The Value of a Sea View: Hedonic Estimates Using 3D Simulation and Natural Language Processing

Dr. Tom Gillespie*  
Prof. Ronan C. Lyons†  
Prof. Stephen Hynes‡

Abstract

This paper develops a novel measure of sea-view breadth and depth, based on a combination of GIS 3D simulation and Natural Language Processing (NLP) techniques. NLP identifies properties that have a sea view, while GIS measures its extent. It then estimates the value of a sea view using a hedonic housing price model and a sample of over 100,000 sales listings in Ireland, 2014-2019. We find that a sea view can add significant value to a property, all else being equal, with an average premium of 8.1% rising to 15% for a wide breadth of sea view.

* Socio-Economic Marine Research Unit, Whitaker Institute, University of Galway.  
Thomas.gillespie@universityofgalway.ie  
† Department of Economics, Trinity College Dublin, and Spatial Economics Research Centre, LSE.  
Ronan.lyons@tcd.ie  
‡ Socio-Economic Marine Research Unit, Whitaker Institute, University of Galway.  
Stephen.hynes@universityofgalway.ie
1 Introduction
Proximity to and views of water – especially the sea – are associated with many positive measures of physical and mental wellbeing, from higher levels of vitamin D to better social relations (Mitchell and Popham, 2008; MacKerron, G. and Mourato, S., 2013; Gascon et al., 2015; Dempsey et al., 2018). However, so-called ‘blue space’ amenities are not fixed in supply, as they can be exposed to pressures such as pollution, rezoning or access restrictions. Thus, knowledge of the benefits from blue spaces needs to be integrated in spatial development and management policies (TEEB, 2011). The estimation of the benefit values associated with blue spaces has long been the focus of research including the estimation of the recreational use benefits of blue spaces (Borger et al., 2021), the health and wellness benefit values associated with proximity to blue spaces (White et al., 2013; Garrett et al., 2019; McDougall et al., 2020) and the aesthetic benefit values bestowed by blue spaces (Bond, Vicky, and Michael 2002; Bourassa, Hoesli, and Sun 2006; Yamagata et al., 2016).

This paper adds to the literature on the aesthetic benefit values associated with blue spaces by first developing a novel measure of sea views that combines GIS 3D simulation and Natural Language Processing (NLP) techniques. Viewshed estimation using GIS techniques are limited to the quality of the underlying Digital Elevation Model (DEM) used, and therefore, depending on the resolution, may overestimate the extent of a view by not taking into account trees, buildings, and other potential blockers of a view. Interacting GIS and NLP techniques for estimating sea views improves measurement error by negating observations which do not have any mention of sea views in the text of the advertisement. To the best of our knowledge, it is the first paper to combine both types of measurement for the purposes of quantifying sea views. While NLP methods are used to screen for dwellings with a sea view, GIS methods are used to estimate the breadth and depth of that view. Those GIS methods are designed to have a fixed
Computational cost, meaning that, in an era where large datasets are increasingly available, the method has a low computational intensity regardless of dataset size. This new approach is applied to a dataset of just over 100,000 Irish properties listed for sale between 2014 and 2019, using a hedonic housing price model to estimate the value of sea view breadth and depth. In so doing, it contributes to the limited literature that uses continuous measures of sea view (breadth and depth) to estimate the aesthetic value of the sea.

The methodological contribution addresses a common issue with measuring environmental aesthetics on large samples and large geographical scales. In their analysis of 328 peer-reviewed articles on environmental amenities 1974-2013, Schaeffer and Dissart (2018) note that it is regrettable that only a few papers attempt to measure environmental views “since perception of the local environment comes first and foremost through people's gaze”. Many of those that do, have significantly smaller sample sizes, meaning computational issues do not arise. Benson et al. (1997), Benson et al. (1998), and Benson, Hansen, and Schwartz (2000) use data on residential property sales in Washington State to estimate the aesthetic value of various natural amenities, in particular sea views. They find not only clear evidence of premiums for better views, ocean frontage and views closer to the coast but also that the value of views increased over the period 1984-1993, due to changes in the market conditions. Likewise, Bond, Vicky, and Michael (2002) also find a substantial premium (of almost 90%) for water views, measured as a categorical variable and again with a small sample size of 190 homes on Lake Erie. Plattner and Campbell (1978) and Bourassa, Hoesli, and Sun (2006) find that properties with water views appreciate at a higher rate than non-view properties.

GIS methods have also been used on larger datasets to examine the effect of blue-space amenities. One example is Samarasinghe and Sharp (2008), who study 2,243 transactions during 2004 in Auckland (New Zealand). In line with the earlier literature, they found a hierarchy of view amenities and a declining premium with distance; they also attempt to control
for spatial dependency by including distance-based variables. A second example is Wallner (2014), who develops four view measures for a database of 24,491 Sydney (Australia) home sales in 2008. Compared to a binary (0-1) view variable for the same sample, the continuous measures improve out-of-sample prediction accuracy, while the results confirm the substantial premium for good water views. Thirdly, Baranzini and Schaerer (2011) construct a continuous measure of view using LiDAR data for approximately 13,000 rental dwellings in Geneva (Switzerland). A final example is Yamagata et al. (2016), who generate a continuous measure of ocean views using LiDAR data for Yokohama (Japan) and employ a spatial multilevel additive regression model, to allow for non-linear price functions of sea view. Unlike previous, however, their results do not show conclusive evidence of a premium for sea views.

This paper has five main contributions. Firstly it uses a combination of NLP and GIS techniques to measure sea views. Secondly, it outlines a GIS method for viewshed generation that is of a fixed cost, after which any sample can be used to obtain a continuous measure of sea view. Previous studies generate continuous measures of sea views for each observation (Benson, Hansen, and Schwartz (2000), Bin et. al., (2008), Baranzini and Schaerer (2011), Wallner (2014), Yamagata et al. (2016)). Thirdly, due to the fixed cost nature of the viewshed generation technique, individual viewsheds do not need to be generated, hence the sample size relative to the existing literature (on estimating values for sea views using continuous GIS generated measures) is large (see appendix table A7). The relatively large sample is advantageous in that there is more within variation in the spatial fixed effects, allowing for the smaller geographical scales of spatial fixed effects to be used, reducing omitted variable bias. Fourthly, it adds to the existing literature on the estimated capitalisation of sea views into property values: 8.1% for an average sea view, and 15% for a wide sea view. This is consistent with the findings of Bourassa et. al., 2005; Yu et. al., 2005; Hui et. al., 2007; Jim & Chen, 2010; Wallner, 2014; and Osland et. al., 2020. However, it is less than estimated in Benson et.
al., 1998; Bond et. al., 2002; Samarasinghe and Sharp, 2008; Baranzini & Schaerer, 2011; Wyman et. al., 2014; Mothorpe & Wyman, 2017; and Filippova et. al., 2020. Finally, it contributes to the literature on the use of listed vs transacted prices in hedonic housing price model. List prices are found to be a satisfactory proxy for final transacted prices, similar to the findings of Semeraro and Fregonara (2013), with more discussion in section 2.3.

The rest of this paper is structured as follows. In the next section, the data used in the study is outlined, including the study area, and the listings dataset. Section 3 describes the natural language processing methods used, and the generation of sea view measures. In Section 4, the empirical approach is described and in Section 5, baseline results and robustness checks are presented. The final section concludes.
2 Data

2.1 Study area
The analysis focuses on the Republic of Ireland's jurisdiction which comprises the majority of the island of Ireland. Ireland is a small open economy and part of the European Union, with a population of 5 million; approximately 1.5 million live in the metropolitan area of its capital, Dublin. Ireland has the highest share of population living in houses in Europe at 92.5% compared to the EU average of 57.6% (Eurostat, 2016). Ireland’s coastline is approximately 1,448km long and roughly two in five of Ireland’s residents live within five kilometres of the coastline; using a European classification, over 90% of the island is considered a coastal zone (Tsakiridis et al., 2019). During the timespan covered here (2014-2019), the Irish economy underwent a period of strong economic recovery. This was preceded by a sharp crash (2007-2012) that followed its ‘Celtic Tiger’ period of growth, starting in the mid-1990s and culminating in a credit-driven property bubble in the mid-2000s. The sample period, 2014-2019, is one of roughly homogenous housing market conditions, typically characterised by strong demand, weak supply and rising prices.

2.2 Housing listings
The housing data used in this paper comes from a national dataset of real estate listings from the property website www.daft.ie, the leading real estate website in Ireland. The daft.ie listings, which cover the period 2014-2019, capture the national market in its entirety, with the company estimating its coverage to be over 95% of all listings in the Irish market. Key property-level attributes for inclusion in a hedonic housing price model include the property’s type, size and location. Property type includes two categories, each with sub-categories: houses (terraced, semi-detached, detached and bungalow) and apartments (regular and duplex). Size is typically measured in Ireland by number of bedrooms (and bathrooms), rather than internal floor area, with universal coverage in listings for bedrooms/bathrooms and approximately 78% coverage for floor area in square metres.
Dwelling location is captured by XY coordinates (according to the Irish National Grid) and latitude and longitude. These locations are automatically generated in the daft.ie system using either a dwelling-specific Eircode identifier or where that is unavailable its address, which may be entered with error. In the sample, only listings that are within Ireland are included, using official definitions of coast, the political border with Northern Ireland and district electoral division (ED) outlines. Given the nature of the study, only listings where the building level accuracy is assured were included, either from the address or the Eircode. Based on a dwelling’s location, it is assigned both to one of almost 3,500 official EDs and to one of 312 micro-markets, which are real estate market segments, manually constructed based on a location’s geographic, demographic and economic factors.

In addition to type, size and location, as discussed above, three other variables are captured for all listings: the listed price, discussed in more detail below; the initial date of listing; and a dwelling’s Building Energy Rating (BER) energy performance certificate, although in certain instances this may be missing. Further variables are generated based on the text of the ad, including acres of land, where included, and other property-specific attributes that may affect its value, such as “double glazed”, “jacuzzi”, “garden” and “views”. The process for capturing views is described in more detail in Section 3 below.

The 2014-2019 dataset in full contains 375,190 listings but is subject to four filters to generate the final dataset used in the analysis. Firstly, properties with unusually high (above €2 million) or low (below €30,000) prices are excluded. Then, any properties that do not include bedrooms or bathrooms are excluded, as are those more than five kilometres from the coastline. Lastly, as described above, only those properties mapped to the exact building are included, giving a final dataset of just over 100,000 dwellings (See Table 1 for breakdown).

[[Insert Table 1 here]]
2.3 Housing transactions
By its nature, the listings dataset does not contain the ultimate transaction price. Nonetheless, the use of listed prices is well established in the hedonic literature, where transaction prices are unavailable; see, for example, Hyland, Lyons, and Lyons (2013) and Malpezzi (2003). While Faller, Helbach, and Vater (2009) find an 8% premium on average for transacted prices above listed prices in North Rhine-Westphalia, Germany, those differences were not explained by observed housing characteristics. More generally, systematic mispricing of housing characteristics has been found to be very costly to the seller (Knight, 2002; Merlo and Ortalo-Magne, 2004), in line with theoretical models of seller behaviour (e.g. Knight, Sirmans, and Turnbull 1994). It increases time on the market and reduces the final transaction price. Both effects make it more likely that measurement errors (differences between listing and transaction prices) are unrelated to housing characteristics (Semeraro and Fregonara, 2013).

Secondly, and related, a body of research finds that listed and transaction prices have the same trend, particularly once hedonic methods are used, including one exploring the relationship between list and transaction prices in Ireland during the period 2006-2012, a period of greater volatility (Lyons, 2019). In an Irish context, the lag between the date of listing and the registration of a transaction price may be in part explained by a delay between the agreement to sell and the registration, due to mortgage approval and legal procedures. As such, the actual date of transaction can lag behind the market, whereas a listed price reflects the market value at that moment in time. Listed prices in Ireland are based on estate agent assessment, rather than an owner’s valuation; estate agents have local market knowledge, including the price of properties recently transacted in the area.

For this analysis, we also matched a sub-sample of almost 50,000 listings with Ireland’s Property Price Register (PPR), a database that records addresses, dates and transaction prices, but no housing characteristics. This enables a direct check of whether the coefficients on the
variables of interest change, when the price metric used is changed; this analysis is undertaken in Section 5.2. The transactions sample is non-random compared to the overall set of listings. In particular, the transactions subsample is somewhat biased spatially to areas where an address match between the daft listing and the property price register can be made, and these listings are more likely to be located in urban areas. Nonetheless the sample is useful to make a comparison between having list prices vs final transacted prices as an outcome variable.
3 Measuring sea views

3.1 Natural Language Processing

The process for calculating continuous measures of sea-view breadth and depth has two main components. The first is categorically identifying those properties with a view. To do this, Natural Language Processing (NLP) is used and the text of daft.ie advertisements. NLP is a field of artificial intelligence that enables computers to analyse and manipulate human language. It has a variety of potential uses, including predictive text, grammar checkers and spam filters, given its ability to summarise long documents, identify key themes or phrases, calculate sentiment and spot trends in a time series, among other things.

NLP is implemented using a combination of Python’s Natural Language Toolkit (NLTK) and the python tool Ruby. Specifically, the Princeton WordNet database, a part of the NLTK library, was utilised to find synonyms for sea, view and related words. The resulting list was augmented by online thesauruses, before a final list of synonyms relevant to the Irish market was chosen. For example, synonyms such as vista for view are not commonly used in Ireland – vista appears fewer than 1 in 200 times as frequently as view.

Through this process of identifying synonyms for both sea and view and calculating their relative frequency, a set of patterns was identified that captures the majority of ads that indicated a property having a sea view. At this point, Ruby was used to scan through the lower-case text of listings looking for three sets of patterns:

1. A listing contains a single phrase, such as: ‘across the sea’; ‘across the Irish sea’; ‘has sea view’; ‘on the sea’; ‘right on the seafront’; ‘see the sea’; ‘views of the sea’; ‘with sea view’.

2. A listing contains pairs of words within 200 characters, such as: bay and either overlook or view. Listings with either ‘overlook’ or ‘view’ and one of 17 terms are considered to have a view under this pattern.
3. Lastly, a listing contains two words but not a third, which would indicate a negation of the view. For example, a listing that contains ‘view’ and ‘beach’ but also ‘drive’ may not have a view of the beach as a drive may be required to enjoy the view.

The three patterns were run on the final dataset. They were also run for listings drawn from non-coastal counties, which are excluded from the dataset, to identify the prevalence of false positives, for example caused by bay windows and island units in kitchens; any such patterns were then corrected in the final dataset. Nevertheless, it is possible that there may be some false negative cases where a sea view is so obvious from the listing’s photos that there is no mention of it in the text of the ad. However, this error is expected to be low as the advertisements are quite descriptive: the median number of words per ad in the sample is 291 (between 69 and 723 when taking the range between the 5th and 95th percentile). Conversely, there may be some false positives in the case where the text of the ad refers to a sea view from some point in the property, the end of the garden for instance, or at a nearby local green space, etc.

3.2 GIS measurement of sea views

Topographical data are used to generate measures of view breadth and depth. For Ireland, 10-meter resolution vector data in the form of contour lines is the highest resolution available nationwide, sourced from the OSi under the national mapping agreement. These contour lines are converted to raster format, necessary for viewshed analysis in ArcPRO. The topo-to-raster tool was then employed for the conversion of contour lines to a raster Digital Elevation Model (DEM). The tool allows the combination of other physical geographical features represented by vector data such as lakes, streams, and coastline in order to more accurately represent the topography close to these features that may not be well represented by the contour lines. This operation had to be done in tiled units as processing time increased exponentially for areas larger than approximately 8,000km2. This split Ireland into 20 different overlapping sections and the tool was run individually on each section. All sections were joined into one DEM using
the mosaic tool at the end. Raster pixels beyond the coastline were automatically assigned no 
data. To allow for viewshed projections into the sea, the DEM was extended 30km out from 
the coastline at a default elevation of zero. This was done using the raster calculator tool. The 
final output from this process was a 10-meter resolution DEM raster of Ireland with a 30 km 
buffer out to sea with a zero elevation. Forestry polygons from OSi data were converted to 10m 
sized rasters and added as extra elevation to the DEM to account for view-blocking forestry. 
This model of Ireland assumes a smooth landscape without buildings, individual trees and other 
potential objects which may block a property's view.

*Viewshed generation*

To create a measure for sea views the initial assumption that is required is that the area of sea 
that can be seen acts as a normal good, that is more is preferred to less. In reality the 
composition of the sea view, and not just the visible amount, might be what drives willingness 
to pay. Sea areas in combination with mountainous areas might be preferred to distant horizon 
sea views out from straight coastlines. Nonetheless the goal of the measurement was to find 
out what area of sea could be seen from each observation in the housing dataset.

Viewshed analysis is a 3D simulation methodology used in GIS software. If topographical 
surface data for a landscape is available, then, given a co-ordinate, a viewshed output can by 
generated by the software, representing, for that coordinate, the areas which can be seen and 
the areas that cannot (see Appendix, Figure A1). Due to the computational challenge of 
generating viewsheds for every observation in the dataset, the process was reversed; the sea 
area's viewshed (what areas of the land could be seen from the sea) was mapped onto the 
landscape. This was done by filling the sea area with evenly distributed points, as viewsheds 
can only be computed for points and lines, not areas/polygons (see Appendix, Figure A3).

Sea buffers extended around the entire coastline of Ireland at a distance of 500m from the shore. 
The buffer was drawn without the small islands that didn't include dwelling observations to
avoid points being far out to sea, the reasoning for this will be made clearer in the following section. The buffer stretched out 500 metres from the shoreline and contained evenly distributed points (inner points) spaced 250m apart.

The visibility tool was used in ArcPRO to generate viewsheds. A maximum of 1000 points could be used in any process and the output would be a raster representing for each pixel the aggregate number of sea points that could be seen (see Appendix Figure A3). Observer height was set to 1.8m, representing the average height of a person standing up, seapoint height was set to 0m as the sea points were on the surface of the water. A default refractive coefficient was also incorporated to represent the bending of light over long flat surfaces due to air temperature differentials.

A maximum visible distance parameter required a decision on how far away a sea point could effectively be seen. Setting this parameter to very large distances increased computational time exponentially after a certain point due to the greater number of raster pixels that needed to be taken into consideration in calculations. The decision was made that 5km inland would be a reasonable cut-off distance for having a 'meaningful' sea view. Therefore, the maximum visible distance was the distance from the land to the furthest edge of the buffer, + 5km, so the inner buffer sea points had a maximum visible distance of 5,500m. The DEM raster for Ireland was clipped relative to the sea points that were being processed and the maximum visible distance that was set, as processing time also increased exponentially the larger the geographical extent of the DEM raster. Separate viewsheds were run for transitional water bodies as well as coastal water bodies as a robustness check. This was because it is difficult to distinguish some transitional water bodies as being either sea or river.
Once each collection of viewsheds (inner coastal, and inner transitional,) were complete they were merged together using the mosaic tool for the entire coastline of Ireland and the values were extracted to the set of housing coordinates.

The final viewshed rasters represented, for each 10m pixel of land in the country, how many sea points (from each category, inner coastal and inner transitional) could be seen. The limitation to this however is that the underlying DEM did not take into account buildings and trees and other potential view obstructions. As a result, it is probably over-estimating the number of points which can be seen in reality. To make these scores more robust, only listings that indicate a sea view from the NLP process are given a sea point score, if NLP is zero, the sea point score will be zero. The assumption here is that any house with a sea view is going to mention it in the text of the ad as it is a unique selling point. There may be cases where the view is partially blocked by buildings or trees and therefore the sea point score would not be a true to reality. Furthermore, there could be listings in which the sea view is so obvious from the location and pictures that it is not mentioned in the text of the advertisement. Therefore, the number of sea points visible from a property should be considered more correctly as a measurement of view potential rather than actual view.

Developing a measure of sea view
A number of previous studies developed non-categorical measures of sea view, including Benson et al. (1998), Samarasinghe and Sharp (2008), Hamilton and Morgan (2010), Barazini and Schaerer (2011), Yasumoto et al. (2011), Wallner (2013) and Yamagata et al. (2016). To measure the area of water visible from a property in these studies, the viewshed for that property was generated and the scope and extent of the sea-view measured. This was done for each individual property and was possible given the small datasets used in these studies. Both the absolute size of the dataset used in the current study and its large geographical scope preclude the use of these techniques.
As discussed in the previous section the number of sea points visible were generated for each 10m parcel of land. The buffer is shown in the left-hand panel of Appendix Figure A3, while the right-hand panel shows the resulting view ‘scores’ for land based on the inner buffer points. As such the ‘view breadth’ measure variable is based on the log of the number of inner buffer points. As this novel methodology generates a score for every parcel of land, it can be used on a dataset of any size, for a fixed computational cost. With the rise of administrative datasets, this should be useful in other settings.

This process also minimises the steps needed to determine the area of sea that can be seen from a property using GIS software. Conversely, if individual viewsheds are computed for each property, there would be four additional steps needed to achieve the same result: simulation of the viewshed, conversion of the viewshed raster to polygon, the intersection of the viewshed polygon with an ocean polygon, and the calculation of the area of sea-view ocean polygon. The computational challenge for computing individual property viewsheds in this case is due to the fact that it is taking a national sample over a large (in the scale of a 10m resolution DEM) geographical area (area of Ireland approximately 84,421 km²). Larger areas can increase viewshed computation time exponentially and therefore a segmentation strategy (splitting the DEM up into smaller areas) is required. This is easier to do with evenly distributed points in the sea than randomly distributed points on land, as sea points have the same elevation and therefore a visible distance to the horizon doesn’t need to be calculated.

Nevertheless, in order to compare both methods, individual property viewsheds were generated for the entire sample. This required a raster DEM segmentation strategy that was based on a property’s visible distance to the horizon (derived from a property’s elevation and used to set the maximum viewshed radius), and clustering observations by area, to minimise the total raster area required to calculate viewsheds. In total 54 DEM raster sections around the coast of Ireland were created. In order to generate the viewsheds for individual properties, and convert to
polygons, computation time took approximately 5 weeks (using ArcGIS Pro on a virtual
machine with 6 virtual processors @2.6GHz clock speed, and 24GB of RAM), with 2 weeks
of post processing of the resulting polygons. By comparison, the original reverse approach took
2 weeks of computer processing time.

The methodology captures, in particular, the (horizontal) breadth of the sea view but may not
capture as accurately the (vertical) depth of the view. This is due to the fact that a large sea
surface area from a bird's eye perspective will not always have the same visual impact when
perspectives are varied by elevation, and therefore the scoring system will not always gauge
the vertical visual impact of the sea. In order to capture view depth, the vertical angle from the
closest coast out to the horizon was used. This will be highest for perspectives that are at a high
elevation and also close to the coast. Euclidean distances to coastal water bodies and
transitional water bodies were calculated to be included in this calculation. The variable ‘view
depth’ is derived to capture the vertical depth of view. Figure A4 in Appendix gives an
illustration of the view depth variable measured in degrees.iv

The limitations of countrywide DSM mean that it is likely to overstate the true number of sea
points that can be seen for many dwellings.v Therefore, in the baseline, only listings which
indicate in the text of the advertisement that a sea view is present are classed as having a view
breadth (either close or distant). If a sea view is not indicated in the text of the listing, its score
will be zero.

The two measures of sea view included for each dwelling, view_breadth (the log of the inner
views score) and view_depth (the horizon angle) are calculated for each listing, \( i \), as follows:

\[
view_{breadth_i} = \ln \ln (inner_{score_i}) * \text{'views'}_i
\]

Where: \( view_{breadth_i} = 0 \) if \( inner_{score_i} = 0 \)
\[
\text{view}_{\text{depth}}_i = \left( \frac{\text{eye level}_i}{\text{distance to coast}_i + 250} \right) \times 57.2958
\]

Where: \(\text{eye level}_i = (\text{elevation}_i + 1.8), \text{view}_{\text{depth}}_i = 0 \text{ if 'views' }_i = 0 \mid \text{innscore }_i = 0\)

4. Model Specifications

4.1 Hedonic Specification

Conceptually, the value of a dwelling takes the following form:

\[
\text{Price} = f(S, L, E) + \varepsilon,
\]

where the natural log of the dwelling’s price is a function of the dwelling’s structural characteristics (\(S\); such as number of bedrooms, bathrooms, or the presence of a garden), its location characteristics (\(L\); such as proximity to CBD, access to transport networks, socio-economic factors) and its environmental characteristics (\(E\); such as proximity to the coast or sea views). The error term, \(\varepsilon\), reflects the gap between the predicted and actual (listed) prices. The dwelling price is thus a function of all attributes relating to the dwelling and the resulting coefficients are the implicit marginal prices of the attributes.

More specifically, this analysis uses ordinary least squares and a semi-log or log-log specification (depending on the variable), as is typical in this type of study. Allowing for the long duration of the sample, and the focus on coastal amenities, the baseline specification is as follows:

\[
\log(\text{price}_i) = \beta_0 + X'_{i1} \beta_1 + X'_{i2} \beta_2 + X'_{i3} \beta_3 + X'_{i4} \beta_4 + X'_{i5} \beta_5 + \varepsilon_i \tag{2}
\]

Where: \(\text{price}_i\) refers to listed or transacted sale price, depending on the specification; \(X'_{i1}\) to a vector of property-specific attributes; \(X'_{i2}\) to the time period (quarterly fixed effects); \(X'_{i3}\) to spatial fixed effect (SFE); \(X'_{i4}\) to a vector of location-specific control amenities (including distance to schools, transport networks, golf courses, etc.\(^{vi}\)); and \(X'_{i5}\) represents the regressors of interest, a vector of variables capturing coast-specific amenities. To account for possible
heteroscedasticity, robust and clustered standard errors (at the level of the SFE unit) are included and are used when calculating statistical significance.

The regressors of interest include the two sea view variables, reflecting aesthetic benefits: view breadth and depth, all as defined in the measurement section. In addition, there are two blue space variables, distance to coast and distance to transitional water bodies. In each case, the log of the distance is used in all specifications.

4.2. Location-specific Controls
Spatially-based relations between properties and/or areas require consideration when the hedonic pricing model is used to explain property values. Omitted variable bias (OVB) is a pervasive issue in hedonic house price modelling and there are many methods used in the literature to try and minimise its influence on identification (see Anselin, Lozano-Gracia, and Lall 2010). There are many unobservable spatial processes that can influence the price function of housing and the onus is on the researcher to outline a strategy for minimising this issue in a hedonic study. The literature is generally split into two principal methodologies for minimising the effect of unobserved spatial processes: spatial econometrics approaches involving the use of spatial weights matrices (such as the spatial error model of the spatial lag model); and the SFE approach which uses spatially delineated geographical entities. Kuminoff, Parmeter, and Pope (2010) show that the SFE strategy can mitigate omitted-variable bias and improve the accuracy of estimates for marginal willingness to pay in both linear and non-linear specifications of the hedonic price function. There is however a trade off in the use of SFE: using geographically smaller units of SFE minimise omitted variable bias at the expense of variation, therefore larger datasets are more suited to the SFE approach. Spatial econometric approaches are less suited to larger datasets as computing spatial weights matrices can quickly become computationally intense with larger numbers of observations and in such cases random within-sampling is used at the loss of the full potential of available variation.
Von Graevenitz and Panduro (2015) argue that the spatial econometric models such as the spatial lag model and the spatial error model are not suited to hedonic analysis. The former implies spillovers between prices and therefore a spatial multiplier in the marginal prices that households pay for an attribute (LeSage and Pace, 2011). Only the interpretation of the spillover as a purely informational effect is consistent with the interpretation of the hedonic function describing a market in equilibrium. The spatial error model assumes that omitted spatial processes causing correlated residuals are uncorrelated with the regressors included in the model. This is unlikely to hold if the regressors include spatially varying characteristics. Because location varies only in two dimensions, spatial variables tend to be correlated with each other (Von Graevenitz and Panduro, 2015). Nevertheless, there is a debate in the literature over the best strategies for reducing OVB and spatial dependence, with Anselin (2013a; 2013b) being one of the main protagonists for the spatial econometrics approach. The large data available for this research, and the spatial nature of the variables of interest, such as sea views, suits the fixed effects approach.

Bishop et al. (2020) outline the best practices when it comes to the implementation of a hedonic house price study for estimating WTP for environmental quality. This paper uses a robust identification strategy for minimising OVB which is in-line with their findings and recommendations. In this paper the approach used to minimise the influence of omitted spatial processes is the use of SFE. These are employed for two reasons. The first is that the sample size is relatively large and therefore is more suited to the use of SFE as outlined above. The second reason is that there are a number of geographical scales of SFE available to choose from in the data and hence it is easier to find an appropriately sized geographical unit which minimises omitted variable bias while maintaining a statistically acceptable amount of within variation. Having a range of SFE to choose from also allows for spatial sensitivity analysis to check for OVB. As a robustness check and as recommended by Von Graevenitz and Panduro
(2015), the estimation of the primary model using all four available levels of SFE, is also reported. Observing the sensitivity of the estimated coefficient on the variable of interest to changes in the level of spatial fixed effect can give an indication as to whether there is some omitted spatial process influencing the variables of interest. Table 2 illustrates the within variation per spatial fixed effect used in the analysis.

[[Insert Table 2 here]]

Remaining spatial correlation in the error term can be accounted for by clustering errors at the level of SFE used to avoid overestimating significance levels. Robust standard errors improve on regular standard errors because the resulting inferences are asymptotically valid when the regression residuals are heteroskedastic, as they almost certainly are when regression approximates a nonlinear conditional expectation function (Angrist and Pitschke, 2008). Housing prices are unlikely to be independent across observations. Prices in the same area tend to be correlated due to the similar levels of amenities and characteristics. If areas can be reasonably identified whereby it is assumed that there is no within explanatory variation due to clustering, then clustering standard errors at the level of spatial fixed effect can correct for serial correlation as long as the number of clusters is sufficiently large. The exact number is difficult to quantify. Angrist and Pitschke (2008) report that studies that cluster at state-level in America have sufficient variation to apply clustering with confidence. In most of the preferred specifications in this paper, the number of clusters exceeds 50 (number of American states), and therefore clustering the error terms should correct any serial spatial correlation in the housing data.

4.3 Causal Identification

In addition to the detailed spatial granularity of the dataset, allowing for controls for locations-specific unobserved variables, identification of causal effects for view breadth and depth
amenities is supported by the topography of Ireland. The island’s geography (an indeed any hilly/mountainous landscape) means that there is variation of the view measures, even at similar distances from the coast\textsuperscript{vii}. This is shown in Figure 1 where summary statistics for 24,166 properties with at least one (inner) point visible and less than 1km from the coast are presented. Specifically, it shows the mean, median and interquartile range for view breadth, by distance from the coast, using the measure of (inner) points as the vertical axis. For properties within 100 metres of the coastline and with a view, the typical number of visible points is 71. For properties between 700m and 800m, the median is 68. These summary statistics suggest that sufficient variation exists to separately measure view-based and distance-based coastal amenities.

[[Insert Figure 1 here]]
5. Results

5.1 Baseline

The results displayed in Table 3 show the six different specifications employed. The first four specifications are descending in order of spatial fixed effect – from most (local market) to least (small area) aggregated. The fifth specification interacts electoral districts by year of listing, thus representing a spatio-temporal fixed effect. The sixth specification uses small area fixed effects and a micro-market by year spatio-temporal dimension.

[[Insert Table 3 here]]

Based on the discussion in Section 4, the preferred specification is with time-invariant spatial fixed effects at the level of the Small Area and spatio-temporal SFEs at the level of micro-market by year. Effectively, for two properties within the same Small Area (in urban areas, typically 200m-300m), the same observable characteristics, and similar access to other location-based amenities, coefficients show the effect of additional view breadth/depth. This sixth specification has the best fit in Table 3, as demonstrated by the lowest root mean square error (RMSE) (0.222) across the six different specifications.

In all seven specifications, the relevant coefficient suggests that the effect of sea view breadth on housing prices is positive and statistically significant. For example, an increase of one on the log scale for the breadth of a sea view, which is equivalent to increasing the area of sea within 500 metres of the coastline that can be seen by 24 square kilometres, is estimated to increase prices by 1.1% in the preferred specification. The depth of a sea-view also has an impact on housing prices. For the preferred specification, a one-degree increase in the angle of sea-view depth is associated with a 1.3% higher price.

Figure 2 summarizes the results from Table 3 graphically. In line with the recommendation of Von Graevenitz and Panduro (2015), it displays the coefficients, with confidence intervals informed by robust standard errors, associated with the two variables of interest: sea view...
breadth and sea view depth. Market (MKT) represents the least granular spatial control, while small area (SA) with micromarket-by-year represents the most granular, as discussed above. The estimated effect of sea view breadth, in particular, is remarkably stable across empirical specification. For sea view depth, the more precise the spatial control, the smaller the estimated effect, suggesting some correlation between otherwise unobserved location-based factors and view depth.

[[Insert Figure 2 here]]

Important controls in the estimation of the value of a sea view are distance to coastline and transitional water bodies.\textsuperscript{viii} In Table 3, across all specifications, the distance coefficients are negative, meaning dwellings closer to the sea are more valuable. As the level of SFE becomes more granular, the statistical significance of the coefficients increases, as well as the magnitude of the effect on distance to coastline. A change in one point of the log scale is the equivalent of comparing a dwelling 400m from the sea/transitional water body to dwelling 1.1km away. Therefore, in the preferred specification (the right most column in table 3), for a property 400m away from the sea, compared to one 1.1km away, the sale price will be 5.9% higher.

As distance to coastline and extent of sea view are likely to be correlated, Specification R2 in Appendix Table A1 presents the results of the coefficients of interest when distances to coast and to transitional water body are removed. While the results for sea view depth are largely unaffected, the coefficient on sea view breadth increases in magnitude by roughly half (from 0.011 to 0.015), with an increase in statistical significance also. This suggests that, while multicollinearity is not affecting the precision of results, there is sufficient correlation between distance and view breadth for view breadth to act a partial proxy for distance to coast when it is not included in the specification.
5.2 Robustness checks
In this section, the robustness of the results are tested in two ways. Firstly, the two components of the view measure – based on 3D viewsheds and NLP – are examined separately and the results compared to the baseline. Secondly, a range of empirical robustness tests are performed, including using transaction prices rather than listed prices, for a subsample for which these are available.

Removing GIS and NLP interaction
Table 4 presents the results of the baseline alongside results from specification that separate out the GIS and NLP measures of sea views. These support the empirical strategy adopted as, in particular for view breadth, this separation affects the results. For view breadth, without the restriction to ads that mention views, the coefficient on view breadth is smaller in magnitude and (when view depth is also included, as per Specification (1) in Table 4) it loses statistical significance. However, if sea view depth is omitted from this regression, sea view breadth remains positive and statistically significant (Specification 2).

[[Insert Table 4 here]]

Specifications 4 and 5 show the categorical impact of views being mentioned in the text of the ad, as determined by NLP, without any control for view breadth or depth. Specification 5 includes a version of the sea view NLP dummy where any listings are scored 0 if the NLP measure is zero but also if the GIS measure is 0 too. Therefore, the coefficient on sea view NLP, in specification 5, is the best estimate (8.2%) for the premium for an average sea view, ceteris paribus.

In an extended specification where the sea view breadth measure is broken into quartiles, the highest quartile of sea view breadth is estimated to command a 15% price premium (See appendix, Table A2.)
Further robustness tests
Appendix Table A1 displays the results of four additional empirical specifications (R1-R4), as robustness tests, presented alongside the preferred specification in the first column. Specification R1 investigates non-linearities, by including a squared term for both measures of sea view. While in both cases, the linear and the squared terms are statistically significant, the model fit does not improve. In line with the desire for specification parsimony, the specification without squared terms is kept as the preferred one. Specification R2 omits the distance variables as discussed in section 5.1.

The final two specifications, R3 and R4, test the effect of changing the outcome variable to the final transacted price as opposed to the listed sale price. This is done on the sample of almost 50,000 listings for which the ultimate transaction price was discoverable on the official Property Price Register, a sample that is non-random compared to the overall set of listings. In particular, the transactions subsample is somewhat biased spatially to areas where an address match between the daft listing and the property price register can be made\(^\text{ix}\), and these listings are more likely to be located in urban areas. R4 confirms this: it shows the results of the preferred specification, with the listed price as the outcome of interest, applied only to listings with a matched transaction price. The RMSE falls substantially, from 0.222 to 0.157, with a related sizeable increase in adjusted R-squared from 89% to 94%. Properties with ‘clean’ addresses that could be matched across listings and transactions datasets had, on average, substantially smaller residuals. While view depth results are largely unaffected by the change of sample, the impact of view breadth more than halves from 1.1% to 0.4%. The transition to the smaller sample throws away properties with harder-to-match addresses, in so doing removing important identifying variation in the variables of interest.

Specification R3 uses the same sample, of almost 50,000 dwellings, and empirical specification, but uses the transacted price as the outcome of interest, rather than the listed
price. Overall, model fit is almost identical across the two outcomes of interest: an RMSE 0.155 for transacted prices, compared to 0.157 in the listed sample. For the variables of interest, the magnitude of the coefficient is greater for sea view breadth in the transacted sample (0.7% compared to 0.4% for listed prices), but lower for sea view depth (0.6%, compared to 1.1%, and of marginal statistical significance, with a $t$-ratio of 1.6).

The switch to a sample of transactions, rather than listings, in this setting provides strong support for the assertion that the use of listings is valid: model fit and the general pattern of results across both datasets are very similar. However, in this setting, with a price register that does not include a dwelling-specific identifier, the use of transactions rather than listings is problematic, as important variation in the variables of interest – sea view breadth and depth – is lost.

It is therefore likely that, with respect to estimating the value of a sea view, the listed sale price is a robust proxy for the final transacted price. The differences are small enough not to warrant concern that they are systematically biased. A true like for like comparison in the two samples is difficult due to the difference in time signatures attached to both types of prices. The time signature for the listed price is the date it was listed on daft.ie, whereas the time signature for the final transacted price is the date that the sale was registered on the Property Price Register, which usually will be several months or longer after the initial list date. The resulting difference in time signatures means that the quarterly time dummy variable will be slightly different for both specifications, despite the fact that the sample is the same.

**Controlling for total viewsed**

One of the limitations of the reverse method of viewsed estimation (mapping the view from the sea on to the land), is that the total viewsed of the property cannot be controlled for. This might give rise to an identification issue whereby there may be a correlation between properties with large views and properties with sea views. Therefore, controlling for the scale of total
viewshed is desirable when estimating the effect of sea views on house prices as the comparison would thus become one between qualitatively similar properties situated similarly in space with the same size viewshed.

In order to test this limitation, total viewsheds were generated for each property in the dataset. This was then added as a control to the main specification. Table 5 displays the results whereby the log of the total area of land that can be seen in square kilometres was added as a control. Encouragingly, when the size of the total viewshed is controlled for, there is little to no difference in the estimated coefficients for both view breadth and view depth: The estimated coefficient for view breadth is 0.011 in the base specification and 0.011 after controlling for total area viewshed, for view depth, the coefficient changes from 0.013 in the baseline specification to 0.012 after controlling for total area viewshed.

[[Insert Table 5 here]]

6. Discussion and Conclusions
The valuation of non-market amenities, such as blue space has implications for policymakers in areas such as environmental protection and urban planning. In this study, the effect of sea view breadth and depth on housing prices in Ireland was estimated, over the period 2014-2019. This is done by combining textual analysis with a method of calculating viewsheds, at fixed computational cost and thus suitable for use on large datasets. Sea views are found to have a significant positive value in the Irish housing market and these effects are large enough to be economically important. Increasing the area of near shore sea sea that can be seen by 24 square kilometres is estimated to increase sale prices by 1.1%, while a one-degree increase in the angle of sea-view depth is associated with a 1.3% higher price. The average effect of a sea view on the listed sale price of a dwelling is 8.2% ceteris paribus, and that premium increases to 15% for a wide breadth of sea view.
We believe the results show causal effects, exploiting meaningful variation in sea view extent, even when controls for distance to coast are included, as in Figure 1. This interpretation is supported by a detailed set of spatio-temporal controls, including over 5,500 time-invariant local intercepts as well as almost 400 time-varying trends by micro-market, as well as a number of additional empirical specifications that speak to the robustness of the results. The results, which use listed price as the outcome of interest across a national sample, are supported by the results of an analysis of a non-random sample of 50,000 more urban listings, for which transaction price is available.

The primary methodological contribution in this paper relates to the development of a novel measures of sea views through the combination of GIS and NLP techniques. Reverse engineering the viewshed process allows for a fixed computational cost, with resulting viewshed data that can be applied to any dataset. Such a method is likely to have benefits as increasing numbers of large datasets become available to researchers. Combining this GIS measure with NLP techniques dramatically reduces measurement error compared to the GIS methodology alone, arising from the available detail of topographical data. The sea view depth measure is also unique to the literature to the best of the researcher’s knowledge.

In relation to sea views, policy makers may choose to draw on elements of these estimated values as a contribution to the costs of maintaining coastal amenities, for example by applying property taxes. Methods such as those used in this study can help inform local infrastructure cost–benefit trade-offs. For example, a practical use for the sea view depth measure developed here could be applied when assessing whether to build a sea wall that would limit the depth of sea views but also reduce future costs associated with coastal erosion (Hynes et al. 2022). Avenues for further research include an examination of the competing relationship between the positive amenity of sea views and the negative amenity of exposure to coastal flood risk, as
well as the differential valuation of amenities like views across income cohorts, tenure and other demographic characteristics.

Acknowledgements
We would like to thank Stephen Ryan for his contribution to the generation of the NLP data used in this research. We would also like to thank the following for their assistance with data and related queries: daft.ie, the Central Statistics Office, An Taisce, Ordinance Survey Ireland and Ireland’s Environmental Protection Agency. Tom Gillespie would like to acknowledge the financial support of Irish Research Council. We would like to thank participants at the 2022 EAERE Conference for helpful comments.
References


https://econpapers.repec.org/paper/tcdtcduee/tep0518.htm


Tables

Table 1: Sample breakdown

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>All listings 2014-2019</td>
<td>375,190</td>
</tr>
<tr>
<td>Excluding price outliers (€30,000-€2,000,000)</td>
<td>348,662</td>
</tr>
<tr>
<td>Excluding missing bedrooms/bathrooms</td>
<td>311,332</td>
</tr>
<tr>
<td>Excluding those &gt;5k from coastline</td>
<td>105,254</td>
</tr>
<tr>
<td>Including only building level accuracy (final sample)</td>
<td>100,030</td>
</tr>
</tbody>
</table>

Table 2. Observations per SFE

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Number of spatial units in sample</th>
<th>Mean Number of Observations Within Spatial Unit</th>
<th>Median Number of Observations Within Spatial Unit</th>
<th>10th Percentile of Number of Observations within Spatial Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Market</td>
<td>50</td>
<td>4,449</td>
<td>3,691</td>
<td>1,717</td>
</tr>
<tr>
<td>Micro-Market</td>
<td>261</td>
<td>871</td>
<td>729</td>
<td>327</td>
</tr>
<tr>
<td>Micro-Market * Year</td>
<td>1,247</td>
<td>152</td>
<td>127</td>
<td>53</td>
</tr>
<tr>
<td>Electoral Division</td>
<td>832</td>
<td>350</td>
<td>265</td>
<td>63</td>
</tr>
<tr>
<td>ED*Year</td>
<td>4,509</td>
<td>62</td>
<td>47</td>
<td>11</td>
</tr>
<tr>
<td>Small Area</td>
<td>5,571</td>
<td>38</td>
<td>29</td>
<td>9</td>
</tr>
</tbody>
</table>

Note the table above gives an indication of the within variation of the level of spatial fixed effects in terms of the number of observations, used in the analysis. The sample is restricted to within 5km of the coastline as specified in the text.
Table 3: Baseline results

<table>
<thead>
<tr>
<th>Treatment of SE</th>
<th>Robust Standard errors &amp; Clustered at level of SFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of spatial FE</td>
<td>Local Market</td>
</tr>
<tr>
<td>Treatment of SE</td>
<td>Robust Standard errors &amp; Clustered at level of SFE</td>
</tr>
<tr>
<td>Level of spatial FE</td>
<td>Local Market</td>
</tr>
<tr>
<td>Coastal Variables</td>
<td></td>
</tr>
<tr>
<td>Distance To Coastline (log)</td>
<td>-0.038</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Distance To Transitional (log)</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>0.34</td>
</tr>
<tr>
<td>Sea View Variables</td>
<td></td>
</tr>
<tr>
<td>Seaview breadth</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td>Seaview depth</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>Controls</td>
<td>YES</td>
</tr>
<tr>
<td>Observations</td>
<td>100,030</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.800</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.289</td>
</tr>
<tr>
<td>Number of absorbed SFE</td>
<td>50</td>
</tr>
</tbody>
</table>

Notes: Robust p-values, clustered at the level of spatial fixed effect, are reported below coefficients. Columns show different levels of spatial fixed effects denoted at header, while controls include location, dwelling and time listed on the market, as discussed in the text.
Table 4: Separating GIS and NLP measures of views

<table>
<thead>
<tr>
<th>Treatment of SE</th>
<th>Robust Standard errors &amp; Clustered at level of SFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of spatial FE</td>
<td>Small Area &amp; Micro-Market*Year</td>
</tr>
<tr>
<td>Specification Number</td>
<td>Baseline 1 2 3 4 5</td>
</tr>
<tr>
<td>Dependent Variable: natural log of the listed sale price</td>
<td></td>
</tr>
</tbody>
</table>

### Coastal Variables

<table>
<thead>
<tr>
<th>Coastal Variables</th>
<th>Baseline 1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance To Coastline (log)</td>
<td>-0.061 -0.062 -0.062 -0.062 -0.061 -0.060</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00 0.00 0.00 0.00 0.00</td>
</tr>
<tr>
<td>Distance To Transitional (log)</td>
<td>-0.014 -0.015 -0.014 -0.015 -0.012 -0.013</td>
</tr>
<tr>
<td></td>
<td>0.01 0.01 0.01 0.01 0.02 0.01</td>
</tr>
</tbody>
</table>

### Sea View Variables

<table>
<thead>
<tr>
<th>Sea View Variables</th>
<th>Baseline 1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seaview breadth</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Seaview depth</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
</tr>
</tbody>
</table>

### Sea View Variables (GIS only)

<table>
<thead>
<tr>
<th>Sea View Variables (GIS only)</th>
<th>Baseline 1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seaview breadth</td>
<td>0.003 0.008</td>
</tr>
<tr>
<td></td>
<td>0.14 0.00</td>
</tr>
<tr>
<td>Seaview depth</td>
<td>0.013 0.015</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
</tr>
</tbody>
</table>

### Sea Views NLP

<table>
<thead>
<tr>
<th>Sea Views NLP</th>
<th>Baseline 1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
</tr>
</tbody>
</table>

### Sea Views NLP (GIS=0)

<table>
<thead>
<tr>
<th>Sea Views NLP (GIS=0)</th>
<th>Baseline 1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
</tr>
</tbody>
</table>

### Controls

<table>
<thead>
<tr>
<th>Controls</th>
<th>YES YES YES YES YES YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>100,030 100,030 100,030 100,030 100,030 100,030</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.890 0.890 0.890 0.890 0.890 0.890</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.222 0.222 0.222 0.222 0.222 0.222</td>
</tr>
<tr>
<td>Number of absorbed SFE</td>
<td>5,571 5,571 5,571 5,571 5,571 5,571</td>
</tr>
</tbody>
</table>

Notes: Robust p-values, clustered at the level of spatial fixed effect, are reported below coefficients. Columns show different robustness specifications numbered in the headers, while controls include location, dwelling and time listed on the market, as discussed in the text.
Table 1: Controlling for the total viewshed of a property

<table>
<thead>
<tr>
<th>Treatment of SE</th>
<th>Robust Standard errors &amp; Clustered at level of SFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of spatial FE</td>
<td>Small Area &amp; Micro-Market*Year</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specification</th>
<th>Baseline</th>
<th>Adding Land Viewshed</th>
<th>Adding Slope</th>
<th>Adding Land Viewshed and Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable: natural log of Listed sale price</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coastal Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance To Coastline (log)</td>
<td>-0.060</td>
<td>-0.061</td>
<td>-0.060</td>
<td>-0.061</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Distance To Transitional (log)</td>
<td>-0.015</td>
<td>-0.016</td>
<td>-0.015</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Sea View Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seaview breadth</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Seaview depth</td>
<td>0.013</td>
<td>0.012</td>
<td>0.013</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Land view and Slope Controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log of Land Area Viewshed</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Log of % Rise of Slope</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.001</td>
<td>0.97</td>
<td>0.83</td>
</tr>
<tr>
<td>Controls</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Observations</td>
<td>99,447</td>
<td>99,447</td>
<td>99,447</td>
<td>99,447</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.890</td>
<td>0.890</td>
<td>0.890</td>
<td>0.890</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.221</td>
<td>0.221</td>
<td>0.221</td>
<td>0.221</td>
</tr>
<tr>
<td>Number of absorbed SFE</td>
<td>5,566</td>
<td>5,566</td>
<td>5,566</td>
<td>5,566</td>
</tr>
</tbody>
</table>

Notes: Robust p-values, clustered at the level of spatial fixed effect, are reported below coefficients. Columns show different levels of spatial fixed effects denoted at header, while controls include location, dwelling and time listed on the market, as discussed in the text.
Figure Titles

Figure 1. Distribution of visible points as a function of distance to coastline

Note: The above figure shows summary statistics of the number of (inner) points visible for properties with at least one inner point visible in their viewshed, by distance to the coastline, with sample sizes for each distance bin reported in the horizontal axis.

Figure 2. Graph of sea view coefficients by level of SFE – Illustrating the results from Table 3.

Notes: The above figure displays point estimates and confidence intervals for the results in table 3.
Hereafter ‘Ireland’ refers to the Republic of Ireland unless otherwise stated.

Specifically, the criteria for a view was the presence of either ‘overlook’ or ‘view’ and any of the following 17 terms: bay, beach, cliff, coast, coastline, cove, estuary, harbour, Irish sea, island, marina, ocean, quay, sea, seafront, seaside, strand.

A similar technique is used in Qiang, Shen, and Chen (2019), however, the original working paper version of this study predates this publication: Gillespie, Hynes, and Lyons (2019). It is also employed and referenced in Dempsey et al. (2018).

A noteworthy limitation of the view depth measure is that it is calculated based on the nearest inner point by Euclidean distance, and not necessarily the nearest visible inner point. This may be a cause for concern for houses close to cliffs or on the crest of mountains.

See Appendix, Table A3 for Type 1 and Type 2 errors comparing the views term and the GIS view measure.

The full list of controls is outlined in Appendix Table(s) A4, A5, and A6.

A concern about this causal hypothesis is that if the variation in view is derived from variation in the hilly and mountainous landscape, how do we know if model estimates are driven by view or are the model estimates driven by the landscape topography? We have undertaken an additional exercise to make an initial examination to address this concern. Specifically, we ran an analysis where a measure of slope was the outcome variable and zonal statistics were used to generate the mean slope in % rise for a 10 km^2 area around the property. Regressing this on the sea view measure whilst controlling for distance to coast and various spatial fixed effects, we found that there was no conditional statistically significant effect on that particular measure of ‘hillyness’. Furthermore, when we include slope (% incline) as a control in the main specification, it has little to no effect on the measures of sea view (see table 5).

The distinction between the two is made by the Environmental Protection Agency (EPA) in Ireland. For the purposes of this study, we treat them separately in regressions but conceptually they are thought of as effectively the same thing. The point in which a river becomes the sea in an inlet is difficult to define and from the perspective of a home buyer a transitional water body would likely be considered as sea.

The sample only includes ‘arm’s length’ transactions. Transactions that are more than 50% different from the list prices are excluded.