Property Valuation

Developing, Assessing and Deploying a Valuation Model for Local Property Tax

March 2013

Local Property Tax
LPT

Research & Analytics Branch
DATA - INFORMATION - KNOWLEDGE

Revenue
Cáin agus Custaim na hÉireann
Irish Tax and Customs

Report Author:

Dr Keith Walsh
Economist,
Research and Analytics Branch,
Planning Division,
Office of the Revenue Commissioners.

Table of Contents

List of Figures

List of Tables

Executive Summary

This is a technical paper that summarises the economic model developed to assist with the valuation of residential property in the context of the introduction of Local Property Tax (LPT).

The model used is a hedonic econometric regression model. This approach disaggregates the value of a property into constituent parts to assess the contribution of different characteristics and provides a weighting for each.

Information on property characteristics is taken from a range of sources:

- Geo-Directory: a list of all properties in Ireland, their type and location;
- Spatially derived data that indicate relative distances of all residential properties from a series of key amenities and services;
- Geographically linked data from sources such as the CSO's 2011 Census results at small area level.

As well as characteristics information, known property values are required to train and test the regression model. Valuation data on property sales since January 2010 are taken from Revenue's electronic stamp duty system and this information is supplemented with additional valuations from NAMA and some commissioned by Revenue from professional valuers.

Effectively, the regression model estimates a value by applying the average for a type of property in an area. Adjustments are made to these average values within in each area, to reflect property specific or neighbourhood-level information, such as distances to amenities or services. The guidance for property owners on Revenue's website shows the average of the estimated values for all properties within each Electoral District.

This type of model is widely used in real estate valuation and property tax administration. Although it is not an exact science, in that no model can accurately value all properties in a country, testing of the model's performance shows that it functions well in the majority of cases. Its performance is within the internationally recognised standards for this type of model.

1 Introduction

This paper outlines the economic model developed to assist with the valuation of residential property in the context of the introduction of Local Property Tax (LPT).

LPT, an annual charge on all residential properties in the State, will come into effect from 1st of July 2013. Announced in Budget 2013 and enacted in the Finance (Local Property Tax) Act 2012, the tax applies on the market value of a property. LPT is a self-assessed tax so it is the responsibility of each of property owner to value their property.

In this context, Revenue has developed a valuation model to provide guidance to property owners seeking to value their property.

Property valuation is not exact science and no model can accurately value all properties in a country. The modelling approach adopted by Revenue is one that is widely used in real estate valuation and property tax administration in other countries. The method aims to develop a model that can systematically value properties, in a manner that generates accurate estimates in the majority of the cases but cannot be expected to generate the correct value for every property.

The paper is structured as follows. Section 2 presents an overview of the modelling approach and data used. Section 3 describes in detail the data sources employed. Section 4 shows the model estimation, the results and performance testing. Section 5 examines the application of the model to value all properties in the State. The final section concludes.

2 Methodology

This section presents an overview of the approach used to create a valuation model in a short summary of the work described in the Sections 3 through 5.

The market value of a property is a function of its characteristics (for example, location and size). Prices also depend on the market at the time of valuation.

Using an econometric technique known as hedonic regression analysis, the value of a property can be disaggregated into constituent parts associated with each characteristic. Essentially, this assesses the contribution of different characteristics to values and provides a weighting for each characteristic.[1]

Analysis of property values, such as the Central Statistics Office (CSO) monthly price index or the quarterly reports on average asking prices published by economists using myhome.ie and daft.ie listings data rely on detailed information on property characteristics for their hedonic regression models.[2] The primary variables normally used in this type of analysis are location, property size (floor size), number of rooms (or bedrooms) and type of house (terrace, detached, etc).

From the perspective of LPT, it is necessary to produce indicative or average values for all properties in the country. Detailed data on characteristics (for example, size or number of rooms) are not readily available for all properties. However, research by the ESRI has shown that property values in Ireland can be predicted with reasonable accuracy with only a few characteristics known.[3]

Geo-Directory provides the most comprehensive listing of residential properties in Ireland. While it does not record information on rooms or floor size, Geo-Directory does have other characteristics that are predictive of values. In addition, it forms a key component of Revenue's newly constructed register of property owners, also assembled for the purposes of LPT.

[1] A hedonic model is more detailed than taking a simple average (mean or median) of all prices. Instead, the hedonic approach attributes each observation (a property's price) to its characteristics (floor size, bedrooms, etc). Such models control for the contribution of these characteristics and reflect the change in underlying property values rather than a change in the mix of properties. For example, an increase in 5 bedrooms houses would increase average prices, everything else being equal, unless controlled for.
[2] See, for example, *The Residential Property Price Index* January 2013 (produced by the CSO), *The Myhome.ie Property Barometer*, Q4 2012 (prepared by C. Kellegher, DKM Economic Consultants) and *The Daft.ie House Price Report*, 2012 Q4 (prepared by R. Lyons).
[3] Mayor, K., Lyons. S, and Tol, R. (2010), *Designing a Property Tax Without Property Values*, ESRI Working Paper 352.

Characteristics information is also assembled from other sources, as discussed in Section 3. The aim is to build up information on every residential property in the country, some of which is property specific, other sources relate only to the area in which the property is located.

As well as property characteristics information, data are needed on valuations to train and test the regression model. Known valuations are extracted from Revenue's stamp duty system and this information is supplemented with additional valuations from NAMA and some commissioned by Revenue from professional valuers.

These valuations are linked to properties on Geo-Directory to provide data on property characteristics. This produces a training dataset, a list of properties with known valuations and their associated characteristics.

Using the dataset, property characteristics are tested to see how well they predict property values using regression analysis. The training allows for comparisons of known and predicted valuations, to judge model performance and accuracy.

The resulting valuation model (the regression model) produces estimated coefficients for each property characteristic in the training dataset. These coefficients are then applied to the same characteristics in the full data listing all residential properties (around 2 million properties) in the country to estimate a value for each individual property.

Average values for areas of the country are then calculated and these are the basis for the Revenue guidance provided to property owners to assist with the self-assessment of the value of their properties for LPT.

3 Data Sources

As described in Section 2, there are two main areas of data assembled for the valuation model. These are known property values (Section 3.1) and data on property characteristics likely to be predictive of value (Section 3.2)

3.1 Known Property Values

To build a model capable of predicting property values, known valuations are required to train the model. This requires sourcing of data on properties where Revenue can be relatively confident that a value is known and represents a reasonable approximation of the market or chargeable value for LPT purposes.

Three sources are used: Revenue's internal stamp duty data, valuation data from NAMA and valuations commissioned by Revenue from professional valuers.

3.1.1 Stamp Duty Transactions [4]

When a property is sold or when there is a transfer of ownership, a stamp duty return must be filed. The return captures the consideration value of the property and the market value (if it differs from the consideration value).

Since 1st of January 2010, Revenue has operated e-Stamping, which allows for electronic paying and filing of stamp duty returns. The data cover the period January 2010 to December 2012 inclusive.[5]

To produce a dataset suitable to assist with LPT valuation, some transactions are filtered from the e-Stamping data and data cleaning is undertaken, to remove:

- All non-property transactions;
- All non-residential property and mixed used property transactions.
- Duplicate records (the most recent record is used);
- Property transactions with no contract type or contract type "none";
- Transactions with a date of execution not between 2010 and 2012;

[4] A separate report documents the analysis of the e-Stamping data in detail – see *Property Valuation – An Analysis of Revenue's e-Stamping Data*, 2013 (by K. Walsh, Research & Analytics Branch).
[5] The same data source forms the basis of the property price register published by the Property Services Regulatory Authority (available at http://www.propertypriceregister.ie/).

In cases where the stamp duty return indicates that the consideration value is not equal to market value, the market value of the property is used if indicated. Further excluded are:

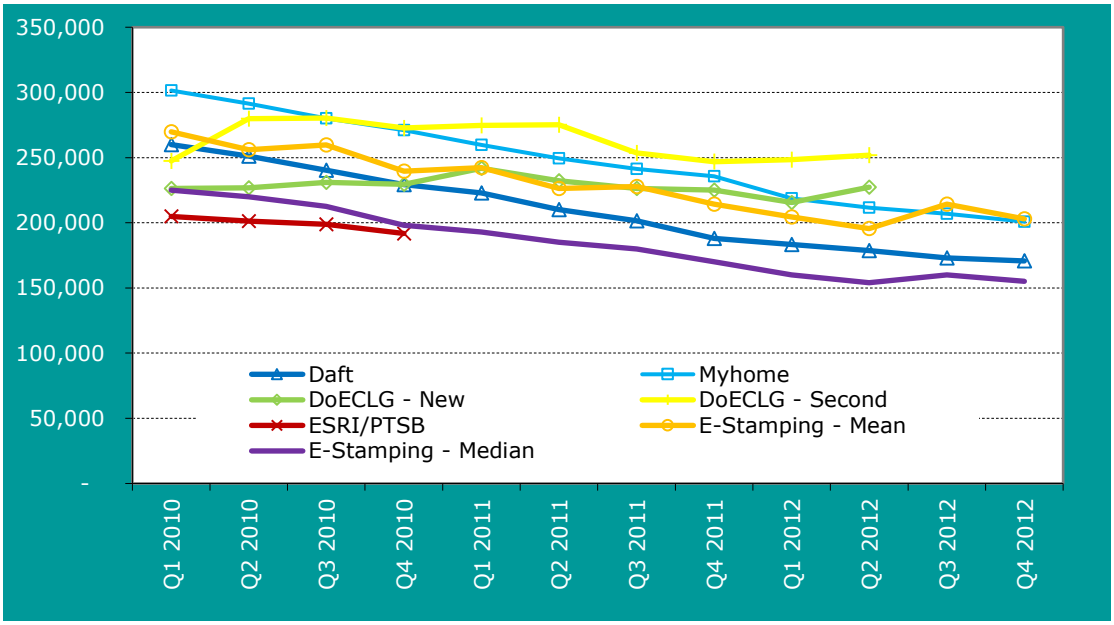
- Any transactions where the return indicates a family or business link between buyer(s) and seller(s);
- Any transactions where the return indicates the transaction is a part of a larger series of transactions;
- Any transactions where the return indicates a fractional interest in the property is being transferred;
- Any transactions where the return indicates shared ownership;
- Any transactions with values below under €25,000.[6]

The result is a dataset that contains only residential property sales where the valuation reflects the market value. The final dataset has 17,400, 15,000 and 19,200 transactions for 2010, 2011 and 2012 respectively.

Figure 1 shows average values from e-Stamping data (mean and median) and a number of independent property price measures. There is a strong correlation across the measures. Differences are a mixture of reasons: daft.ie and DoECLG[7] prices for new houses are VAT inclusive[8]; the measures use different approaches to calculate prices; possible undervaluation to evade stamp duties; and differences between asking and sale prices.[9]

Figure 1 – Average Price Comparisons

[6] A few of the largest transactions are also manually removed, as they appear to be data errors.
[7] Department of Environment, Community and Local Government.
[8] Data from e-Stamping are adjusted for VAT in the case of new builds.
[9] *The Daft.ie Analysis of the Residential Property Price Register,* 2012 Q3, suggests a differential of 10 to 15 per cent between asking and sales prices in 2012.

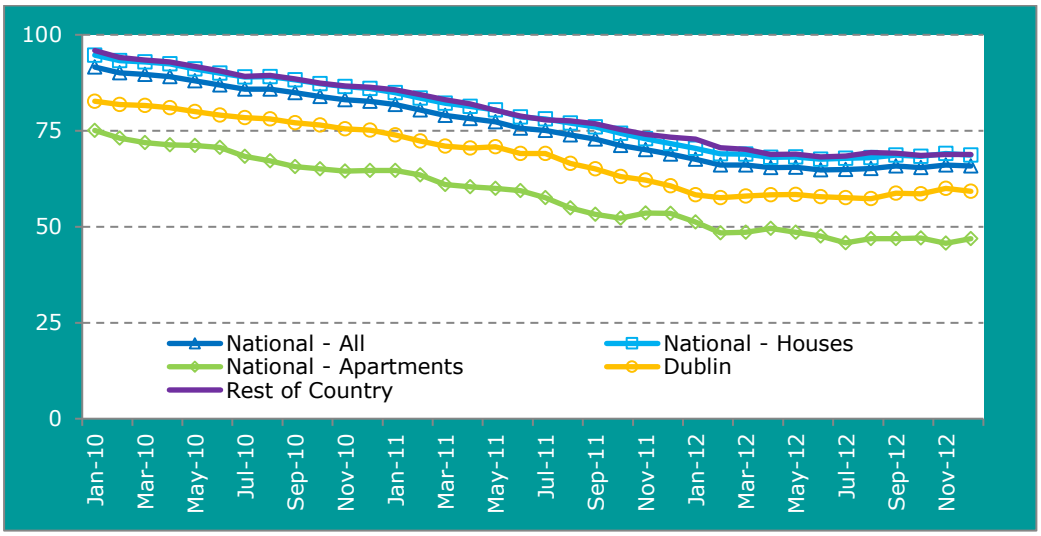


Source: Revenue analysis of data sources cited in Figure.

The data in Figure 1 and more detailed analysis suggests that these stamp duty data are representative of market values of residential properties. Or, at least, are comparable to the existing measures tracking ongoing property sales.

As Figure 2 indicates, property values in Ireland have declined over the period covered by the stamp duty data. Therefore, all e-Stamping data are deflated using CSO and daft.ie price indices to approximate values for end June 2012. Prices have stabilised since June 2012 in most categories and no further adjustments are made. It is assumed that all resulting values used in the analysis are reflective of current market prices in quarter one 2013.

Figure 2 – Residential Property Prices Changes 2010-2012



Source: Revenue analysis of CSO data. Note: Base January 2005 = 100.

Following extraction of the data from the e-Stamping system, the properties are matched to Geo-Directory records. Using address matching and geo-coding software, 34,400 (67 per cent) of the 51,600 e-Stamping properties are successfully matched to a Geo-Directory address point.

Table 1 shows the distribution of matched properties across the LPT valuation bands.[10] For comparison, the Table also shows analysis conducted by Department of Finance for the interdepartmental group examining the design of LPT [11] Differences in data sources probably underlie the variations. The Department of Finance estimates are based on CSO data from financial institutions' mortgage records (stamp duty returns capture all transactions regardless of whether mortgaged or not). These data are for 2005 through 2009 and adjusted for price changes to mid-2012 compared to the stamping data adjusted since 2010.

Table 1 – Distribution of e-Stamping Property by Value Band

Band	e-Stamping	Interdepartmental Group
0 to 100K	17.9%	7.26%
100K to 150K	24.9%	33.65%
150K to 200K	21.3%	32.05%
200K to 250K	13.0%	14.91%
250K to 300K	7.3%	6.21%
300K to 350K	4.8%	2.72%
350K to 400K	3.0%	1.37%
400K to 450K	2.0%	0.72%
450K to 500K	1.3%	0.39%
500K to 550K	1.0%	0.23%
550K to 600K	0.7%	0.12%
600K to 650K	0.5%	0.10%
650K to 700K	0.5%	0.07%
700K to 750K	0.3%	0.03%
750K to 800K	0.3%	0.03%
800K to 850K	0.2%	0.03%
850K to 900K	0.1%	0.02%
900K to 950K	0.1%	0.03%
950K to 1M	0.1%	0.01%
Over 1M	0.7%	0.04%

Source: Revenue analysis and Interdepartmental Group data.

3.1.2 NAMA Valuations

The e-Stamping transactions form the core of the valuations data. These have been supplemented by seeking data from the National Assets Management Agency (NAMA).

In accordance with Section 204 of the National Asset Management Agency Act 2009 and the amendment of this by Section 154 of the Finance Act 2010, Revenue is entitled to exchange data with NAMA. Following a request from Revenue, NAMA supplied data on

[10] There are twenty LPT bands: €0 to €100,000, 18 bands of €50,000 intervals from €100,000 to €1,000,000 and a single band for all properties above €1m.
[11] See *Design of a Local Property Tax*, Report of the Interdepartmental Group, published in 2012.

valuations of residential properties related to loans transferred from financial institutions to NAMA.

The NAMA valuation data are cleaned and the observations reduced, primarily to remove multiple properties within specific locations. For example, if the NAMA data included 20 houses in one estate, one property is selected as the representative property for that estate.

All of the valuations conducted for NAMA were established at market prices in November 2009. In the same way as the stamp duty data, these valuations are deflated to end June 2012 levels using CSO indices.

Using address matching software, 1,000 NAMA valuations are successfully matched to an address point in Geo-Directory.

3.1.3 Professional Valuations

From early iterations of the valuation model, it become clear that observations (valuations) were lacking or limited in certain areas of the country. To address this, a professional valuer was contracted to carry out around 1,300 valuations in locations specified by Revenue, targeted at areas with limited valuations from the sources discussed in Sections 3.1.1 and 3.1.2.[12]

The valuations specified by Revenue focused on the Dublin and the other urban areas. They were conducted in late January and early February 2013 and merged with stamp duty and NAMA valuations data described above.

This approach of using professionally valued properties to supplement sales or other valuation data, is used elsewhere. For example, in Northern Ireland, such "Beacon" valuations were combined with sales records in the revaluation of residential property values conducted in 2007.[13]

3.1.4 Valuations Data Summary

From the three sources described above, the final dataset of known valuations linked to properties in Geo-Directory totals 36,800, all with valuations approximated to current

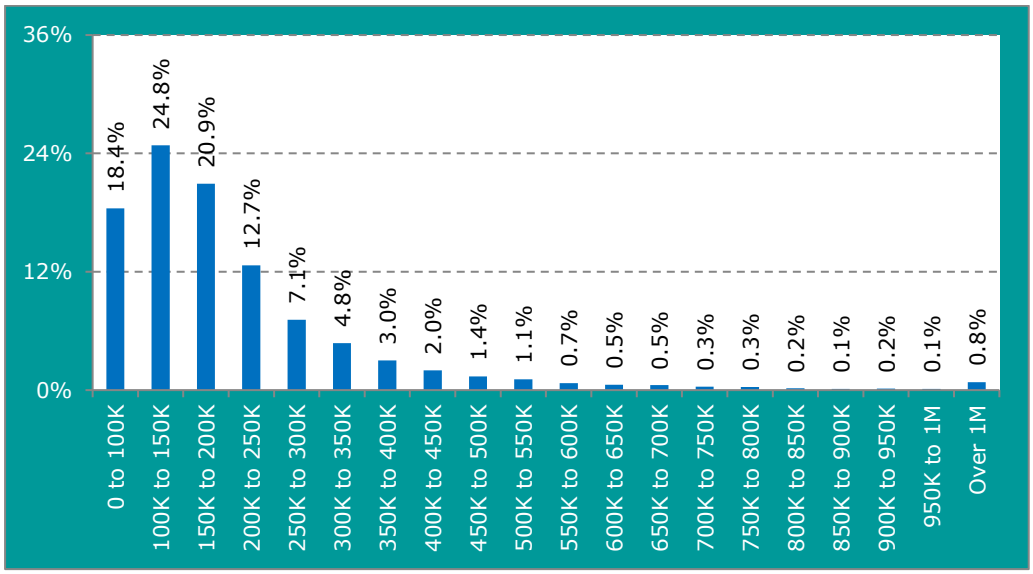
[12] Revenue did not issue a list of individual property addresses to be valued, rather a list of property locations and types were submitted to the valuer with the request to return an average or indicative value for that type of property in the specified location.
[13] See Gloudemans, R. and Montgomery, E. (2008), *Domestic Revaluation in Northern Ireland*, Journal of Property Tax Assessment and Administration, Volume 5, Issue 4 or McCluskey, W. and Woods, N. (2010), *Property Tax Reform: The Northern Ireland Experience*, Lincoln Institute of Land Policy Working Paper.

market values in early 2013. This dataset forms the basis for the valuation model implemented in Section 4.

The coverage of the valuation data is good. For example, at county level, the lowest number of observations in a county is 171. The comparisons to third party property price information show the data are consistent. However, a caveat that needs to be borne in mind, in relation to all of the analysis, is that e-Stamping valuations form the majority of the dataset used in the model. These are based on transfers of properties in the last three years. It is an underlying assumption of the approach that such transaction based data are reflective of the overall stock of properties.

Figure 3 shows the distribution of the training data valuations across the twenty LPT valuation band.

Figure 3 – Distribution of Training Data Properties by Valuation Band



Source: Revenue analysis.

3.2 Property Characteristics Information

Property characteristics information is drawn from a number of sources. The aim is to assemble information on characteristics of properties or areas that are likely to be predictive of the value. The choice of variables is largely driven by the availability of data. For example, Major *et al*. (2010) adopt a similar modelling approach, also in Irish context, but use five main characteristics variables (floor size, type, room number, bedroom number and age). However, they employ survey data, these characteristics data are not available for all properties in the country and so the approach cannot be adopted in this model.

3.2.1 Geo-Directory – Property Specific and Area Information

Geo-Directory is a database produced and maintained by An Post and Ordnance Survey Ireland.[14] It provides a list of addresses and locations for all properties in Ireland. It strength lies in the detail of the location information – each property is geo-coded. It covers both residential and commercial buildings, as well as vacant and derelict properties.

Geo-Directory provides information on each property relating to type (terraced, detached, etc), location (to numerous levels), use (residential, commercial or mixed) and other variables.

[14] See www.geodirectory.ie and Fahey, D. and Finch, F. (2012), *Geo-Directory Technical Guide*, by Ordnance Survey Ireland and An Post.

3.2.2 Spatially Derived Data – Distances to Amenities and Services

Using Geo-Directory as a register of all properties, a series of spatial variables are derived for each property. Revenue worked with geo-specialist consultant to link data on relative distances of all residential properties (using GeoDirectory geo-coded locations) to a series of key amenities and services (such as transport, health, education, emergency services, coastlines).

Effectively, given relatively small distances between properties in many locations, while these are property specific variables, their impact is only differentiated when moving from street to street or similar.

3.2.3 Geographically Linked Data – Area Based Characteristics

Also based in Geo-Directory location data, properties are geographically linked to publicly available sources such as the Central Statistic Office Census 2011 results at small area level and the Pobal HP Deprivation index (Haase and Pratschke, 2012).[15]

Other data are derived by small area from different sources.[16] For example, vacancy rates in each small area are calculated from Geo-Directory data. Average floor size per small area is calculated from Building Energy Rating (BER) data.[17] Although these measures are not property specific, given that small areas generally cover between 50 and 200 properties, this allows for a fairly refined degree of granularity for the characteristics variables included at this level.

3.2.4 Summary of Property Characteristics Data

As a summary, Table 2 shows all property characteristics variables used in the model. The characteristics are a mixture, some are specific to the property, others relate to the location (usually at small area level). These characteristics data are assembled for all 2 million residential properties in Geo-Directory.

[15] See http://www.cso.ie/en/census/ and www.pobal.ie/Pages/New-Measures.aspx respectively.
[16] "*Small Areas are areas of population comprising between 50 and 200 dwellings created by The National Institute of Regional and Spatial Analysis on behalf of the Ordnance Survey Ireland in consultation with CSO*" (Source: CSO website). There are about 18,500 small areas in the country.
[17] Revenue received BER data from the Sustainable Energy Authority of Ireland (SEAI) under Section 151 of the Finance (Local Property Tax) Act 2012.

Table 2 – Property Characteristics Variables

Variable(s)	Source	Description
Property Type	Derived from Geo-Directory	Type of building: detached, semi-detached, terraced, bungalow or apartment/flat. A binary variable is included for each property type.
Locality Type	Geo-Directory	Type of Locality: urban, rural, housing estate, village or industrial estate. A binary variable is included for each locality type location.
Year of Construction	Derived from Geo-Directory	Derived from the "create date" field. May not necessarily reflect the year of the construction and is only included as a binary variable: 0 for pre-2000 and 1 for post-2000.
Residential / Commercial Shared Building	Geo-Directory	Derived from commercial delivery points. This is included as a binary variable: 0 if the building is only residential and 1 if the building includes a commercial property.
Location	Geo-Directory	Location of the property: Electoral Districts are used for Dublin (City and County) and within the other 4 cities.[18] Local authority is used elsewhere in the country. A binary variable is included for each location.
Residential Mix	Derived from Geo-Directory	Using the Geo-Directory use marker (residential, commercial, both or unknown), the percentage of residential properties in each small area is calculated.
Vacancy Rate	CSO Census 2011	The census results indicate the percentage of vacant residential properties in each small area.
Room Number Mix	CSO Census 2011	The census results indicate the percentages of properties with 1 room, 2 rooms,… , up to 8 or more rooms for each small area. A variable is included for each room number.
Property Type Mix	CSO Census 2011	The census results indicate the percentages of properties of different type (detached, bungalow, etc) for each small area. A variable is included for each property type.
Property Age Mix	CSO Census 2011	The census results indicate the percentages of properties constructed in each of the last 5 decades and pre-1960 for each small area. A variable is included for each decade.
Ownership Type Mix	CSO Census 2011	The census results indicate the percentages of properties by ownership (owner occupied, rented, etc) for each small area. A variable is included for each ownership type.
Utilities Mix	CSO Census 2011	The census results indicate the percentages of properties by type of sewage system (public scheme, septic tank, etc) for each small area. A variable is included for each type.
Relative Distances to services and amenities	Spatially Derived	Distance of each property to services / amenities: airport, dentist, GP, hospital, retail, Garda & fire station, primary & secondary school, pharmacy, train / Luas station and coast. A separate variable is included for each amenity or service.
Deprivation Index	HP Pobal	For each small area, the index measures the degree of relative affluence or disadvantage based primarily on demographic, social and labour market indicators.
Floor Size Average	Sustainable Energy Authority of Ireland	For each small area, the average floor size of properties is estimated from BER certificate data.
Extension Indicator	Sustainable Energy Authority of Ireland	From BER certificate data, a rough estimate of the proportion of properties that have constructed an extension to the original property is derived for each small area.
Energy Rating	Sustainable Energy Authority of Ireland	For each small area, the average energy rating is included from BER certificate data.

Source: Revenue analysis of data sources listed in the Table.

[18] "*There are 3,440 Electoral Divisions (EDs) which are the smallest legally defined administrative areas in the State*" (Source: CSO website). There are 34 Local Authorities: 4 in Dublin, 1 in each of the four Cities (Cork, Galway, Limerick, Waterford) and 26 based on the remaining administrative counties (with North and South Tipperary separately distinguished) (Source: CSO website).

4 Model Results and Performance

4.1 Choice of Modelling Approach

A hedonic regression model is used to estimate values based on a variety of property characteristics. The accuracy of the model depends on the coverage and quality of the data available. The value of a property is regressed on a series of variables representing characteristics (some specific to the property itself, others related to its location) and an implicit value is estimated for each characteristic.

This type of approach is often referred to as an Automated Valuation Model (AVM) or Computer Assisted Mass Appraisal (CAMA) when applied to systematic valuations of properties. AVM is defined as "*a mathematically based computer software program that produces an estimate of market value based on market analysis of location, market conditions, and real estate characteristics from information that was previously and separately collected. The distinguishing feature of an AVM is that it is an estimate of market value produced through mathematical modeling*".[19]

AVM / CAMA approaches are commonly used and accepted in real estate valuation and property tax administration. They are widely employed at state level in the US and the 2007 revaluation of residential property in Northern Ireland also used AVM.[20]

In an Irish context, hedonic regression is used by researchers in the Economic and Social Research Institute (ESRI) and elsewhere.[21] As noted above, the main Irish property prices indices from the CSO, myhome.ie and daft.ie are based on hedonic models, as was the now ceased ESRI/ Permanent TSB index.

4.2 Model Implementation

As described in Section 3, the data assembled contain 36,800 properties where the valuation is known. These valuations, combined with the characteristics information for those properties, form the training dataset.

The regression model is estimated on the data for the 36,800 properties. The dependent variable is the log of value. The explanatory variables, that are assumed to be predictive of value, are those shown in Table 2.

---

[19] International Association of Assessing Officers (IAAO) (2003*), Standard on Automated Valuation Models (AVMs)*.
[20] Gloudemans and Montgomery (2008).
[21] See Mayor et al. (2010), Callan, T., Keane, C., Savage., M. and Walsh, J. (2012), *Analysis of Property Tax Options – A Report to the Interdepartmental Expert Group on Property Tax*, Keane, C., Walsh, J., Callan, T. and Savage, M. (2012), *Property Tax in Ireland: Key Choices*, ESRI Renewal Series Paper 11., or Lyons, R. (2012), *East, West, Boom & Bust: The Spread of House Prices and Rents in Ireland, 2006-2012*.

In effect, there are three sub-models (separate regressions), dividing the country in three different areas based around Local Authorities:[22]

- Dublin (Dublin City Council, Dun Laoghaire-Rathdown County Council, Fingal County Council and South Dublin County Council) – 15,693 properties;
- Other cities (Cork Corporation, Limerick Corporation, Galway City Council and Waterford Corporation) – 3,039 properties;
- Rest of the country (all remaining county councils) – 18,014 properties.

The variables listed in Table 2 are those used in the final iteration of the model. Some other variables were tested initially but not found to be predictive of value. Not all of those included in the final iteration of the model are statistically significant in every regression but all contribute in one of the three sub-models.

Regression analysis can be influenced by outliers – observations (property values in this case) that vary considerably from the average. Property, in particular in certain parts of the country, can be quite heterogeneous. It is possible for one property to sell at a high price and distort model results if the area is sparsely populated in the training data. For this reason, a fairly aggressive approach to outliers is taken in the regression analysis. The top 10 per cent and bottom 10 per cent of properties in terms of valuation are dropped.

The model is a linear regression implemented using OLS. An ordered logit model, that seeks to predict the probability that a property's value falls into defined bands, rather than specific valuations, was also tried. The results were very close to the linear regression, confirming the model is not method dependent. Several other statistical, non-regression techniques were also benchmarked against the linear regression with similar results.

4.3 Performance

Table 3 summarises a series of measures regarding the performance of the model (effectively the three sub-models).

Table 3 – Model Performance (Training Data)

Measure	Dublin	Other Cities	Rest of Country	*Overall*
Number of Observations	12,254	2,436	13,941	*28,631*
R – Squared	64.9%	56.6%	45.0%	--
Median Absolute Value Error	€28,167	€20,384	€20,016	*€22,879*
Median Assessment/Sales Ratio	0.995	0.994	0.989	*0.995*
Coefficient of Dispersion	16.98	19.81	19.36	*18.33*
Price Related Differential	1.07	1.07	1.09	*1.07*
Value In Correct LPT Band	41%	53%	53%	*48%*

[22] As noted in Table 2, the location variable included in the regression equation differs depending on the sub-model. Electoral District is used for Dublin and the other cities, Local Authority for the rest.

Source: Revenue analysis.

From an econometric perspective, the R-squared measures the percentage of variation in values in the training dataset explained by the regression. All three are reasonable for this type of model.

The absolute value error is the average of the differences in predicted (from the regression model) and true values (from training data), in absolute terms. While the average errors are low in the training data, the spread is relatively wide. The median error of €22,879 overall indicates that half of all errors are less than this amount. Quartile analysis shows 25 per cent of errors are less than €10,600 and the same percentage are over €42,300.

Assessing model performance is difficult without an appropriate benchmark. The International Association of Assessing Officers (IAAO) is one of the leading practitioner bodies in property valuation and taxation globally. They recommend a number of indicators to measure model performance.

The Assessment to Sales (A/S) ratio compares the assessed value of property (from the regression model) to the known market value (from training data). A median level of 0.9 to 1.1 is considered acceptable by the IAAO. A ratio just below 1 indicates a small degree of under-valuation (over 1 would indicate over-valuation). The three sub-models and the overall model easily meet the required standard for this indicator.

The Coefficient of Dispersion (COD) measures the average deviation of A/S ratio from median ratio, effectively the spread of errors around the average error. IAAO guidelines suggest the COD should lie between 5 and 20 depending on the property mix.[23] In the sub-models and overall, the COD is below 20, indicating the valuation model functions within acceptable standards for such models. Although the Cities and Rest of the Country models are close to 20, the IAAO guidelines suggest models for rural areas (given the heterogeneous nature of the properties) can have higher COD.

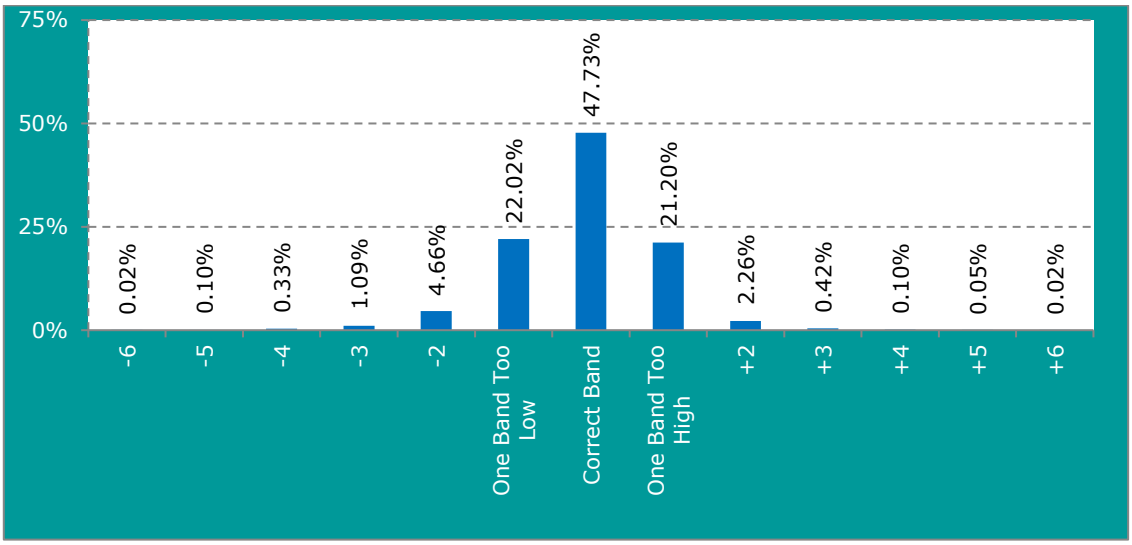
The Price Related Differential (PRD) measures the progressivity or regressivity of the model. A PRD above 1.03 indicates a degree of regressivity, a PRD below 0.98 indicates progressivity (IAAO, 2010). All three sub-models are slightly in the regressive range, indicating low value properties tend to be more over-valued, in relative terms, than higher

[23] See International Association of Assessing Officers (IAAO) (2010), *Standard on Ratio Studies*. IAAO (2010) notes that in new or homogenous areas the COD should be between 5 and 10, older or more heterogeneous areas between 5 and 15 and rural residential, recreational / seasonal, multi-family dwellings and income producing property areas between 5 and 20. Given the range of areas covered by the model, effectively all of these categories, the reasonable standard to apply is 5 to 20.

value properties. This reflects the difficulty in creating a model to capture the full range of property values.

As a final measure, the last line in Table 3, also shown in Figure 4, indicates the allocation of training data cases to the correct LPT value band by the model. Zero indicates the property is the correct band, +1 the property is placed one band too high, etc. Overall, 91 per cent of properties from the training data are placed in the correct valuation band or either one band above or below overall.

Figure 4 – Valuation Band Prediction Accuracy (Training Data)



Source: Revenue analysis.

By way of comparisons on model performance, Mayor *et al*. (2010) analyse a 2002 ESRI representative survey of 40,000 Irish households that includes information on property values.[24] Mayor *et al*. test options to model property values using limited characteristics information on this data.

Mayor *et al*. find an absolute value error of €3,888 to €12,353 depending on the choice and combination of characteristics variables used. This is the lower than the model results in this paper but the standard deviations of the errors are quite large. Depending on model choice, the Mayor *et al*. model places 75 to 84 per cent of property value predictions are in correct band. However, Mayor *et al*. use bands of €150,000 based on the Commission on Taxation report.[25] The Revenue generated model places 75 per cent in correct band when the same ranges are used. The COD in Mayor *et al*. is around 27.

[24] The models used in the analysis are based on about 25,000 observations.
[25] Report of the Commission on Taxation (2009).

Overall, the model performs within acceptable international standards for both the A/S ratio and the COD. Irish based analysis of this type is limited but the performance is in line with those in Mayor *et al*. (2010) and Callan *et al*. (2012). This is despite the use of variables (bedroom number, floor size) in those studies that are traditionally expected to be more predictive of value than the characteristics used in the model in this paper.

4.4 Analysis of Errors

It appears from the training dataset that in about 10 per cent of properties, the model is not very accurate (the prediction is more than one band out). It is useful to examine these cases to identify if any systematic error occurs.

Examining performance by building types, the model performs better on semi-detached and terraced houses and weaker on detached and apartments. Most likely, this arises because a mixture of reasons – limited numbers of apartments in the training data and a wide range of variation in detached properties.

Analysis of errors by area shows that the weakest performance occurs in Dublin City Centre and Dun Laoghaire Rathdown County Council, perhaps to be expected given the greater variation in properties (and their valuations) in these parts of the capital. There are significant numbers of properties in these locations and weak performance in these areas does weigh on national outcomes.

Model performance is slightly better with more recently constructed properties (post-2000) than older properties.

Although small in number, properties that are mixed use (commercial as well as residential) have relatively high errors rates.

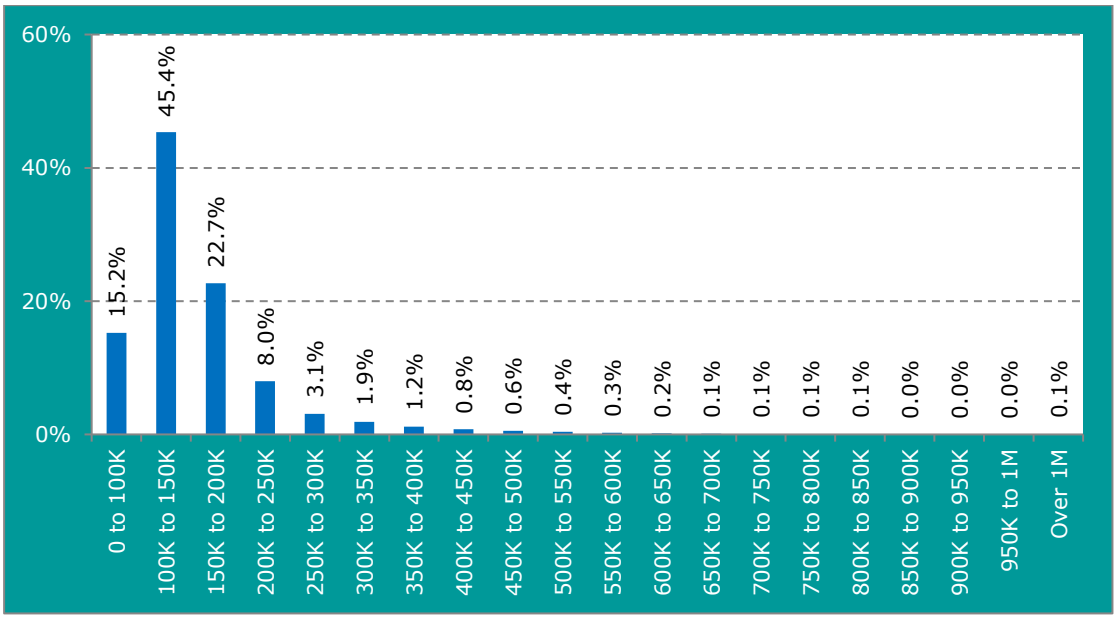
5 Application of the Model

5.1 Valuing All Properties

Using regression results, the hedonic model values the full dataset of properties in Geo-Directory based on their characteristics.[26] A value is predicted for each property and this is converted into an LPT value band.

As Figure 5 shows, a substantial share of cases are placed in the €100,000 to €150,000 band. For comparison, it is worth noting that in the Interdepartmental Group analysis (shown in Table 1) 34 per cent of properties are in the €100,000 to €150,000 band and 32 per cent in the €150,000 to €200,000 band.

Figure 5 – Number of Properties per Value Band



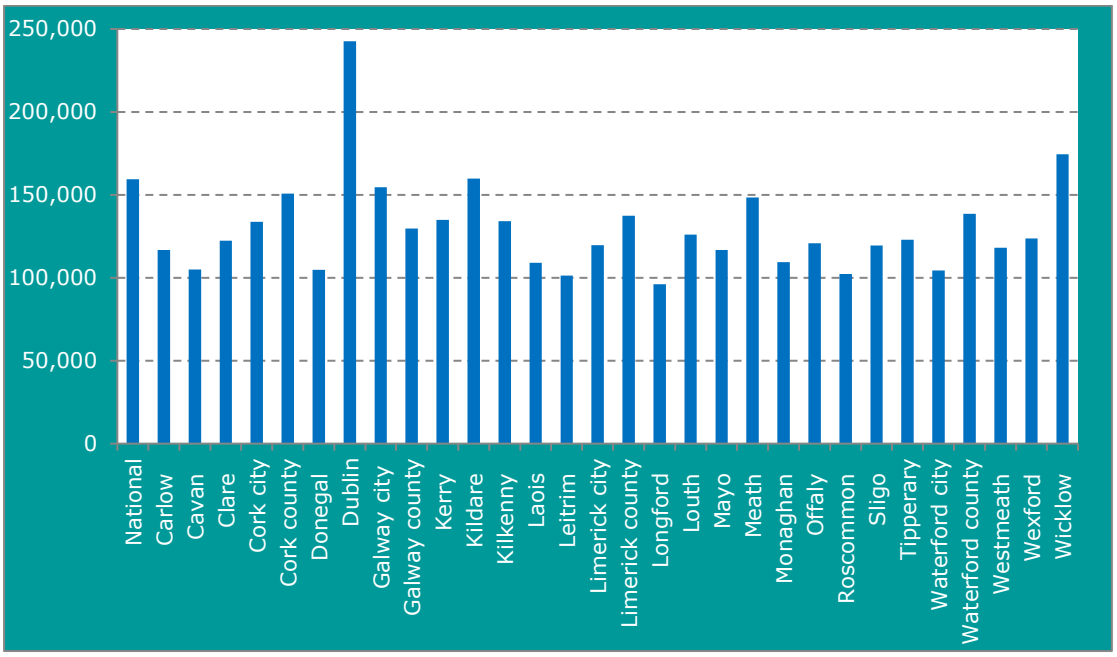
Source: Revenue analysis.

Figure 6 displays average values by location. The average national value is €161,000.[27] Comparisons to listing data from third party sources confirm that patterns in values across most counties are broadly consistent (when adjusting for differences in asking and sales prices).

[26] All properties marked as residential or mixed (residential and commercial) in Geo-Directory are valued. This covers 2.05 million properties. The CSO Census results indicate there are 1.99 million permanent dwellings or housing units in the State: 1.7 million occupied and the remainder vacant or holiday homes (Census 2011 Results *Profile 4: The Roof Over Our Heads – Housing in Ireland*).

[27] Estimates of average prices vary. In Q4 2012, average asking values from daft.ie and myhome.ie listings data were €170,700 and €200,800 respectively.

Figure 6 – Average (Mean) Values in Certain Locations



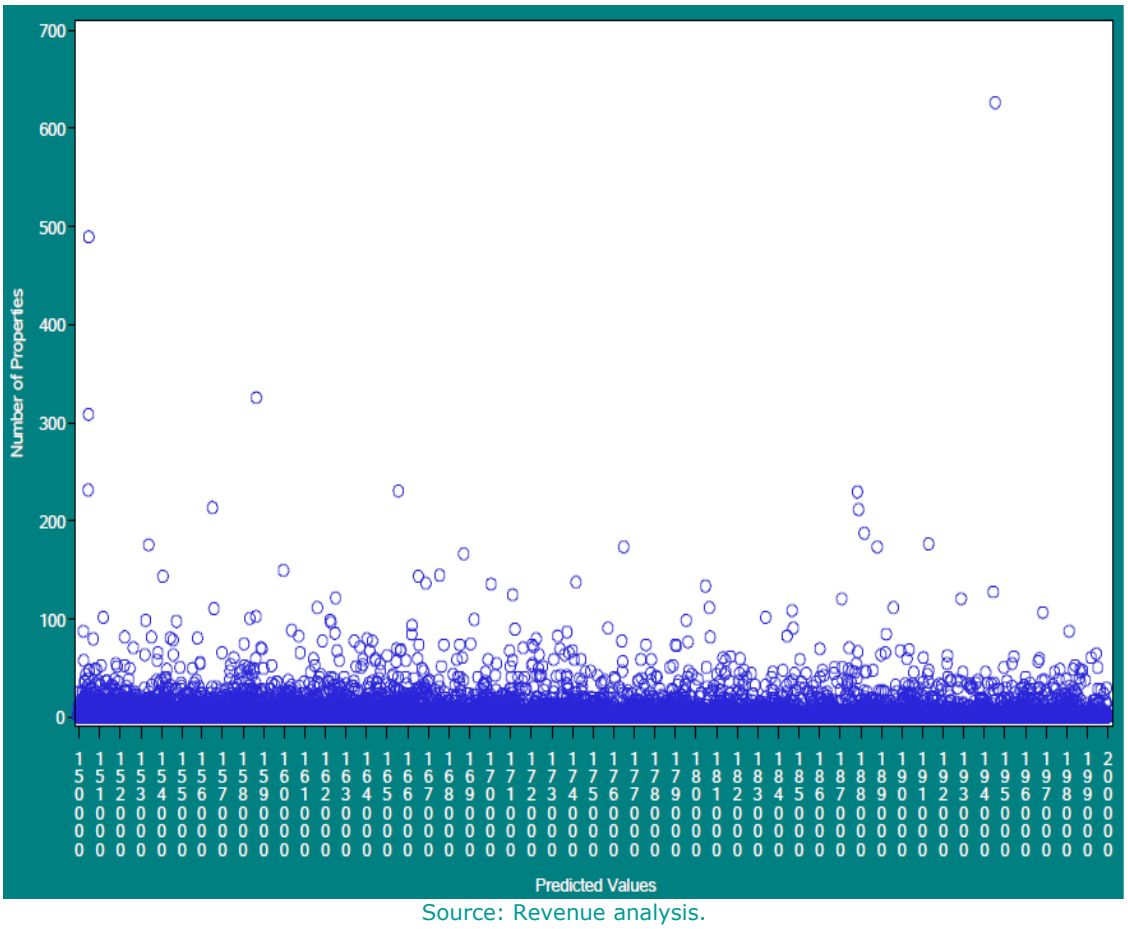
Source: Revenue analysis.

5.2 Spread of Properties within Bands

The spread of properties within bands is fairly even. From visual assessment, there does not appear to be any major clustering of valuations around band thresholds. For example, Figure 7 shows the distribution of properties within the third band (€150,001 and €200,000).

Within each band, the 10th and 90th percentiles are the levels below and above which the bottom and top 10 per cent of cases lie. With bands of €50,000, if evenly distributed, the percentiles should be at €5,000 and €45,000 respectively (e.g., in the second band the percentiles should be €105,000 and €145,000). Table 4 shows the percentiles for each band and confirms this is the case in most bands. Although there is variation, there is no significant clustering in any band. In general, 90th percentiles are slightly further from thresholds than the 10th, suggesting minor weighting towards the lower end of bands.

Figure 7 – Scatter Plot of Property Values (Valuation Band 3)



Source: Revenue analysis.

Table 4 – Valuation Band Percentiles

Value Band	10th Percentile	90th Percentile
100K to 150K	105,291	144,023
150K to 200K	153,089	191,597
200K to 250K	203,224	241,598
250K to 300K	253,838	293,681
300K to 350K	304,169	344,709
350K to 400K	353,808	394,471
400K to 450K	404,181	443,332
450K to 500K	454,587	494,419
500K to 550K	503,822	543,709
550K to 600K	554,504	592,641
600K to 650K	604,309	645,085
650K to 700K	653,419	694,066
700K to 750K	703,968	744,986
750K to 800K	752,735	793,745
800K to 850K	805,213	837,591
850K to 900K	855,072	895,993
900K to 950K	904,289	944,092
950K to 1M	953,800	993,307

Source: Revenue analysis. Note: Percentiles for the first and last valuation bands are not shown given the differences in structure of these.

5.3 Valuation Guidance

The model applied to the full population of properties in Geo-Directory, as described above, forms the basis for Revenue guidance provided to property owners in March 2013.

Where properties are not recorded in Geo-Directory or non-unique addresses make identification and valuation of the specific properties difficult, average values for the thoroughfare or another locality level are used.

As LPT is a self-assessed tax, it is the responsibility of each property owner to make a return to Revenue and indicate the value band in which their property is located. To assist property owners, Revenue guidance will be available from March 2013.

This guidance will be presented in the form of average valuation bands by Electoral District, based on the average of the model results to that level. Average values are distinguished by property type (detached, semi-detached, terrace, apartment/flat and bungalow), age (pre- or post 2000) and location.

Property owners will need to consider the specifics of their property, their knowledge of their local neighbourhood and any other factors influencing value to determine the appropriate valuation band for their own property. If a property is smaller or larger than the average for the area, is in a significantly poor state of repair or has exceptional or unique features, this will have to be factored into the assessment of the valuation band of the property.

6 Conclusion

This paper summarises the development of a property valuation model to assist with the implementation of LPT. The data sources, implementation, results, performance and subsequent use of the model to value all properties in the State are described.

At its simplest, the model values properties by applying the average value for that type of property in that area, with some adjustments for property specific or neighbourhood information, such as distances to amenities/services.

The model generates the guidance on Revenue's website, of average valuation bands within each Electoral District, to assist owners in assessing the value of their property.

The model performs within industry and international standards. There remains a group of outlier cases, around 10 per cent of properties in the training dataset where the model does not accurately predict values (valuations are more than one LPT band out). It is interesting to note that this percentage has remained stable through all recent iterations of the model. Despite considerable improvements over time, as new valuations and characteristics data sources became available, this percentage remained broadly the same.

One disadvantage of this model is the lack of property specific data on room number or floor size, usually among primary variables predictive of value. Given the persistence of the 10 per cent outlier figure, it may be that further data would not significantly improve accuracy. If that 10 per cent reflects a very heterogeneous block of properties, no model based on average relationships between values and characteristics may be able to fully capture their presence.

Any focus on potential shortcomings of the model should bear such points in mind. Property valuation on the scale of an entire State is not an exact science. Even with the best data, no model can be expected to accurately predict the values of all properties in Ireland. However, the model and the average valuations produced are good guides for the vast majority of properties, particularly in the context of the requirement of LPT for property owners only to assess the correct valuation band, not the precise value, of their property.