Urban and socio-economical correlates of property price evolution: Application to Dublin’s costal area

Understanding the characteristics of the housing market is essential for both sellers and buyers. However, the housing market is influence by multiple factors. In this paper, the urban and socio-economical structre of an area is used to predict the price of 10389 houses sold in 2018 in the city of Dublin. More precisely, the direct distance to every houses to 172 amenity types taken from OpenStreetMap is calculated and a gradient boost linear regression performed. Using these feature, the model explains 43% of the housing price variance. Most important factors in this model are the proximity to an embassy and to a grassland. In addition, results of the population census is also used to correlate with the price of houses. From this census, 48 features are used as the input of a gradient boost linear regression model. All together the socio-economic features are explaining 42% of the housing price variance as well. The most important socio-economic features are the density of houses having more than eight or more rooms in the area, the density of young children and the density of individual reporting having no religion. By taking into account either urban or socio-economic feature, it is possible to accurately estimate housing prices and to predict their evolution.

(ref:distrib-plot) Overall (A) and geographic (B) distribution of the houses prices density in the city of Dublin in 2018. The geographic distribution was obtained by using a generalized additive model with soap film smooth parameter to take into account the influence of coastal boundaries. Geographic estimates outside of the 95% Confidence Interval were removed. (ref:osm-features-xgb) Property price prediction accuracy (A) and Property price prediction error (B) using urban features with XGBoost. (ref:osm-features-table) Urban features importance (higher than 1%). (ref:census-features-xgb) Property price prediction accuracy (A) and Property price prediction error (B) using socio-economic features with XGBoost. (ref:census-features-table) Socio-economic features importance (higher than 1%).

# Introduction

First time property acquisition is an important achievement in individuals’ lifetime. It provides not only a housing security but also the feeling of being a landowner. However the access to the status of landowner is a challenge for low to middle classes as it corresponds an expensive spending which will potentially have budget impacts for years (Savage et al. 2014). For this reason, understanding property price markets as well as factors driving price is essential to identify investment opportunities.

Beside their inherent characteristics such as size, design and materials (Bourassa, Cantoni, and Hoesli 2007; Liu 2013), the property valuation is dictated by multiple external factors explaining their spatial autocorrelation (Basu and Thibodeau 1998). Some of these factors are macroeconomic (Antonakakis and Floros 2016), related to the evolution of population general wealth while others are related to the characteristics of the neighbourhood (Dubin 1998). These characteristics include the urban and socio economic structure of the neighbourhood (Goodman and Thibodeau 1998). However, it is difficult to evaluate the influence of such external factors on property valuation (Clapp, Kim, and Gelfand 2002). By using machine learning algorithms, prediction of house prices became more and more accurate. Techniques used usually include artificial neural network due to the volatility of the market (Limsombunchai 2004; Feng and Jones 2015; Selim 2009).

In this paper, an alternative machine learning algorithm is used to identify urban and socio-economical correlates of property price evolution in the area of Dublin, Ireland in 2018. Due to the high number of potential features, an eXtreme Gradient Boosting (XGBoost) regression model is believed to obtain the best results with low parameter customisation and high speed.

# Method

Since 2010, under the Irish Property Services (Regulation) Act, all individuals acquiring a property in Ireland has to declare it to Property Services Regulatory Authority (PSRA). It includes details such as the date of sale, the price and the address of all residential properties purchased in Ireland as declared to the Revenue Commissioners for stamp duty purposes (“Residential Property Price Register,” n.d.). It must be noticed that data is filed electronically by persons doing the conveyancing of the property on behalf of the purchaser and errors may occur when the data is being filed.

Because of the evolution of the urban planning and the evolution of socio-economic features measured by the census, only houses sold in 2018 where processed to avoid potential temporal incoherences. In 2018, a total of 10395 in the area of Dublin were sold and constitute the database used to identify the urban and socio-economical correlates of property prices.

In order to evaluate the spacial distribution of the property sold, property addresses were geocoded (i.e., converted to latitude and longitude) using the OpenStreetMap API. OpenStreetMap is a collaborative project which aims to create and provide access to free editable maps of the world (Haklay and Weber 2008).

To estimate the geographical distribution of the price density, multiple methods such as multiple regression (McCluskey et al. 2000) or Bayesian smoothing (Clapp, Kim, and Gelfand 2002) have been employed. Even if these methods are efficient in a non-restricted space, they have limitations when they are used in coastal areas. To deal with the influence of boundaries on price estimation, a generalised additive model using soap film smoothing has been used (Wood, Bravington, and Hedley 2008).

## Distance to urban features

OpenStreetMap combines information about more than 400 urban features including road information and building information to categorize amenities, leisure or tourism structure for example (“OpenStreetMap Wiki: Map Features,” n.d.). Among these features only 160 are available or relevant to the Dublin Area. The distance between each property and the closest point corresponding to each of the 160 urban features.

## Density of socio-economic features

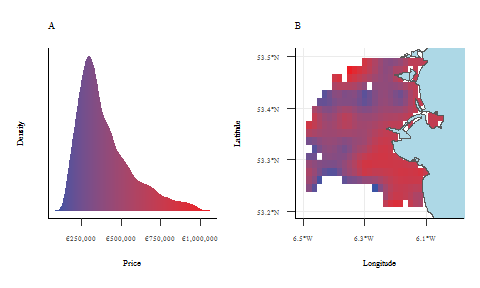
Additionally to the distance to every urban feature, each house sold in 2018 was related to the socio-economic features of the small area including the house measured by the Irish census 2016 (“Central Statistics Office: Census 2016 Small Area Population Statistics,” n.d.). The results of Irish 2016 census consultation is accessible through both the Central Statistics Office and All-Island Research Observatory. The data obtained can be mapped over Ireland small area boundaries which are fraction of Irish Electoral Division map. The social features extracted are corresponding to population information, religion, carers and health. Economic features correspond to the type of each small area including the proportion of housing type, rooms number, occupancy and tenure per small area. Each property is then associated to the value corresponding to its small area. For anonymisation purpose, the result from small area having less than five respondents to the census were converted to a proportion of five respondents.

## Gradient Boosting Regressions

Lastly, an eXtreme Gradient Boosting (XGBoost) regression model was calculated to examine the extent to which house prices relies on urban and on socio-economic features. The XGBoost algorithm optimizes the prediction accuracy by performing iterative least-squares regressions (set to 100 iterations as the best trade-off between accuracy and speed), thereby minimizing the root mean squared error (Friedman 2001; Chen and Guestrin 2016). The original dataset has been randomly split to 80% for training and 20% for testing the models. Models accuracy is estimated using the coefficient of determination (), the Root Mean Square Error () and the Mean Absolute Error () of the predicted property price values of the test dataset.

# Results

In 2018, the average price of a sold property in Dublin is €330,364 (SD = €180,448). In order to remove potential human errors and outliers, prices higher or lower than 1 SD were removed from the original dataset.

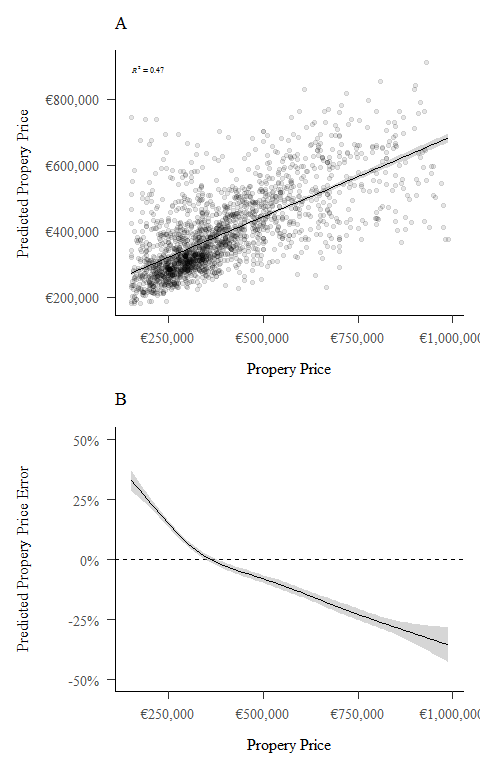


(ref:distrib-plot)

The density of housing prices reveals an unimodal distribution of property prices (Figure @ref(fig:distrib-plot)A). In addition, the geographical distribution of properties prices indicates a higher valuation in the South-West of the city as well as on the coast line (Figure @ref(fig:distrib-plot)B).

## Correlates with urban features

Using the XGBoost algorithm, the 160 urban features are explaining 47% of the property price variance (, ) with a of €124,364 and a of €83,963.61.



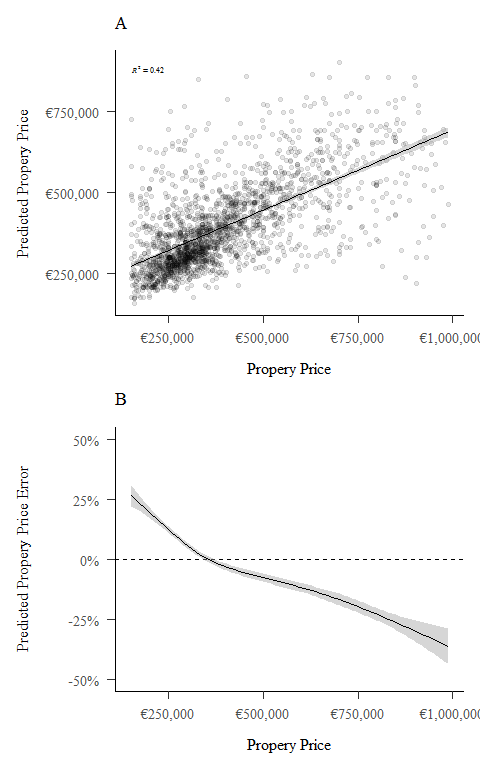
(ref:osm-features-xgb)

The overall correlation of the predicted property prices with urban features is shown Figure @ref(fig:osm-features-xgb)A. It can be noticed that property prices situated at the low and the high end of the range are the most difficult to predict (Figure @ref(fig:osm-features-xgb)B). A possible explanation is the absence of distinctive and recurrent patterns in urban features for these houses. However, the prices higher than €300,000 are potentially driving down the prediction accuracy due to outliers.

By analysing their importance (Table @ref(tab:osm-features-table)), the most relevant geographical features to predict housing prices are the distance to an embassy (18.2%) and the distance to natural grasslands such as parks and gardens (6.7%).

## Correlates with socio-economic features

Using the XGBoost algorithm, the 48 socio-economic features are explaining 42% of the property price variance (, ) with a of €131,322 and a of €87,328.79.



(ref:census-features-xgb)

The overall correlation of the predicted property prices with socio-economic features is shown Figure @ref(fig:census-features-xgb)A. Similarly as the model used for urban features, the model based on socio-economic features reveals that prices lower than €300,000 lead to the highest errors in over valuation while prices higher than €300,000 lead to systematic under valuation errors (Figure @ref(fig:census-features-xgb)B).

According to the analysis of the relative importance of socio-economic features (Table @ref(tab:census-features-table)), the most important socio-economic features is the proportion of large houses (i.e., houses with eight or more rooms) in the small area containing the property (29.0%). It also appears that areas having a high proportion of young children (3.8%) as well as the proportion of people reporting having no religion (5.0%) influence the model.

# Discussion

While Dublin is a very specific city due to its low population density spread on a large surface, its housing market is one of the most expensive in Europe. The main reason is the structure of the market, mainly made of residential properties and few apartment buildings. In this study, the prices of 10387 properties sold in 2018 in the area of Dublin were analysed from the Property Services Regulatory Authority database in order to identify the potential factors correlating with these prices. The enforcement of a housing price registry has been shown as being an essential tool to regulate the housing market (Tomson 2016). However, a major limitation of this study is the absence of characteristics for the property itself. Indeed, prices are mainly determined factors such as the size of the property or the number of rooms.

Despite the absence of property characteristics, urban and socio-economic features were used to identify spatial correlates to the housing prices. Results revealed that the proximity to embassies and grasslands is a driver of house prices. Embassies are in general located in the richest areas of cities which are also the nicest and the most secure part of cities. In addition, the density of embassies and properties allocated to embassy staff, reduces the density of property available on the market and drive their price up. The distance to a park or a natural area is also a very important factor for house prices. Results also revealed that the density of large houses in the area and the proportion of individuals reporting having no religion are very important. Again, the size of the property sold was not included in the database; however, the higher the density of large houses in the area, the higher the probability for the property to be a large house. Second most important feature is the density of young children. While the literature is not conclusive of the link between wealth and family size, a future research could investigate the mediation effect of houses size between family size and property price. Finally, the relation with the expression of religious belief, or more precisely its absence, is a very interesting feature. Again, the relationship between religion and income is not confirmed by the literature. Whereas some studies reveal a positive correlation between religion and income (Guiso, Sapienza, and Zingales 2003; Elgin et al. 2013) others reveals no evidence of such relationship (De La O and Rodden 2008). Here, it appears that prices of houses are higher is era with a high density of people indicating having no religion.

# Conclusion

The evolution of housing prices is a real problem in most of European capitals and specially in Dublin. Given their significant increase, houses are less and less affordable for individuals. By preforming a feature analysis with urban and socio-economical feature it is possible to evaluate and predict the potential price of a house. Indeed features such as presences of embassies or parks are criteria that influence significantly the price of houses. Similarly, the characteristics of inhabitants in the area such as religion, health and age is correlated to the evolution of housing prices. These results allow to understand why an area have higher prices than others.

# References

Antonakakis, Nikolaos, and Christos Floros. 2016. “Dynamic Interdependencies Among the Housing Market, Stock Market, Policy Uncertainty and the Macroeconomy in the United Kingdom.” *International Review of Financial Analysis* 44: 111–22.

Basu, Sabyasachi, and Thomas G Thibodeau. 1998. “Analysis of Spatial Autocorrelation in House Prices.” *The Journal of Real Estate Finance and Economics* 17 (1): 61–85.

Bourassa, Steven C, Eva Cantoni, and Martin Hoesli. 2007. “Spatial Dependence, Housing Submarkets, and House Price Prediction.” *The Journal of Real Estate Finance and Economics* 35 (2): 143–60.

“Central Statistics Office: Census 2016 Small Area Population Statistics.” n.d. <http://census.cso.ie/sapmap>.

Chen, Tianqi, and Carlos Guestrin. 2016. “Xgboost: A Scalable Tree Boosting System.” In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–94.

Clapp, John M, Hyon–Jung Kim, and Alan E Gelfand. 2002. “Predicting Spatial Patterns of House Prices Using Lpr and Bayesian Smoothing.” *Real Estate Economics* 30 (4): 505–32.

De La O, Ana L, and Jonathan A Rodden. 2008. “Does Religion Distract the Poor? Income and Issue Voting Around the World.” *Comparative Political Studies* 41 (4-5): 437–76.

Dubin, Robin A. 1998. “Predicting House Prices Using Multiple Listings Data.” *The Journal of Real Estate Finance and Economics* 17 (1): 35–59.

Elgin, Ceyhun, Turkmen Goksel, Mehmet Y Gurdal, and Cuneyt Orman. 2013. “Religion, Income Inequality, and the Size of the Government.” *Economic Modelling* 30: 225–34.

Feng, Yingyu, and Kelvyn Jones. 2015. “Comparing Multilevel Modelling and Artificial Neural Networks in House Price Prediction.” In *2015 2nd Ieee International Conference on Spatial Data Mining and Geographical Knowledge Services (Icsdm)*, 108–14. IEEE.

Friedman, Jerome H. 2001. “Greedy Function Approximation: A Gradient Boosting Machine.” *Annals of Statistics*, 1189–1232.

Goodman, Allen C, and Thomas G Thibodeau. 1998. “Housing Market Segmentation.” *Journal of Housing Economics* 7 (2): 121–43.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2003. “People’s Opium? Religion and Economic Attitudes.” *Journal of Monetary Economics* 50 (1): 225–82.

Haklay, Mordechai, and Patrick Weber. 2008. “Openstreetmap: User-Generated Street Maps.” *IEEE Pervasive Computing* 7 (4): 12–18.

Limsombunchai, Visit. 2004. “House Price Prediction: Hedonic Price Model Vs. Artificial Neural Network.” In *New Zealand Agricultural and Resource Economics Society Conference*, 25–26.

Liu, Xiaolong. 2013. “Spatial and Temporal Dependence in House Price Prediction.” *The Journal of Real Estate Finance and Economics* 47 (2): 341–69.

McCluskey, William J, William G Deddis, Ian G Lamont, and Richard A Borst. 2000. “The Application of Surface Generated Interpolation Models for the Prediction of Residential Property Values.” *Journal of Property Investment & Finance*.

“OpenStreetMap Wiki: Map Features.” n.d. <https://wiki.openstreetmap.org/wiki/Map_Features>.

“Residential Property Price Register.” n.d. <https://propertypriceregister.ie>.

Savage, Michael, James Barlow, Peter Dickens, and Tom Fielding. 2014. *Property Bureaucracy & Culture*. Routledge.

Selim, Hasan. 2009. “Determinants of House Prices in Turkey: Hedonic Regression Versus Artificial Neural Network.” *Expert Systems with Applications* 36 (2): 2843–52.

Tomson, AIVAR. 2016. “Property Sales Register as a Tool to Improve the Quality of Valuation and Market Research: Lessons Learned from Selected Countries.” In *World Bank Conference on Land and Poverty, Washington, Dc, March*, 14–18.

Wood, Simon N, Mark V Bravington, and Sharon L Hedley. 2008. “Soap Film Smoothing.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70 (5): 931–55.