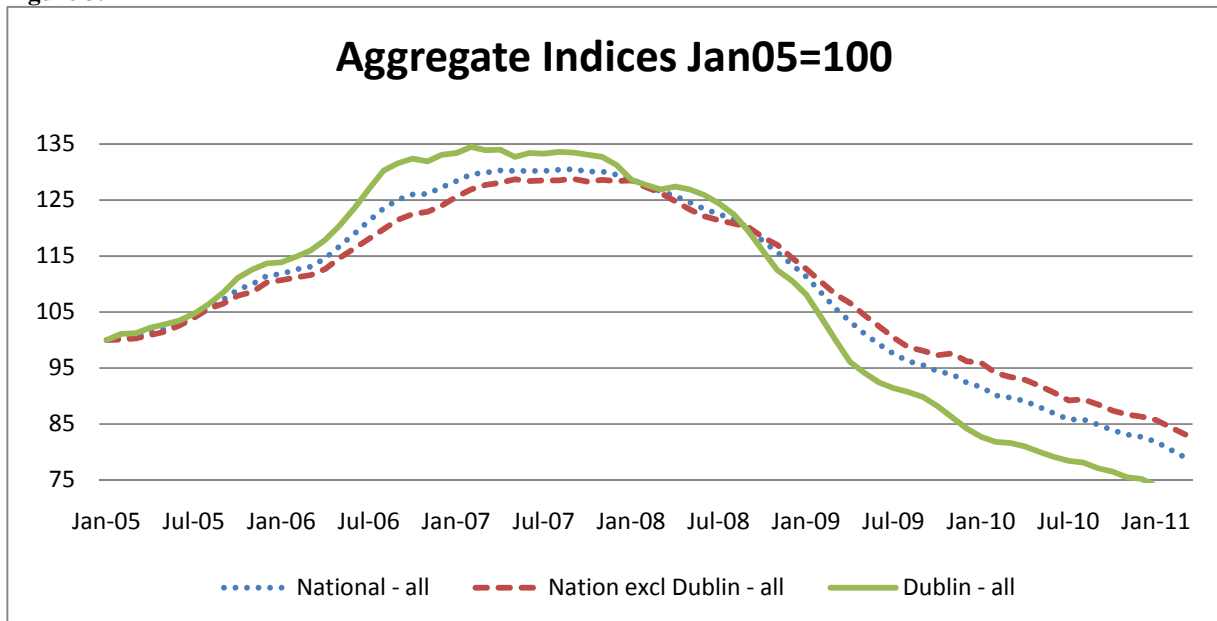
Until the discontinuation of the PRSB/ESRI In May 2011 there were 4 well-known indices of property price change in Ireland produed by; Daft.ie, Myhome.ie, PTSB/ESRI and Sherry-Fitzgerald.

Figure 4 shows the CSO RPPI to be somewhat in the middle of the others all following a similar rate of decline to the Daft.ie and MyHome.ie indices in recent quarters.

Figure 4.
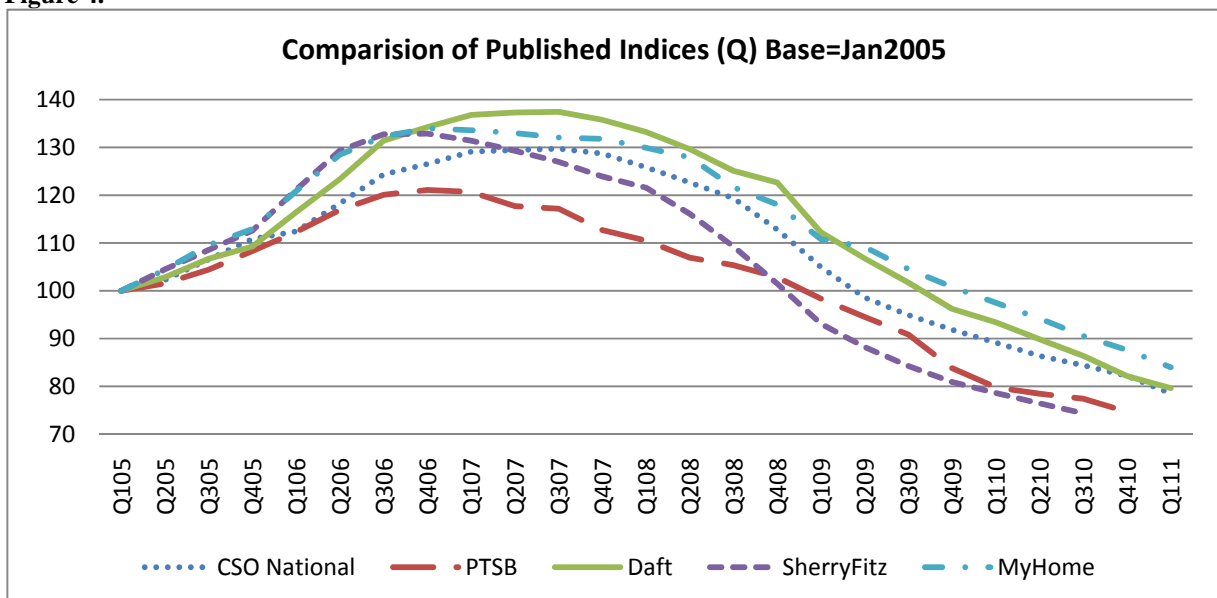

Comparision of Published Indices (Q) Base=Jan2005

Table 16 compares the various indices in respect of their price obervation type, when they achieved their highest level and the magnitude of decline from that level. Three of the five indices reached their highest levels in late 2006 but the Daft.ie and CSO indices lag several months behind in mid 2007. However it should be noted that with the exception of the Sherry-Fitzgerald index, all were quite flat in mid 2007.

Table 16. Comparison of selected published indices

Index	Price Observation	Highest level	Decline from highest level
PTSB/ESRI	Transactions	Qtr 4 2006	38%
Daft.ie	Asking Prices	Qtr2/ 3 2007	43%
Myhome.ie	Asking Prices	Qtr 4 2006	37%
Sherry Fitzgerald	Repeated Valuations	Qtr 4 2006	51%
CSO	Transactions	Q2 /Q3 2007	40%

There is undoubtely a time lag between when a transaction price is agreed and the associated mortgage is drawn down . It is not surprising therefore that the CSO index achieved its highest level a number of months later than some of the other indices. It is possible that this lag could be reduced, or even eliminated, if mortgage approvals data were used instead of mortgage drawdowns. As noted earlier the lending insitutions supply the CSO with data on mortgage approvals, as well as drawdowns. The CSO plans to compile an index based on mortgage approvals (perhaps as a complimentary once-off analysis) and to compare it to the RPPI. This should allow us to estimate the lag from approval to drawdown. However, it must be remembered that the impending EU legislation covering OOH indices requires that the index must be based on transactions. Approvals cannot be assumed to result in tranactions. Similarly, one transaction may have generated several approvals (from different lenders). Therefore, the RPPI will continue to be based on mortgage drawdowns.

7. CONCLUDING REMARKS: FUTURE WORK AND CHALLENGES

7.1 *Keeping models kept under constant review*

Improvements in the quality of the mortgage data offer the possibility for adding additional characteristics (age and plot-size) to the regression models. Their potential inclusion will be examined on an annual basis. This examination will go hand in hand with a general review of the performance of the models each year.

7.2 *Securing access to and utilising better address detail*

The lack of more detailed address information (particularly outside Dublin) is undoubtedly the most serious weakness within the dataset. It places a severe limitation on the explanatory power of the location coefficients and excludes the potential for "quality of neighbourhood" analysis. Furthermore, it is possible that additional characteristics like plot size and age may improve the model only when micro-location is also accounted for. It is reasonable therefore that the CSO should seek additional address detail. Alternatively, the provision of the property folio number may allow for matching to other data sources where full address detail is present.

However, the absence of a nationwide system of postcodes (ideally location specific) complicates enormously the potential provision and usage of better address detail. The introduction of nationwide postcode may well be a required before more detailed address information is provided and/or used.

7.3 *Measuring the size of the market*

Timely and robust data generated by the new online Stamp Duty returns system should deliver accurate detail on the size of the residential property market. This will benefit the external weighting of the acquisitions indices in particular and also provide regular volume and value measures of the market which are in themselves of significant importance.

CSO has had some initial discussions with the Revenue Commissioners regarding these data. However, it is not yet certain when they will be available. It should be noted that until then it will not be possible to accurately measure the total value of the market.

7.4 Compliance with EU requirements

As well as the requirement to produce indices on the ownership of houses which are outside the scope of this paper, the CSO must comply with requirements in respect of the acquisition of houses. The introduction of the RPPI release is a very significant milestone in this regard. However the level of disaggregation currently proposed by Eurostat (new houses, houses new to the sector, self-builds and major renovations) will pose significant challenges for the CSO (and other NSI's) especially in the context of the very small size of the current market. The current data on mortgage drawdowns simply cannot support finer levels of stratification and more detailed models. These challenges will require flexibility and pragmatism on behalf of Eurostat, Ireland and other member states. For example the introduction of minimum weight thresholds (ideally in respect of the total EU weight) for reporting of data would be helpful. The use of higher level indices as proxies for more detailed reporting requirements might also be considered.

Bibliography

Conniffe, D. and D. Duffy (1999), "House Price Indices – Methodological Issues", ESRI, Dublin

Diewert, W.E. (2010), "Alternative Approaches to Measuring House Price Inflation", Working Paper 10-10, University of British Columbia, Vancouver

Diewert, W.E., S. Heravi and M. Silver (2009), "Hedonic Imputation versus Time Dummy Hedonic Indexes", pp. 161-196 in *Price Index Concepts and Measurement*, W.E. Diewert, J.S. Greenlees and C.R. Hulten (eds.), Studies in Income and Wealth 70, Chicago: University of Chicago Press.

Duffy, D. (2009), "Measuring House Price Change, Working Paper No. 291", ESRI, Dublin.

Eurostat (2003), "Item 4.2: Methodological documents – definition of quality in Statistics", Working Group "Assessment of Quality in Statistics"

Fleming, M.C. and .J.G. Nellis 1984. *The Halifax House Price Index — Technical Details*, Halifax: Halifax Building Society.

Fleming, M.C. and .J.G. Nellis 1985. "The Application of Hedonic Indexing Methods: A Study of House Prices in the United Kingdom", *Statistical Journal of the United Nations,* ECE 3, pp. 249-270.

Fleming, M.C. and .J.G. Nellis 1992. "Development of Standardised Indices for Measuring House Price Inflation Indexing Methods Incorporating Physical and Locational Characteristics", *Applied Economics*, Vol. 24, pp. 1067-1085.

Hill, Robert (2011), "Hedonic Price Indexes for Housing", OECD Working Paper NO.36, OECD, Paris.

Li, W., M. Prud'homme and K. Yu (2006), "Studies in Hedonic Resale Housing Price Indexes," paper presented at the OECD-IMF Workshop on Real Estate Price Indexes held in Paris, November 6-7, 2006. http://www.oecd.org/dataoecd/2/25/37583404.pdf

Laferrere A., (2003), "Hedonic housing price indexes: The French experience", Bank of International Settlements Paper No. 21 http://www.bis.org/publ/bppdf/bispap21v.pdf

Shimizu, C, H.Takatsuji, H.Ono and K. G. Nishimura (2010), "Structural and Temporal Changes in the Housing Market and Hedonic Housing Price Indices," *International Journal of Housing Markets and Analysis*, Vol.3 Issue 4.

Appendix 1. Mortgage Dataset variables

Financial Institution	Borrower	Loan	Property
Institution Code	Number of Male	Amount of Approval	**Transaction Type**
Sequence Number	Number of Female	**Property Acquisition**	**Price of Property**
File Month	Age of Main	PA Nature	**Location County**
	Age of Second	NPA Nature	**Location Dublin**
	Gender Main	Loan Term	Location Abroad
	Gender Second	Initial Gross IR	**New or Second Hand**
	Marital Status Main	Rate Type	**Year Built**
	Marital Status Second	Years Fixed	**Dwelling Type**
	Employment Status Main	Loan Type	**Construction Type**
	Employment Status Second	Means of funding Gap	**Floor Area**
	Employment Sector Main	LA Clawback	**Plot size**
	Employment Sector Second	Loan Approval	**Rooms**
	Occupation Main	LA Date	**Bedrooms**
	Occupation Second	Drawdown	**Price of Property at Drawdown**
	Buyer Status Main	DD Date	
	Buyer Status Second	Total Loan Amount	**Use of Property**
	Gross A Income Main	Other Costs	
	Gross A Income Second	Indemnity Bond	
	Net M Income Main		
	Net M Income Second		
	Other Non Rental Income Main		
	Other Non Rental Income Second		
	Rental Income Main		
	Rental Income Second		
	Current Tenure Main		
	Current Tenure Second		
	Location County Main		
	Location Dublin Main		
	Location Abroad Main		
	Location Dublin Second		
	Location County Second		
	Location Abroad Second		

Appendix 2. Threshold rules

Table A. Thresholds for Number of Bedrooms by Property Type

Property Type	Allowable Range for number of Bedrooms
Detached House	between 2 and 6
Semi Detached House	between 2 and 5
Bungalow	between 1 and 6
Terraced House	between 1 and 5
Apartment/Flat	between 1 and 4

Table B. Thresholds for Floor area by Property Type and Bedrooms

Property Type	Number of Bedrooms					
	1	2	3	4	5	6
	Allowable Range for Floor area (metres squared)					
Detached House	-	40-239	60-279	100-339	120-419	120-479
Semi Detached House	-	40-179	60-199	80-219	80-339	-
Bungalow	10-219	20-2019	40-239	80-319	100-379	100-419
Terraced House	20-159	40-159	60-159	60-259	80-379	-
Apartment/Flat	20-119	40-139	40-159	40-219	-	-

Table C. Thresholds for Number of Rooms by Property Type and Bedrooms

Property Type	Number of Bedrooms					
	1	2	3	4	5	6
	Allowable Range for number of Rooms					
Detached House	-	4-8	5-12	6-14	7-16	8-17
Semi Detached House	-	4-8	5-10	6-12	7-14	-
Bungalow	3-8	4-8	5-10	6-13	7-15	8-15
Terraced House	3-6	4-7	5-10	6-12	7-13	-
Apartment/Flat	3-5	4-7	5-8	6-10	-	-

Appendix 4. Indices January 2005-March 2011

	National - All Properties	National - Apartments	National - Houses	Dublin - All Properties	Dublin - Apartments	Dublin - Houses	National excl Dublin - All Properties	National excl Dublin - House
Jan-05	100.0	100.0	100	100.0	100	100	100.0	100
Feb-05	100.4	100.1	100.5	101.1	101.7	100.8	100.1	100.3
Mar-05	100.6	99.9	100.8	101.2	101.3	101.2	100.3	100.6
Apr-05	101.3	99.7	101.6	102.2	101.5	102.4	100.9	101.2
May-05	102.0	99.7	102.3	102.8	101.1	103.4	101.5	101.9
Jun-05	102.9	99.6	103.4	103.5	101.5	104.2	102.6	103.1
Jul-05	104.3	100.0	105	104.7	102	105.6	104.0	104.7
Aug-05	105.9	101.5	106.6	106.4	103.5	107.3	105.6	106.2
Sep-05	107.2	103.8	107.8	108.5	106	109.4	106.5	107.1
Oct-05	109.0	106.2	109.5	111.1	108.3	112	107.9	108.3
Nov-05	110.0	106.9	110.5	112.6	109.2	113.7	108.6	109.1
Dec-05	111.5	108.3	112	113.7	110	114.9	110.3	110.7
Jan-06	111.8	108.4	112.4	113.9	109.9	115.2	110.7	111.1
Feb-06	112.6	109.1	113.1	114.9	110.5	116.4	111.2	111.6
Mar-06	113.1	109.0	113.8	116.0	110.6	117.8	111.6	112
Apr-06	114.6	111.0	115.2	117.9	113.4	119.4	112.7	113.2
May-06	116.8	112.9	117.4	120.5	115.3	122.1	114.7	115.3
Jun-06	119.0	114.7	119.7	123.6	117.4	125.6	116.4	117
Jul-06	121.3	115.7	122.2	127.0	118.2	129.9	118.1	118.7
Aug-06	123.5	117.8	124.4	130.3	120.9	133.3	119.8	120.4
Sep-06	125.1	119.9	125.9	131.6	123.1	134.4	121.5	122.1
Oct-06	126.0	120.9	126.8	132.4	123.7	135.3	122.5	123
Nov-06	126.1	120.7	126.9	131.9	123.4	134.6	122.9	123.4
Dec-06	127.3	120.8	128.3	133.1	123.7	136.2	124.0	124.7
Jan-07	128.4	122.5	129.3	133.4	125.9	135.7	125.6	126.4
Feb-07	129.6	123.9	130.5	134.5	127.4	136.7	126.9	127.6
Mar-07	129.9	123.2	130.9	133.9	126	136.4	127.7	128.4
Apr-07	130.3	121.6	131.7	134.0	123.3	137.8	128.1	128.9
May-07	130.2	121.7	131.5	132.7	123.1	135.9	128.7	129.4
Jun-07	130.2	122.5	131.4	133.4	123.9	136.6	128.4	129.1
Jul-07	130.2	123.0	131.4	133.3	125.4	135.9	128.5	129.2
Aug-07	130.4	122.1	131.7	133.6	124.1	136.8	128.5	129.3
Sep-07	130.5	121.5	132	133.5	123.4	136.9	128.8	129.7
Oct-07	130.1	121.5	131.4	133.1	122.6	136.6	128.3	129
Nov-07	130.1	120.7	131.6	132.7	121.3	136.7	128.6	129.3
Dec-07	129.5	121.3	130.8	131.2	122.6	134.1	128.4	129.1

191

	National - All Properties	National - Apartments	National - Houses	Dublin - All Properties	Dublin - Apartments	Dublin - Houses	National excl Dublin - All Properties	National excl Dublin - House
Jan-08	128.7	120.3	130.1	128.6	122.3	130.6	128.4	129.4
Feb-08	127.6	120	128.8	127.7	123.4	128.9	127.2	128.3
Mar-08	126.6	119.4	127.7	126.9	122.1	128.2	126.2	127.2
Apr-08	125.7	118.5	126.9	127.4	121	129.6	124.7	125.6
May-08	124.6	115.9	126	126.9	117.7	130.6	123.3	124.2
Jun-08	123.4	113.4	125.1	125.9	114.7	130.7	122.1	123
Jul-08	122.5	112.4	124.3	124.4	113	129.3	121.4	122.3
Aug-08	121.5	111.6	123.2	122.5	111.8	127.1	120.7	121.4
Sep-08	120.1	109.9	121.9	119.5	109.2	123.9	120.1	120.8
Oct-08	117.7	105.2	120	116	103.6	121.4	118.2	119
Nov-08	115.7	101.8	118.2	112.5	99.3	118.3	116.8	117.7
Dec-08	113.5	99.1	116.2	110.6	96.7	116.7	114.6	115.6
Jan-09	111.3	97.6	113.8	108.1	96.4	113.2	112.5	113.6
Feb-09	108.5	94.7	111	104.1	93.3	108.9	110.3	111.4
Mar-09	105.6	92	108.2	100	90.1	104.4	108.2	109.2
Apr-09	103.3	89.1	106	96.1	86.7	100.2	106.6	107.7
May-09	101.1	87.5	103.6	94.1	86	97.5	104.4	105.6
Jun-09	99.2	84.9	101.9	92.4	84.5	95.6	102.4	103.8
Jul-09	97.5	83.8	100.1	91.4	83.1	94.8	100.4	101.7
Aug-09	96.2	81.3	99	90.7	80.8	95.1	98.7	100.1
Sep-09	95.5	78.9	98.6	89.8	78.2	95.2	98	99.5
Oct-09	94.4	77.2	97.7	88.2	76.6	93.7	97	98.6
Nov-09	93.9	76.7	97.2	86.2	76	90.9	97	98.7
Dec-09	92.4	76	95.4	84.2	75.1	88.2	95.7	97.3
Jan-10	91.6	75	94.7	82.7	74.1	86.5	95.4	96.9
Feb-10	90.1	73.1	93.3	81.8	71.7	86.3	93.6	95
Mar-10	89.7	72	92.9	81.6	70.9	86.3	92.9	94.4
Apr-10	89.1	71.5	92.4	81	70.3	85.7	92.3	93.9
May-10	88	71.4	91.1	80	70.7	84.3	91.3	92.8
Jun-10	86.9	71.1	90	79.1	70.2	83.2	90.2	91.6
Jul-10	85.8	68.6	89	78.4	67.4	83.1	88.9	90.4
Aug-10	85.8	67.4	89.1	78.1	66.2	83.2	89	90.5
Sep-10	84.9	65.7	88.3	77.1	64.3	82.4	88.1	89.7
Oct-10	83.9	65.1	87.3	76.5	64.4	81.5	86.9	88.5
Nov-10	83.1	64.5	86.5	75.5	63.4	80.6	86.2	87.7
Dec-10	82.7	64.7	86	75.2	63.9	79.9	85.8	87.4
Jan-11	81.8	64.9	84.9	73.9	63.7	78.4	85.2	86.6
Feb-11	80.4	63.5	83.5	72.3	63.2	76.4	83.9	85.4
Mar-11	79	63.5	82.2	71	63.2	75.2	83.9	85.4

FIRST VOTE OF THANKS PROPOSED BY DAVID DUFFY, ESRI.

The launch of the CSOs residential property price Index is a very welcome development. A feature of mix adjustment using hedonic regression is a very onerous data requirement. We have seen the implications of this for similar indices, for example the Permanent TSB house price index which was based on data from one lender. The CSO index utilises data from 8 of the main mortgage lenders and so hopefully will have sufficient data to deal with fluctuations in the number of transactions associated with housing market cycles.

The index confirms the growth in house prices during the latter stages of the boom and shows that peak to trough house prices are down by nearly 40 per cent. The new indices also provide us with information on the movement of prices in the apartment market, down from their peak by over 50 per cent. I also note that the new indices suggest that the housing market peak occurred later than that recorded by other housing market measures.

When using any dataset the identification and exclusion of extreme outliers is important - prices that seem way outside the range that could be expected given the recorded characteristics of the dwelling. This may be the result of data errors or some undocumented characteristic. Unlike many papers on hedonic regressions this paper contains a detailed description of the data and the process the author worked through to arrive at the final dataset to use when constructing the index. This is not usually given much attention in the literature and so the work by Niall provides an insight into the process undergone to construct a dataset and provides a useful guide to those undertaking this work in the future.

As outlined in the paper the index is based on transactions, although data is gathered on approvals as well. This reflects the requirement by Eurostat to have a transaction based measure. This means that there will be a lag in the index, reflecting the difference between when the price was agreed and when the mortgage is drawn down. While approvals data may be more timely, use of approvals data is not without its difficulties. Loan approval does not always become a transaction, approval might be received for one property but the ultimate sale might be a different property. During the boom we had the experience of people "shopping around" and getting multiple approvals. So, while there is a lag I think this represents a reasonable trade-off between timeliness and quality.

I would have some concerns about the address variable. Generally, location is one of the most important variables in explaining the price of a dwelling. Currently the definition of location is quite broad – county, or Dublin postcode. We know that the market consists of a series of "local" markets that are not defined by county or postcode boundaries. These variables may not capture variation that exists within the boundaries. By using a broad location the index may not be capturing changes in the "quality" of dwellings location. Hedonic indices weight to a base period. Without precise location identifiers the CSO index may not be fully capturing changes in the location of dwellings that sell and so may still reflect some changes in the mix of properties selling in different periods.

Bourassa *et al* (2003) conclude that price predictions are found to be most accurate when based on the housing market segmentation used by appraisers. The use of a broad variable may not fully capture the influence of location on price, and the influence of other variables, such as age or plot size, may be diminished without more precise location identifiers. In addition, more precise location variables would allow the index to use other variables such as distance to school, work or transport, all of which have been shown to influence the price.

I was interested to see that the regression includes a variable for first-time buyers. When calculating the Permanent TSB index a dummy variable identifying whether or not the purchaser is a first-time buyer was included, arguing that the variable "can function as a proxy picking up omitted factors" (see Conniffe and Duffy, 1999). The inclusion of such a variable must, initially at least, be regarded as suspect given that it is a characteristic of the purchaser rather than the dwelling. However, the use of non-hedonic variables is not uncommon in the literature with a first-time buyer variable being used as a proxy to capture the effect of other characteristics for which there is no information. For example, Goodman and Ittner (1993) include occupant characteristics for the reason that "some household characteristics, especially income, may proxy for unobserved housing characteristics". Kiel and Zabel (1999) examine what constitutes a "neighbourhood" in the housing market. In the absence of immediate neighbourhood data they include the characteristics of the house owner as proxies for local neighbourhood quality. Turnbull and Sirmans (1993) use buyer characteristics in a hedonic regression as proxies for the impact of variations in the level of buyer information and search costs on house prices. It is evident that the range of variables included in hedonic regressions has broadened as researchers seek to gain further insights into the factors that influence house

prices. Indeed, Conniffe and Duffy (1999) found that there was not a significant difference between indices if the first-time buyer variable was included or excluded, albeit with data for a short time frame.

Finally, the paper also points to some promising developments for the future – data on the size of the mortgage market. This new CSO release provides data on index levels and rates of change. Perhaps the introduction and publication of standardised prices based on the indices might also be considered so that we have information on price levels as well as price changes. In addition, the mortgage dataset variables suggest that there is a rich dataset with details of the borrower, the mortgage and the property which hopefully in the future will become a valuable resource for those of us involved in housing market research.

It is my pleasure to propose a vote of thanks to Niall O'Hanlon for his interesting and informative paper.

References:

Bourassa, S.C., M. Hoesli and V.S.Peng, 2003, "Do Housing Submarkets Really Matter?", *Journal of Housing Economics*, Vol 12, pages 12-28.

Conniffe, D., and D.Duffy, 1999, "Irish House Price Indices: Methodological Issues" *Economic and Social Review*, 30, 403-423.

Goodman, J.l., and J.B. Ittner, 1993, "The Accuracy of Homeowners' Estimates of House Value", Board of Governors of the Federal Reserve System Working Paper No, 131, March.

Kiel, K.A. & Zabel, J.E., 1999, "The Impact of Neighborhood Characteristics on House Prices: What Geographic Area Constitutes a Neighborhood", Department of Economics, College of the Holy Cross Faculty Research Working Paper Series, No. 99-05, January

Turnbull, G.K., and C.F. Sirmans, 1993, "Information, Search, and House Prices", Regional Science and Urban Economics, 23, pp.545-557.

SECOND VOTE OF THANKS PROPOSED BY MARIAN FINNEGAN, CHIEF ECONOMIST, SHERRY FITZGERALD GROUP

Can I begin this evening by first of all thanking Seán Lyons and his fellow society members for the opportunity of speaking at this evenings' meeting.

The publication of the new Residential Property Price Index by the CSO was widely anticipated and warmly welcomed by the property industry at large and I think in particular by those of us who spend our days analysing and discussing the various performance indicators for the wider property market. I am therefore delighted to have this opportunity to commend Niall and his team for the excellent work done in compiling the new Irish Residential Property Price Index. It is truly a very positive step forward for an industry starved of vital statistics.

Before I begin to discuss the report I might just share a personal anecdote, to illustrate how far behind the curve the property industry is in terms of vital statistics. I remember when I joined Sherry FitzGerald some 15 years ago in 1996 with the remit of setting up a research department. I began by spending some time with the property research division of our sister company DTZ International in London. The head of research in London, Dr Peter Evans, gave me a lot of very good advice about the task ahead of me – a task he had begun some 25 years previously in the early 1970s. Bear in mind that Sherry FitzGerald were one of the first estate agents in Ireland to set up a research function and this really illustrates how badly served we are in Ireland, in terms of vital property related statistics, not just the residential market but also the commercial property market.

On that very note, having lived through what will be described as one of the strongest booms in a property market in recent decades, we are now enduring what I fear may be the worst property recession in the post war era and until last week we did not have a nationally recognised index on the performance of the industry.

Having said that, the gap in market was filled by various industry bodies all providing a variety of analyses of market trends. The vast majority of these indices had a significant value however confusion did emerge in the interpretation of these barometers. A generalisation emerged that if all barometers were analysing the same thing – the residential property market – then surely the results should be showing the same thing. Unfortunately the fact that they were not either analysing the same data or indeed the same markets got lost in the message.

With the benefit of hindsight it is often useful to look back at the results of the different indices and how we can interpret what they mean for the property market.

The new Residential Property Price Index and indeed its predecessor the Permanent TSB ESRI Index both analyse transaction prices of dwelling based on mortgage drawn downs. This in some ways is similar to the analysis by Daft and Myhome except for the fact that the portal sites are analysing asking prices while the other indices are analysing achieved prices. However, the binding characteristic that links all four indices is that they are all in one guise or another, analysing transactions.

The strength of the new CSO index over other indices obviously lies in the breath of its coverage and the fact that it is analysing achieved prices. In other words real live data rather than the aspirations of hopeful vendors, which can fluctuate a lot. This is particularly true during a challenging period in the market.

Just to look at that issue for a moment. Our own analysis of the Dublin market in the period 2005 to 2010 shows that that the average deviation between achieved and asking prices for each calendar year ranges from +11.9% in 2005 to -9.0% in 2010. Furthermore, it is probably worth noting that analysing markets at different point in a cycle brings with it different challenges - challenges that should be borne in mind, when interpreting results.

When one is analysing transactions in a normal market, one is probably typically analysing a proportionate representation of the entire market. The same however may not be true of a dysfunctional market. This is particularly notable during a property collapse as we are enduring at the moment. Because of a variety of factors, in particular though not exclusively, the limited liquidity in the market place, there is a concentration of activity around a certain type of property and a reduction in activity in other property types.

Firstly, first time buyers are in effect punching above their weight, they are favour in the mortgage market and they represent approximately 45% of all purchasers. This is significantly greater than their proportionate representation in the overall population. FTB are favouring a particular type of property, now they are typically buying 2 -3 bed houses. For a variety of reasons demand for smaller units and in particular apartments is particularly subdued. They would have largely been the property of choice of investors whose presence in the market is significantly reduced now. Also there has been a notable shift in general consumer sentiment away from apartment living of late.

Secondly, the upper end of the market, by which I mean properties valued over €750,000 and perhaps in particular the premium upper end, properties valued more than €1.5 million are particularly inactive. To illustrate this, > 90% of all sales through Sherry FitzGerald were for properties valued at less than €500,000. As such, properties in the upper end and if you like lower end of the market are unlikely be proportionally represented in an analysis of transactions at the moment. By that very fact, deflation levels in those markets are greater than overall averages. This is borne out by Niall's findings in relation to the apartment market.

As such overall levels of deflation determined through an analysis of transactions may not truly reflect the full extent of the market correction. One is analysing the most active end of the market and as such the area that is perhaps least impacted by the reduction in demand. This point should be taken into consideration in the interpretation of statistics from the new index.

Another point worth bearing in mind in our interpretation of statistics on the market lies in the lag which exists between the time the sale closes on a property and when the mortgage is drawn down. Niall, I understand, believes that the range here is 1 to 3 months. As such, when we discussing the results from March for the new Residential Property Index we should note that it largely reflects market activity in January for example. This is again particularly reflective during a period of crisis rather than a normal functioning market place.

With all of this in mind I might just for a moment comment on the Sherry FitzGerald index of prices which I am responsible for. This index has been in place since 1996 for Dublin and 1999 for the national market. It analyses trends in the second hand market only based on an analysis of a basket of properties in our locations nationwide. Each basket of properties was chosen based on a weighted profile of properties in each location. The basket extends to over 1,500 properties. Despite the different natures of the dataset there is still a very strong correlation between the analysis produced by the CSO on both the Dublin and Ireland housing markets in the period 2005 to March 2011 and that of the Sherry FitzGerald Index which has been in existence since 1996. Comparing the results for the Dublin market shows a correlation of 0.9695, while the correlation for the Irish market is somewhat weaker at 0.9590. That said the Sherry FitzGerald barometers are showing stronger levels of deflation. The national index shows price deflation of 51.1% from peak with a 55.8% deflation recorded in the Dublin. I would suggest however that the extent to which our figures exceed those of the new CSO Index could be explained by my previous comments in terms of the active an inactive elements of the market.

Before I conclude, this evening, it would I think be remiss of me not to return to my opening point in relation to the dearth of data on the property market. This new index is a welcome first step in producing good quality data on the Irish residential market. But let us hope that it is not the only step. For true transparency in the industry we need other statistics to be publicly available.

Firstly, the price achieved for all residential sales to be matter of public record. It should be stored in a database that was freely available to the public to search. Then and only then can we truly allow the public who are after all in most cases making one of the biggest investment decision of their lives access to the necessary information to allow them to make an informed decision. This data is the only data that will give consumers comfort in knowing the true value of houses in an area that they are considering investing.

Secondly, the volume of activity that is taking place in the market should be published on a quarterly basis. This too would prevent harmful speculation as to activity levels and allow full transparency in the market.

Finally, we need to ensure that the work of commenced by the Department of the Environment on the supply of houses in the market place is updated on a regular basis with further clarity supplied on the profile of those units.

To conclude, therefore Ladies and Gentlemen, can I once again commend Niall and his team on the work done on this new index and say that we look forward to a future of greater information in an area that is such a vital element for our economy.

Thank you very much

DISCUSSION

Steve MacFeely:- I would like to congratulate Niall on his paper. The Residential Property Price Index is a very important development for CSO. Both the index and this paper will I hope contribute to further informing policy makers and the public alike. In his paper, Niall has given an excellent outline of the relative strengths and weaknesses of the possible data sources available, the challenges in compiling a RPPI in a state where no postal codes are available and the consequent methodology used to compile the index. Very importantly, Niall has also contextualised his decisions in view of impending EU legislative requirements.

Niall raised a very important point during his presentation: that of data infrastructure. If the datasets listed contained spatial or geo-coded information or contained the transaction folio number, our ability to develop more refined analysis, through data matching or linkage would be significant. The importance of post codes or spatial identifiers from a statistical perspective doesn't require any explanation at this society but it bears repeating that it is a serious gap in our infrastructure. In contrast a transaction folio number already exists but it is not captured on all of the datasets listed by Niall. If this single issue were addressed our data infrastructure would be so much richer.
.