REM WORKING PAPER SERIES

On the uncertainty of real estate price Predictions

João A. Bastos, Jeanne Paquette

REM Working Paper 0314-2024

March 2024

REM – Research in Economics and Mathematics
Rua Miguel Lúpi 20,
1249-078 Lisboa,
Portugal

ISSN 2184-108X

REM – Research in Economics and Mathematics

Rua Miguel Lupi, 20
1249-078 LISBOA
Portugal

Telephone: +351 - 213 925 912
E-mail: rem@iseg.ulisboa.pt

https://rem.rc.iseg.ulisboa.pt/

https://twitter.com/ResearchRem

https://www.linkedin.com/company/researchrem/

https://www.facebook.com/researchrem/

On the uncertainty of real estate price predictions

João A. Bastos* Jeanne Paquette

Lisbon School of Economics & Management (ISEG) and REM
University of Lisbon

Abstract

Uncertainty quantification associated with real estate appraisal has largely been overlooked in the literature. In this paper, we address this gap by analyzing the uncertainty in automated property valuations using conformal prediction, a distribution-free procedure for constructing prediction intervals with valid coverage in finite samples. Through an empirical study of property prices in the San Francisco Bay Area, we find that prediction intervals obtained using conformal quantile regression have exact coverage. In contrast, prediction intervals obtained from non-conformal quantile regressions severely undercover the data. Furthermore, we show that the intervals adapt to various characteristics of the dwellings, which is crucial given the heterogeneous nature of real estate data. Indeed, we observe that larger and older properties, those in both low and high-income neighborhoods, as well as those on the market for less than one year are more challenging to evaluate.

Keywords: Real estate; Automated valuation model; Conformal prediction; Quantile regression; Machine learning.

1 Introduction

Real estate brokers and investment advisers use predictive tools like automated valuation models (AVMs) to offer guidance and advice to prospective property buyers, investors, and sellers. These tools allow them to estimate property values, analyze the impact of various property attributes on prices, and provide informed recommendations to their clients. The conventional method for estimating house prices is hedonic, meaning it relies on the attributes of the dwellings. Typically, this involves using a linear regression model estimated through least squares to model house prices as a function of these attributes (Goodman, 1978). In recent years, with the growing accessibility of large databases,

*Corresponding author: jbastos@iseg.ulisboa.pt

machine learning models such as neural networks and ensemble methods have become increasingly prevalent for property valuation tasks, gradually replacing linear hedonic models (see, inter alia, Peterson and Flanagen, 2009; Guan et al., 2011; Ho et al., 2021; Hjort et al., 2022). These methods possess strong capabilities to identify and model nonlinear patterns, making them more suitable for property valuations.

Surprisingly, the literature on real estate appraisal has predominantly focused on point predictions, overlooking the quantification of the associated uncertainty. When a real estate company lists a new property, the real estate broker may not only aim to provide an expected selling price to the seller but also an interval of prices within which the property is likely to sell with a high degree of confidence. On the other side of the transaction, we may have individuals who intend to sell a property and reinvest the proceeds. In addition to wanting to know the expected selling price, these individuals may also seek a reliable range of potential selling prices to effectively plan a future investment. Consider a bank providing a home mortgage loan. The bank may seek a conservative estimate of the appraised value when assessing the debtor's default risk. This value can be obtained from the lower bound of a prediction interval. In all the examples above, the agents need to quantify the uncertainty of the prediction provided by their valuation models. This uncertainty may be quantified via prediction intervals.

The problem at hand involves a set of n property prices $\{Y_i\}_{i=1}^n$ and corresponding explanatory variables $\{\boldsymbol{X}_i\}_{i=1}^n$, such as the size, the number of bedrooms, and location. Our objective is to construct a prediction interval $\mathcal{C}(\boldsymbol{X}_{n+1}) \subseteq \mathbb{R}$ for the unknown price Y_{n+1} of a property with known characteristics $\boldsymbol{X}_{n+1}$. Given a significance level α, the goal is to ensure that the prediction interval has high probability of containing the unknown property price

$$\Pr(Y_{n+1} \in \mathcal{C}(\boldsymbol{X}_{n+1})) \geq 1 - \alpha. \tag{1}$$

Furthermore, this relationship should hold for any joint distribution of the data $P_{\boldsymbol{X},Y}$.

Regrettably, the conventional method for constructing prediction intervals for a linear hedonic model, based on the studentized least squares estimators, relies on the assumption of normality for the error term. Because this assumption rarely holds, studentized prediction intervals do not ensure exact coverage, meaning they do not include the observed price at a rate aligned with the desired level. A distribution-free alternative is to construct prediction intervals using regression quantiles (Zhou and Portnoy, 1996). Training two models to learn the quantile functions $q_{\frac{\alpha}{2}}(\boldsymbol{X})$ and $q_{1-\frac{\alpha}{2}}(\boldsymbol{X})$ enables us to obtain a prediction interval for Y_{n+1} with a nominal coverage level of $1-\alpha$, given by

$$\mathcal{C}(\boldsymbol{X}_{n+1}) = \left[q_{\frac{\alpha}{2}}(\boldsymbol{X}_{n+1}), q_{1-\frac{\alpha}{2}}(\boldsymbol{X}_{n+1})\right]. \tag{2}$$

Regarding the functional form for $q_\alpha(\boldsymbol{X})$ we could consider the conventional quantile regression of Koenker and Bassett (1978). But this would impose a linear association between the conditional quantiles of the target variable and the predictors, whereas property

prices typically exhibit nonlinear relationships with the dwelling's characteristics. Furthermore, these intervals based on quantile regressions are not guaranteed to provide exact coverage, even if the true relationship between prices and property characteristics is linear.

Alternatively, we could employ nonlinear machine learning models. Indeed, some approximate solutions have been proposed for constructing prediction intervals for neural networks (Papadopoulos et al., 2001) and ensemble methods (Wager et al., 2014). However, the prediction intervals derived from these solutions cannot be constructed without making strong assumptions or relying on large-sample asymptotic approximations that may not be easily justifiable. Another approach could involve training two machine learning models for conditional quantiles to construct prediction intervals, following the principle outlined in Equation 2. Indeed, many machine learning algorithms can be adapted to learn conditional quantiles (see Taylor (2000), in the case of neural networks and Meinshausen (2006), in the case of random forests). However, as the empirical results below demonstrate, this approach gives intervals that strongly undercover the data, resulting in intervals that fail to include the observed price more often than the nominal level.

In this paper, we investigate the uncertainty in real estate price predictions using conformal prediction – a method for constructing statistically rigorous prediction intervals, ensuring valid coverage in finite samples without relying on distributional assumptions (Papadopoulos et al., 2002; Vovk et al., 2005, 2009; Lei et al., 2013; Lei and Wasserman, 2014). Specifically, we employ conformal quantile regressions (Romano et al., 2019) to construct prediction intervals of the type described in Equation 2. The method is distribution-free, meaning that prediction intervals with valid coverage are obtained regradless of the distribution of the data. In contrast to alternative conformal prediction techniques, conformal quantile regression constructs flexible and adaptive intervals, capable of capturing the varying levels of uncertainty associated with the prices of different dwellings. Furthermore, it can be used with any model for conditional quantiles.

The empirical analysis is based on data from the dynamic and highly competitive real estate market of the San Francisco Bay Area, where property valuations are influenced by a multitude of complex factors. San Francisco has some of the highest real estate prices in the United States, with median property prices well above the national average. The dataset contains information on sold properties over the period from 2020 to 2023. San Francisco has a diverse array of neighborhoods characterized by varying income levels and amenities. The city's geography, featuring hills and waterfront on three sides, has also contributed to the development of neighborhoods with distinct characteristics. Indeed, as we show below, housing prices vary greatly across different neighborhoods of the city. To account for the city's heterogeneity, we enriched our dataset with socio-economic, geographic, and demographic data from the city's survey records and the United States Census Bureau. We collected socio-economic data such as median income, employment

rate, and poverty rate. The discrepancies in socio-economic indicators between the different areas of San Francisco are very high. Therefore, explanatory variables of this nature must be included in the property valuation process. We also collected data on amenities such as the number of schools and parks, and the easiness of access to public transportation.

The machine learning models used for predicting property prices are based on tree ensembles. Preliminary results showed that these models predict more accurately property prices in our dataset than neural networks. This is not surprising since tree-based model still outperform neural networks on many problems with tabular data (Grinsztajn et al., 2022; Curth et al., 2024). In particular, we used a gradient boosting machine with a modified loss function for predicting conditional quantiles.

We are aware of two previous applications of conformal prediction in the area of real estate valuation. Bellotti (2017) applied conformal prediction to AVMs using data for the UK housing market. As acknowledged by the author, these data include a rather limited set of explanatory variables, lacking features such as the number of bedrooms and property size. Hence, Lim and Bellotti (2021) expanded the work of Bellotti (2017) by applying conformal prediction to a richer dataset, namely the public Ames Housing dataset. They also introduced and compared different variations of a conformal prediction algorithm. Our analysis introduces several novel elements to the study of uncertainty quantification in real estate markets. First, by using conformal quantile regression, we construct prediction intervals with widths that are adaptive to different property characteristics. Second, we do not limit our analysis to the study of the *marginal* coverage guarantee shown in Equation 1. Instead, we also study the behaviour of the *conditional* coverage guarantee

$$\Pr\left(Y_{n+1} \in \mathcal{C}(\boldsymbol{X}_{n+1}) | \boldsymbol{X}_{n+1} = \boldsymbol{x}\right)) \geq 1 - \alpha, \tag{3}$$

which should be satisfied for any $\boldsymbol{x}$ and any joint distribution of the data $P_{\boldsymbol{X},Y}$. This is a stronger property than the marginal coverage property stated in Equation 1. In fact, conditional coverage is rarely achieved across all $\boldsymbol{x}$ (Vovk, 2012), and our empirical findings corroborate this observation. However, evaluating the conditional coverage allows us to assess the adaptability of prediction intervals. Our findings reveal that prediction intervals adjust to distinct dwelling characteristics, a critical aspect given the heterogeneous nature of real estate data. Specifically, we observe that larger and older properties are more challenging to evaluate. The non-trivial relationship between a property's age and its market value introduces uncertainty. Indeed, older properties tend to undergo renovations over the years, a factor not readily accounted for in our dataset. This lack of information increases the difficulty in accurately evaluating older properties. Additionally, properties located in both low and high-income neighborhoods, as well as those on the market for less than one year, present increased difficulty in evaluation. The increased uncertainty at both ends of the income distribution can be attributed to the significantly

different variables influencing property prices at the opposite ends of San Francisco's real estate market.

Third, our dataset incorporates both private data sourced from a real estate marketplace company and public data obtained from the city's survey records and the United States Census Bureau Data. Therefore, our study includes not only housing attributes but also macro-level indicators that significantly influence the real estate market and drive housing prices. By integrating a diverse range of variables, our analysis aims to provide a more holistic understanding of the factors driving the uncertainty in housing prices. Finally, our study focuses on the unique real estate market of San Francisco, characterized by its large share of high-income tech workers, the diversity of neighborhoods in terms of income and amenities, and the geographical constraints and strict zoning laws. These factors ultimately lead to a wide variety of housing offers and prices.

The remainder of this paper is structured as follows. The following section presents our methodology for quantifying the uncertainty of real estate prices. Section 3 exposes the data used in the empirical study. In particular, we examine the most important determinants of real estate prices in the city of San Francisco. In Section 4, we present the main results of our analysis, focusing on both marginal and conditional prediction intervals. Section 5 provides a summary of the results and some concluding remarks.

2 Methodology

Let Y represent a property price, and $\boldsymbol{X}$ denote a vector of explanatory variables, such as the size, number of bedrooms, location, and other relevant factors. We assume that a real estate broker has a sample of n sold properties, denoted as $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^{n}$, which are used to train an automated valuation model. This agent wants to list a new property with known characteristics $\boldsymbol{X}_{n+1}$ and, naturally, unknown selling price Y_{n+1}. We assume that all observations in the dataset, including the new property, $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^{n+1}$, are exchangeable and drawn from a joint distribution of the data $P_{\boldsymbol{X},Y}$.[1] The agent aims to establish a prediction interval for the selling price $\mathcal{C}(\boldsymbol{X}_{n+1}) \subseteq \mathbb{R}$ that fulfills Equation 1 for any chosen coverage level $1-\alpha$. Moreover, this relationship should remain valid for any joint distribution $P_{\boldsymbol{X},Y}$.

2.1 Conformal prediction

A straightforward approach for constructing conformal intervals is the *split conformal prediction* method (Papadopoulos et al., 2002). This method entails dividing the training data into two subsets: $S_1 = (\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_1$ for estimation and $S_2 = (\boldsymbol{X}_i, Y_i) : i \in \mathcal{I}_2$ as the calibration set for obtaining *conformity scores*. A regression model $Y = f(\boldsymbol{X})$ is then trained on S_1, where any regression function can be employed. Subsequently, conformity

[1] This assumption holds automatically if the $(\boldsymbol{X}_i, Y_i)$ are independent and identically distributed.

scores are computed for each observation in the calibration set using the absolute residuals of the trained model $\hat{f}(\boldsymbol{X})$:

$$\hat{\varepsilon}_i = |Y_i - \hat{f}(\boldsymbol{X}_i)| \;\; \forall\, i \in \mathcal{I}_2. \tag{4}$$

For a given significance level α, a quantile $q_{1-\alpha}(\mathcal{I}_2)$ is computed from the empirical distribution of the conformity scores,

$$q_{1-\alpha}(\mathcal{I}_2) = \frac{\lceil (n_2+1)(1-\alpha) \rceil}{n_2} \text{th-quantile of } \hat{\varepsilon}_i : i \in \mathcal{I}_2, \tag{5}$$

where n_2 denotes the number of observations in the calibration set. Finally, the prediction interval for the unknown property price Y_{n+1} is constructed as:

$$\mathcal{C}(\boldsymbol{X}_{n+1}) = \left[\hat{f}(\boldsymbol{X}_{n+1}) - q_{1-\alpha}(\mathcal{I}_2),\; \hat{f}(\boldsymbol{X}_{n+1}) + q_{1-\alpha}(\mathcal{I}_2) \right]. \tag{6}$$

One limitation of this approach is that it generates constant prediction intervals, failing to capture the empirically observed varying width of intervals for property prices. However, the split conformal prediction method can be extended to address this limitation by introducing a locally adaptive approach known as *locally adaptive conformal prediction* (Papadopoulos et al., 2008). This method scales the absolute residuals $\hat{\varepsilon}_i$ by their dispersion at $\boldsymbol{X}_i$, allowing for non-constant prediction intervals. However, Romano et al. (2019, pp. 8) identifies several limitations in this approach.

Considering the limitations of early conformal prediction methods, Romano et al. (2019) propose the use of *conformal quantile prediction* to obtain intervals that are well-calibrated and adaptive to variations in the data. In conformal quantile prediction, the training dataset is also divided into two subsets: one for estimation and the other for calibration to obtain conformity scores. Given a coverage level of $1-\alpha$, two regression models for quantiles, namely $\hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X})$ and $\hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X})$, are trained using subset S_1. Then, conformity scores are calculated using:

$$\hat{\varepsilon}_i = \max \left[\hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X_i}) - Y_i, Y_i - \hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X_i}) \right], \; \forall\, i \in \mathcal{I}_2. \tag{7}$$

Next, Equation 5 is employed to compute the quantile $q_{1-\alpha}(\mathcal{I}_2)$ of the empirical distribution of these conformity scores. Finally, the conformalized prediction interval for Y_{n+1} is given by:

$$\mathcal{C}(\boldsymbol{X}_{n+1}) = \left[\hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X}_{n+1}) - q_{1-\alpha}(\mathcal{I}_2),\; \hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X}_{n+1}) + q_{1-\alpha}(\mathcal{I}_2) \right]. \tag{8}$$

Romano et al. (2019) demonstrates that if the observations $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^{n+1}$ are exchangeable, the prediction interval $\mathcal{C}(\boldsymbol{X}_{n+1})$ given in Equation 8 satisfies Equation 1. Furthermore, if the conformity scores $\{\hat{\varepsilon}_i : i \in \mathcal{I}_2\}$ are almost surely distinct, then $\mathcal{C}(\boldsymbol{X}_{n+1})$ is nearly perfectly calibrated:

$$\Pr(Y_{n+1} \in \mathcal{C}(\boldsymbol{X}_{n+1})) \leq 1 - \alpha + \frac{1}{1+n_2}. \tag{9}$$

2.2 Machine learning model for quantiles

To obtain the conformalized prediction intervals described in Equation 8, a regression model for quantiles is required. Here, we employ a modified version of a *gradient boosting machine* (Friedman, 2001) as the quantile regression model. The gradient boosting machine combines multiple base models to form a robust 'committee' of models. Usually, the base models are decision trees. The gradient boosting machine's prediction $\hat{Y}$ is derived from the sum of the predictions of a set of K decision trees $\{f_k(\boldsymbol{X})\}_{k=1}^{K}$:

$$\hat{Y} = \sum_{k=1}^{K} f_k(\boldsymbol{X}). \tag{10}$$

The first tree, $f_1(\boldsymbol{X})$, is trained on the original data. The decision trees that follow $\{f_k(\boldsymbol{X})\}_{k=2}^{K}$ are added to the committee sequentially. Every new tree in the committee, though, is trained using the errors made by the set of previous trees. The goal of this procedure is to correct the error made by the current committee. During each iteration, the tree selected for addition to the committee is the one that minimizes the regularized loss function:

$$\sum_{i=1}^{n} L\left(Y_i, \hat{Y}_i^{(k-1)} + f_k(\boldsymbol{X}_i)\right) + \gamma T + \frac{1}{2}\lambda ||\boldsymbol{w}_k||^2. \tag{11}$$

When aiming to predict the mean response, $L(\cdot)$ is the squared-error loss,

$$L\left(Y_i, \hat{Y}_i^{(k-1)} + f_k(\boldsymbol{X}_i)\right) = \left(Y_i - \hat{Y}_i^{(k-1)} - f_k(\boldsymbol{X}_i)\right)^2. \tag{12}$$

However, when it comes to predicting quantiles, we use an alternative loss function referred to as the 'pinball loss', which is defined as

$$L\left(Y_i, \hat{Y}_i^{(k-1)} + f_k(\boldsymbol{X}_i)\right) = \begin{cases} \alpha\left(Y_i - \hat{Y}_i^{(k-1)} - f_k(\boldsymbol{X}_i)\right), & \text{if } Y_i \geq \hat{Y}_i^{(k-1)} + f_k(\boldsymbol{X}_i) \\ (1-\alpha)\left(\hat{Y}_i^{(k-1)} + f_k(\boldsymbol{X}_i) - Y_i\right), & \text{otherwise.} \end{cases} \tag{13}$$

The last two terms in Equation (11) act as regularization penalties designed to discourage complex trees, thus mitigating overfitting of the training data by the committee. Parameter γ imposes a penalty on the number of terminal nodes in a tree, represented by T, while λ penalizes the magnitude of the tree weights $\boldsymbol{w}_k$. A gradient descent algorithm is used to minimize the loss function. Several efficient implementation of gradient boosting have been proposed. In this paper, we use the 'Light Gradient Boosting Machine', or LightGBM (Ke et al., 2017). LightGBM is known for its speed and efficiency.[2] Similar to many other machine learning algorithms, LightGBM has a variety of 'hyperparameters' – parameters not learned during training but requiring optimization.

[2]source: https://lightgbm.readthedocs.io/en/latest/Experiments.html

3 Data

Our empirical results are based on property prices from the San Francisco Bay Area, California, USA. San Francisco real estate is expensive, with property prices typically exceeding those elsewhere in the United States. We focus on residential properties. Our dataset comprises 8,249 observations of sold properties spanning from 2020 to 2023. This dataset was sourced from Zillow, a real estate marketplace company.

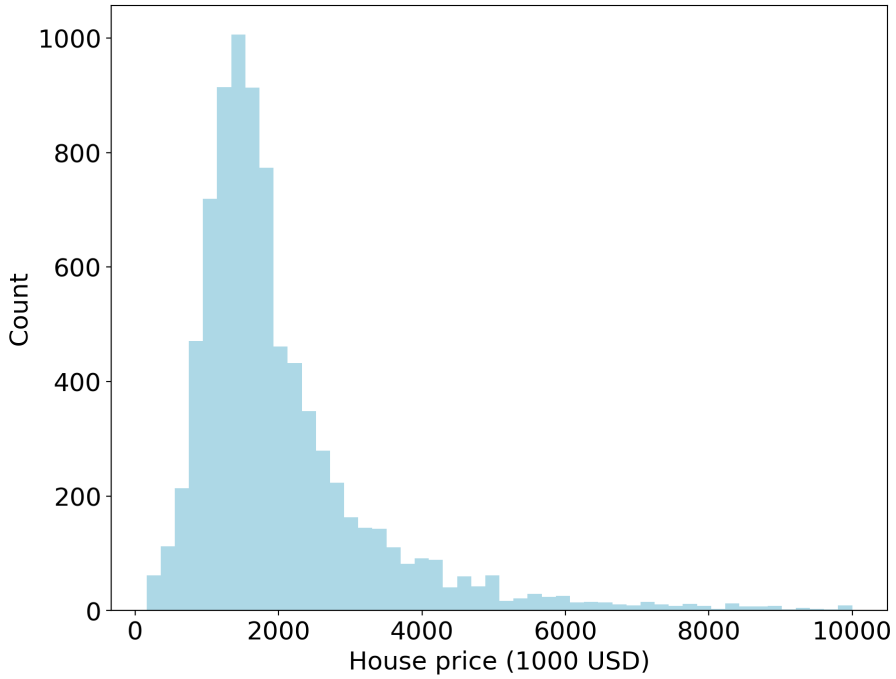

Figure 1: Distribution of prices, in thousands of USD, for sold properties in the dataset.

Figure 1 shows the distribution of sold property prices within the dataset, revealing a slight right skew. The mean price stands at $2,150,000, while the median price is $1,680,000. In principle, the requirement that the data must be exchangeable – as explained in Section 2 – could be compromised by increasing property prices. However, prices have remained relatively stable during the period covered by the data: the median selling price in 2020 stood at $1,702,000, which remained relatively stable in 2022 at $1,700,000. Furthermore, there was a decrease in the price of properties sold in 2023, with a median selling price of $1,380,000. Nonetheless, these variations in median prices did not affect the exact coverage guarantees provided by conformal prediction, as we show in the following section.

Zillow provided several explanatory variables detailing the characteristics of the housing units and their locations. The numeric explanatory variables and their summary statistics are listed in Table 1. Additionally, we were provided with a categorical variable indicating the property type, which includes categories such as single-family, townhouse, condo, and apartment. Single-family homes are the most expensive properties, boasting a median price of $1,700,000, while apartments represent the least expensive option, with

Variable	mean	$q_{0.05}$	median	$q_{0.95}$
House size (sq ft)	2004	900	1729	3931
Lot size (sq ft)	4207	1643	2790	6680
No. bedrooms	3.2	2	3	5
No. bathrooms	2.4	1	2	5
Year built	1936	1900	1931	2001
Days on market	525	41	542	1007

Table 1: Summary statistics of the numeric variables provided by the real estate company: mean, median, and the 0.05 and 0.95 quantiles.

a median price of \$900,000. Given that the vast majority (95%) of properties in the dataset are single-family homes, we group the remaining property types into a single category. The dataset further includes a categorical variable denoting the type of heating, distinguishing between central, radiant, electric, or no heating. Additionally, it includes dummy variables indicating whether the property features garage, parking space, and fireplace. The average property in the dataset is a single-family home with an area of approximately 2004 square feet, built around the mid-1930s, comprising 3 bedrooms and 2 bathrooms, with both a parking space and a garage, and equipped with central heating.

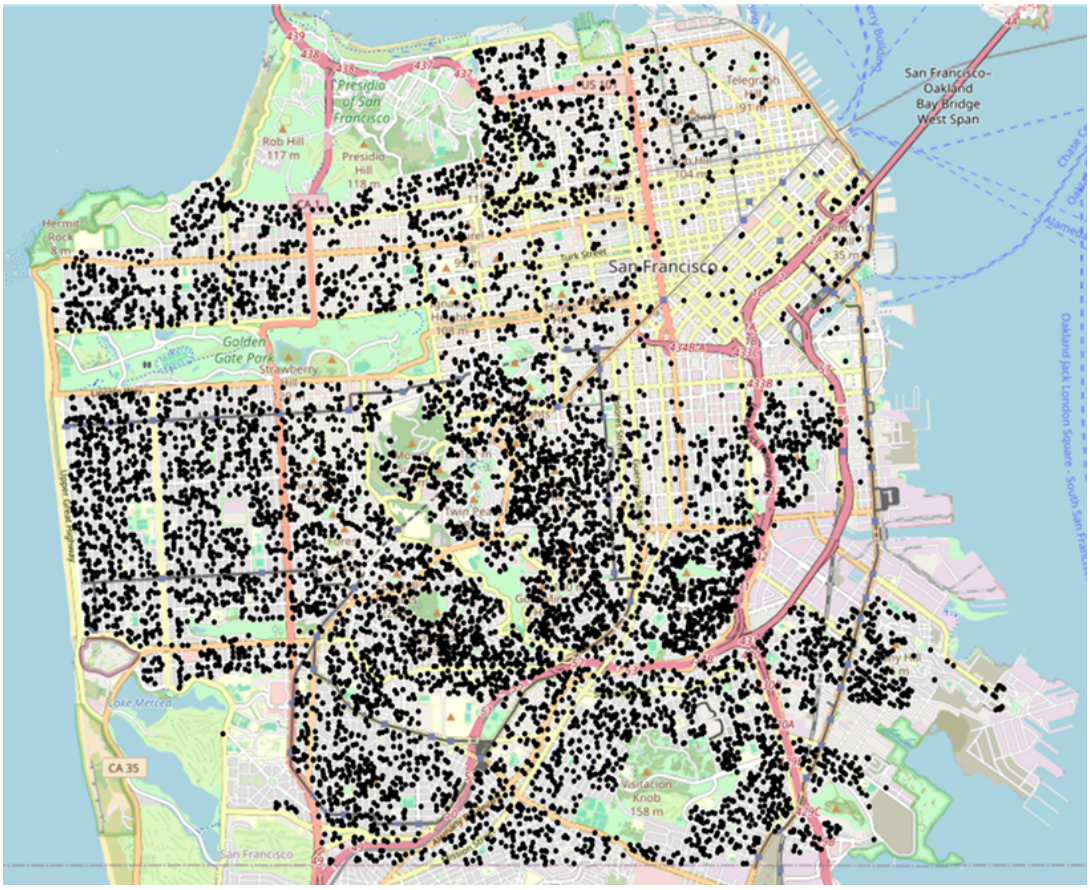

Figure 2: Location of the properties in San Francisco.

We were also provided with the latitude and longitude coordinates of the properties. As illustrated in Figure 2, the properties are evenly distributed across the residential areas of San Francisco. Empty spaces on the map correspond to parks or unzoned areas, such as Golden Gate Park and Lake Merced Park in the west, and the Presidio National Park in the northern tip of the peninsula. Alongside the coordinates, zip codes for the properties were also provided. We augmented our dataset with data obtained from the United States Census Bureau website, which offers a wide array of socio-demographic attributes. For each of the 24 neighborhoods in San Francisco, defined by zip codes, we obtained the following socio-demographic data: median income, median rent, racial demographics, poverty rate, employment rate, home-ownership rate, and number of housing units. Additionally, for each neighborhood, we collected variables defining the quality of amenities such as schools, parks, safety, transportation, and cleanliness. We created a dummy variable indicating whether a neighborhood has shoreline. Finally, we created four additional variables based on existing variables: the number of housing units per capita, the average room area (house area divided by the number of bedrooms and bathrooms), an interaction term between the number of days on the market and the house size, and another interaction term between the number of days on the market and the lot size. The rationale for creating these interaction variables is that larger properties may spend more time on the market.

The median annual income in San Francisco stands at approximately $132,000, but there is a significant income variation across neighborhoods, as reported in Table 2. Furthermore, Table 2 reveals a positive relationship between income levels across neighborhoods and the median property prices within those areas. For instance, the Van Ness/Civic Center neighborhood has the lowest median household income at $55,888. Furthermore, this neighborhood contends with a poverty rate of 18.5%, rendering it the second most impoverished area in San Francisco, trailing only behind Lower Nob Hill/Chinatown/Downtown. As expected, property prices in the Van Ness/Civic Center neighborhood are well below the city's average. Conversely, the Financial District South is the neighborhood with the highest annual median income, reaching $244,662, and it boasts elevated employment and home ownership rates, standing at 76% and 41.8%, respectively. Interestingly, properties in the Financial District South are not particularly expensive when compared to neighborhoods such as the Marina, which is the neighborhood with the second-highest median income. The spatial heterogeneity observed in San Francisco, characterized by extreme disparities in socio-economic indicators across various areas, underscores the importance of incorporating such variables into the property valuation process.

neighborhood	Median price	Median income	Poverty rate	Employment rate	Home ownership
Van Ness/Civic Center	1294	55888	18.5	62.3	9.2
South of Market	733	93143	16.2	71	18.4
Financial District South	1080	244662	11.5	76	41.8
Mission Bay	1350	161391	10.7	73.1	39.7
Lower Nob Hill/Chinatown/Downtown	2027	65392	19.7	58.9	9.5
Polk/Russian Hill (Nob Hill)	1680	104476	12.4	70	16.3
Inner Mission	1750	143938	9.9	72.9	38.3
Embarcadero	985	135735	18.2	54.9	21.3
Mission Terrace	1247	112795	9.1	63.2	67.8
Castro	2625	169459	5.8	74.7	43.9
Zion District/Lower Pacific Heights	3139	138023	12.4	68.2	26.2
Parkside/ Sunset District	1610	134652	5.9	61.5	74.5
Buena Vista Park	2853	174419	7.2	79	26.8
Inner Richmond/Richmond District	2965	139043	5.9	67.1	35.6
Outer Richmond	2000	116970	9.7	63.8	43.7
Marina	4000	194098	4.4	79.6	28.6
Bayview-Hunters Point	960	66618	18	56	51.1
Westwood Highlands/Twin Peaks West	1978	180768	4.5	62.3	81.1
Diamond Heights/Twin Peaks West	1885	181329	5.3	71.7	54.1
Stonestown	1475	93995	13	60.3	45.8
Marina District	1190	71063	14.6	60.5	16.8
Portola	1100	93068	11.4	61.3	61.9
Central Sunset/ Sunset District	1675	130708	8.4	65.2	50.4
Potrero Hill	1650	164289	7.8	76.7	37.5

Table 2: Socio-demographic data for San Francisco neighborhoods: Median housing prices are expressed in thousands of USD, while median incomes are stated in USD.

4 Empirical analysis

4.1 Predictive accuracy of the models

We begin the exposition of our empirical results by examining the accuracy of the trained models. We compare the accuracy of the gradient boosting machine with the linear model typically used in hedonic valuation models. Additionally, as a robustness check, we compare the gradient boosting machine to an alternative tree-ensemble model, the random forest (Breiman, 2001). A random forest averages the predictions of various trees trained on bootstrap samples of the original dataset. Furthermore, additional diversity in the individual trees is introduced by randomly selecting a subset of regressors at each split attempt during the tree growth process.

We employ several metrics to compare the predictive accuracy of the models. Common

accuracy metrics in regression problems include the root mean squared error (RMSE) and mean absolute error (MAE), defined as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}, \tag{14}$$

and

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|. \tag{15}$$

Due to the observed right-skewed distribution of the absolute residuals, we also take into account the median absolute error,

$$\text{MdAE} = \text{median}\left(\left|Y - \hat{Y}\right|\right). \tag{16}$$

Lastly, to account for the influence of the price scale on accuracy metrics, we also consider the scale-independent mean absolute percentage error,

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100, \tag{17}$$

and the R^2 metric,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}. \tag{18}$$

Note that the R^2 metric compares the sum of squared residuals of the model with that of a naïve model that always predicts the average property price, $\bar{Y}$.

model	**RMSE**	**MAE**	**MdAE**	**MAPE**	$\mathbf{R^2}$
Linear model	733.8	442.7	278.7	26.5	0.683
Random Forest	582.6	333.6	188.1	19.9	0.801
Gradient boosting machine	548.7	314.1	175.2	18.1	0.823

Table 3: Out-of-sample model accuracy given by a 10-fold cross-validation. The RMSE, MAE, and MdAE are measured in thousands of USD.

Table 3 presents the out-of-sample accuracy metrics obtained through a 10-fold cross-validation. The RMSE, MAE, and MdAE are expressed in thousands of USD. We observe that both tree-based ensembles outperform the linear model typically used in hedonic models. While the tree-based ensembles exhibit comparable accuracy, the gradient boosting machine holds a slight edge. The gradient boosting machine yields a mean absolute error of \$314k and a median absolute error, which is not influenced by large errors, of \$175k. Additionally, the mean absolute percentage error is 18%. These results indicate that there exists further heterogeneity in sales prices within the San Francisco real estate market, which is not fully captured by the extensive set of explanatory variables included in the dataset.

4.2 A look into the determinants of predictions

Based on the preceding findings, the gradient boosting machine emerges as the most effective model for predicting housing prices within our dataset. Now, we look deeper into this model to identify the variables that exert the most significant influence on its housing price predictions. One widely recognized methodology for assessing variable importance is SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017; Lundberg et al., 2018). SHAP relies on the concept of Shapley values, which originated within cooperative game theory. The Shapley value method distributes payoffs by quantifying the marginal contributions of each individuals within a group, allocating the payoff based on each player's respective contribution to the collective effort.

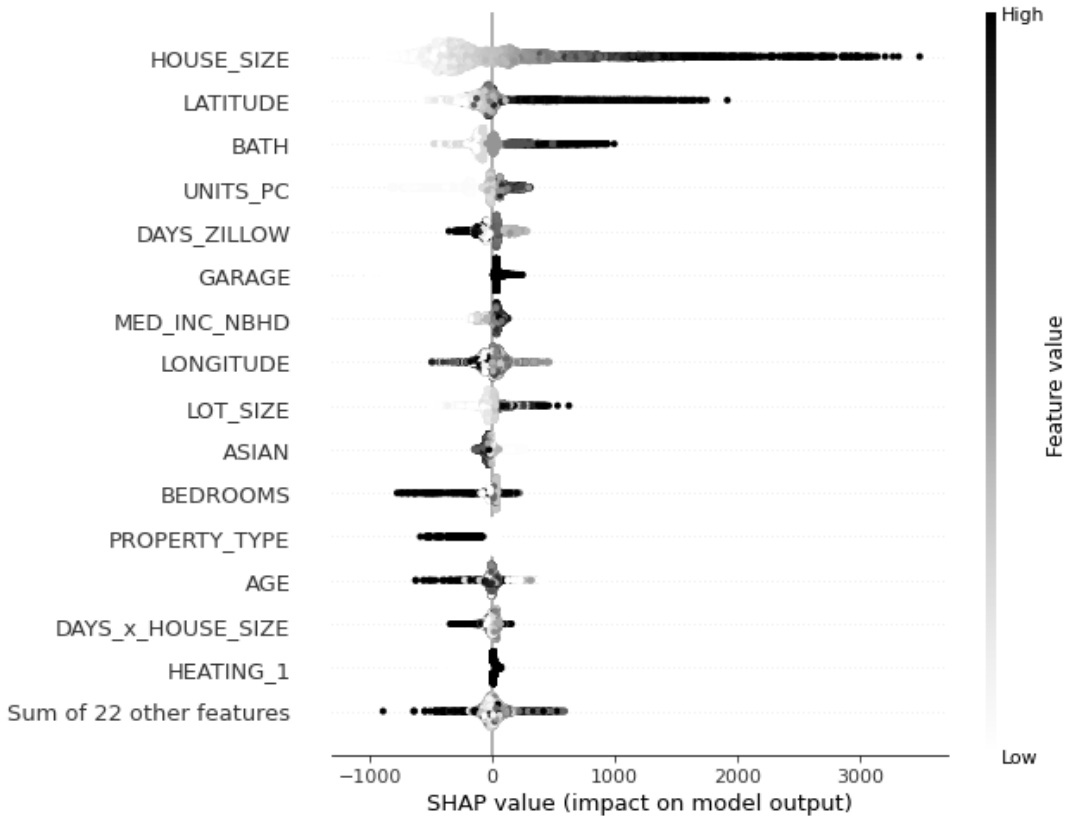

Figure 3: Beeswarm plot for the most influential features according to the SHAP methodology. The model is a gradient boosting machine trained to learn median house prices.

To conserve space, we just provide a brief overview of the methodology here; interested readers are directed to Molnar (2021) for a comprehensive introduction to the subject. The methodology assigns a SHAP value to each observation and explanatory variable, with variables having large absolute SHAP values considered important to the model. The global importance of a variable is determined by calculating the average of its absolute SHAP values across the dataset. SHAP values for individual observations are typically visualized using a beeswarm plot. Figure 3 depicts the beeswarm plot that we obtained for a gradient boosting machine trained to learn median house prices. In this plot, variables

are arranged according to their global importance. For instance, in our dataset, the house size emerges as the most influential variable in determining prices on average. The horizontal positions of the dots represent the SHAP values for individual observations. To aid in understanding the distribution of SHAP values, overlapping points are vertically offset. Large positive SHAP values indicate a substantial positive effect on the model's output, while large negative SHAP values indicate a significant negative effect. Different shades of gray are employed to depict the value of the explanatory variable – dark shades represent high values, while light shades indicate low values. Large house size values correspond to large and positive SHAP values, while small house size values correspond to negative SHAP values. This confirms the expected positive relationship between house size and selling price.

Latitude is the second most important variable, while longitude occupies the eighth position in importance. This aligns with the well-know emphasis on location by real estate brokers. In particular, we note a positive effect of latitude on prices, which is partly influenced by the affluent Richmond-Presidio-Marina area in northern San Francisco. Interestingly, the number of days on the market exhibits a negative effect on house price predictions. This suggests that if a house remains on the market for a long period, it is likely that the asking price will decrease over time. Moreover, a negative and seemingly counterintuitive relationship between the number of bedrooms and house price appears to exist. However, this relationship is primarily driven by a small number of very large houses with more than five bedrooms that were sold for lower prices. Once these large houses are excluded, a positive relationship between the number of bedrooms and selling price emerges.

4.3 A look into prediction intervals

After training the machine learning model, we can obtain conformal prediction intervals for the test data. To assess the adaptability of the prediction intervals, we examine how their width changes as a function of explanatory variables. In this section, we visualize how the prediction intervals behave as a function of some of the variables highlighted in the beeswarm plot in Figure 3. In the plots below, we considered a nominal coverage of 0.9.

Figure 4 depicts the conformal prediction intervals plotted against house size (in sq. ft.) for a random sample of 200 properties. The gray bars represent cases where the prediction interval failed to cover the actual value, while the dots represent actual property prices. With a nominal coverage of 0.9, the instances where the prediction interval fails to provide adequate coverage should account for approximately 10% of the plotted data points. The adaptability of the prediction intervals to model uncertainty is evident, as significant variations in the interval size can be observed across the house size range. The larger and more expensive the house, the wider the range of potential selling

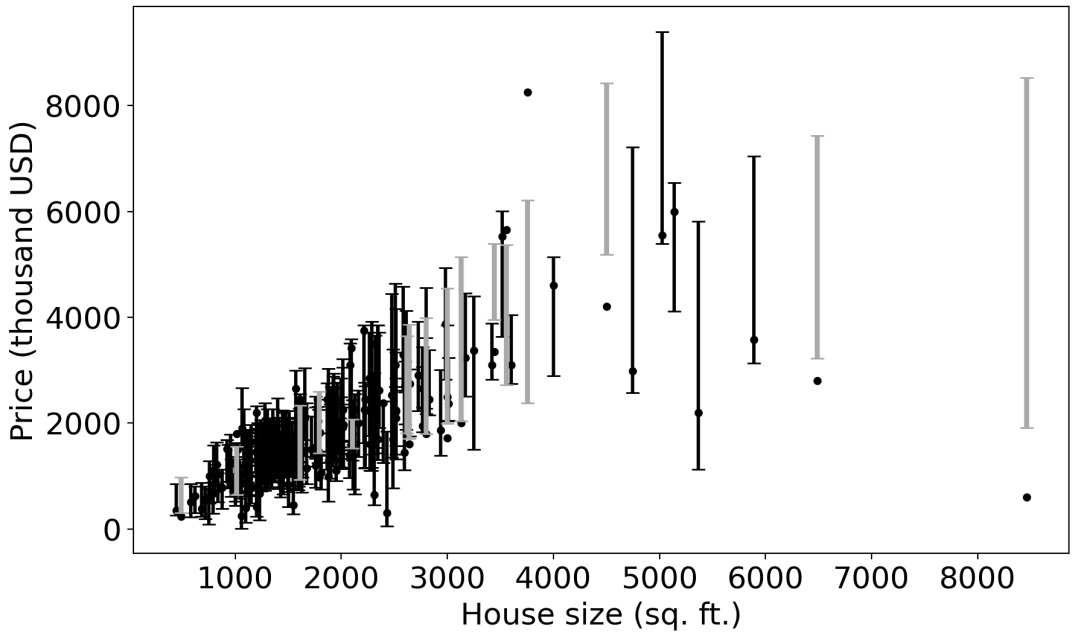

Figure 4: Conformal prediction intervals against house size (in sq. ft.) for a random sample of 200 properties. The bars in gray represent the cases where the prediction interval failed to cover the actual value. The dots represent actual property prices. The nominal coverage is 0.9.

prices.

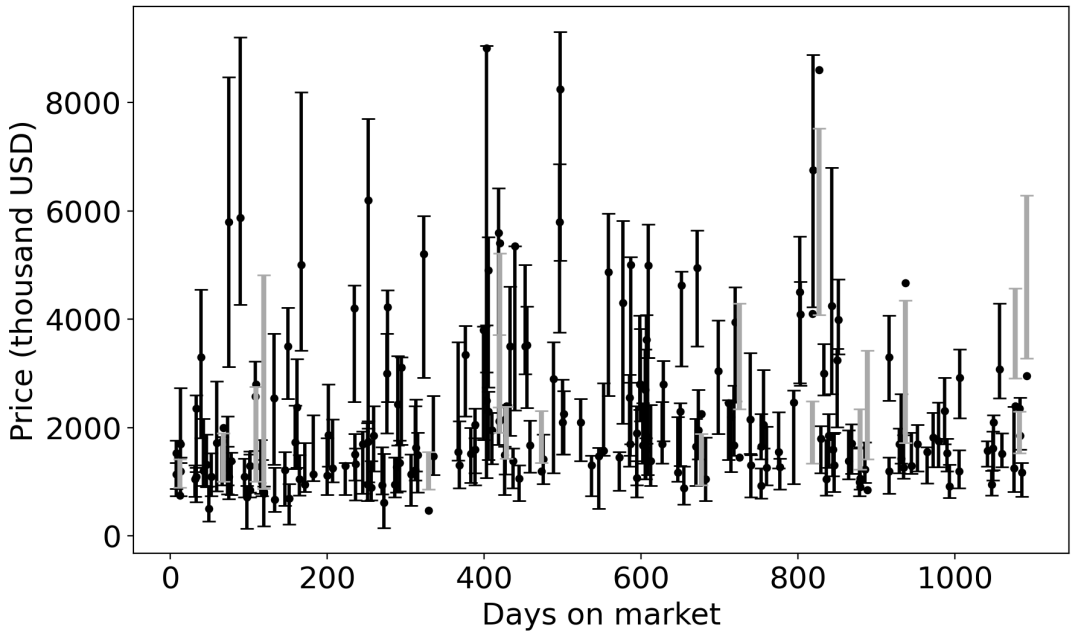

Figure 5: Conformal prediction intervals against the number of days the property was on the market for a random sample of 200 properties. The bars in gray represent the cases where the prediction interval failed to cover the actual value. The dots represent actual property prices. The nominal coverage is 0.9.

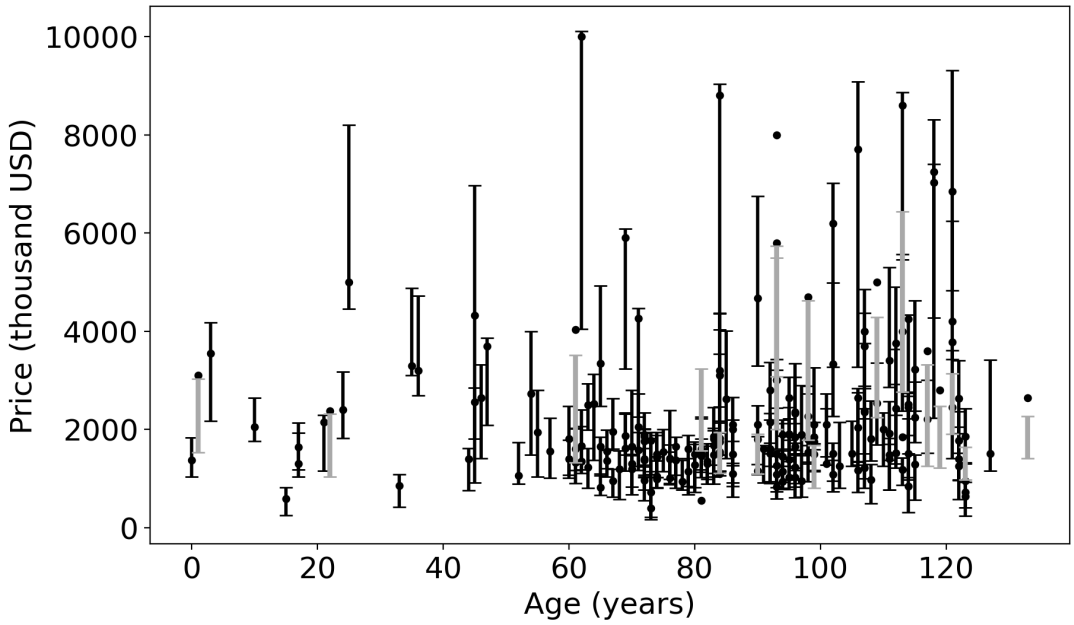

Figure 6: Conformal prediction intervals against the age of the property for a random sample of 200 properties. The bars in gray represent the cases where the prediction interval failed to cover the actual value. The dots represent actual property prices. The nominal coverage is 0.9.

Figures 5 and 6 display the conformal prediction intervals plotted against the number of days the property was on the market and the age of the property, respectively, for different random samples of 200 properties. In these cases, there is no strong pattern in the magnitude of the intervals.

It is evident from Figure 4 that larger properties, which typically command higher prices, correspond to wider prediction intervals. Therefore, we divide the absolute interval's width by the actual property price to obtain a relative measure of uncertainty. This can provide a more meaningful comparison of widths across properties with different price levels. Considering the limits of the conformal intervals in Equation 8, we have

$$\begin{aligned}\text{relative width} \equiv & \frac{\left[\hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X}) + q_{1-\alpha}(\mathcal{I}_2)\right] - \left[\hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X}) - q_{1-\alpha}(\mathcal{I}_2)\right]}{Y} \\ = & \frac{\hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X}) - \hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X}) + 2q_{1-\alpha}(\mathcal{I}_2)}{Y}.\end{aligned} \tag{19}$$

We note that, because the $(1-\alpha)$-quantile of the conformity scores, $q_{1-\alpha}(\mathcal{I}_2)$, is positive for any α, the width of the conformal intervals will always be larger than that of the non-conformal intervals $[\hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X}), \hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X})]$.

Figure 7 displays the distribution of relative interval widths. This distribution exhibits an asymmetric shape with positive skewness. The uncertainty for some predictions can be very high, reaching up to 5 times the actual selling price.

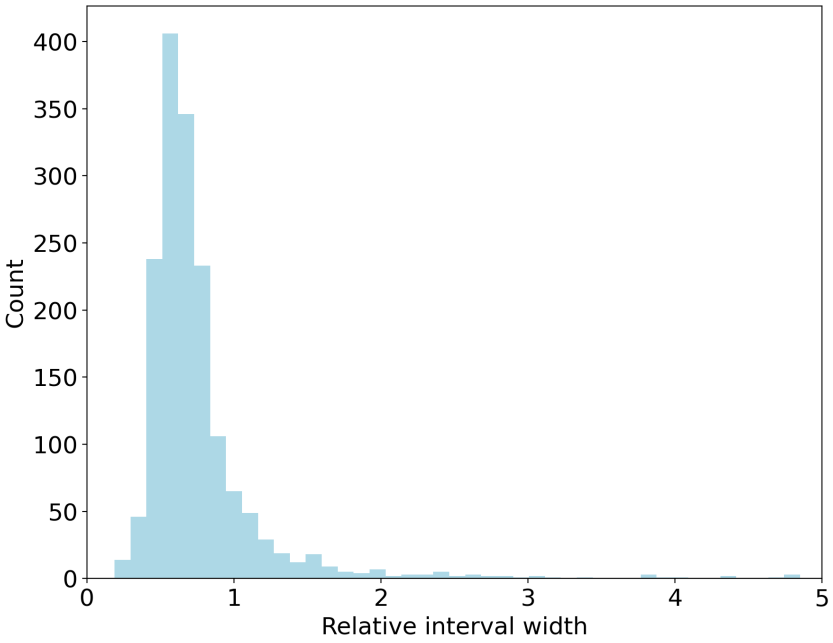

Figure 7: Distribution of relative interval widths for property prices in the test set.

4.4 Empirical coverage

To evaluate the reliability of the models' prediction intervals, we calculate the *empirical coverage*. This is defined as the proportion of actual property prices that fall within the prediction intervals. This proportion should be evaluated using an independent test set. Let n_{test} denote the number of observations in the test set, and $\mathcal{I}_{\text{test}}$ denote the set of indices of the observations in this set. The empirical coverage of the prediction intervals is

$$\frac{1}{n_{\text{test}}} \sum_{i \in \mathcal{I}_{\text{test}}} \mathbf{1}(Y_i \in \mathcal{C}(\mathbf{X}_i)), \tag{20}$$

where the indicator function $\mathbf{1}(\cdot)$ is equal to 1 if the actual selling price Y_i falls within the prediction interval $\mathcal{C}(\mathbf{X}_i)$, and 0 otherwise.

A real estate broker may also be interested in the width of the intervals as they reflect the uncertainty in the selling price. Since the distributions of relative interval widths in Figure 7 exhibit significant skewness, we report the median value of relative interval widths for the test set. This provides a comprehensive measure of interval widths that is less influenced by the extreme values observed in Figure 7 and captures the typical range of interval widths observed in the data.

However, to compute the empirical coverage and the median relative widths for the real data, we need to consider sampling variability. While conformal regression offers non-asymptotic coverage guarantees, there are two factors contributing to finite-sample variability. The first factor arises from the random partitioning of the original data into training and test sets. This affects the out-of-sample performance metrics of any predictive model, whether conformal or not. The second source of finite-sample variability

is due to the random selection of calibration samples from the training data. This source of variability only affects conformal prediction models, as calibration data is required to obtain sample conformity scores. Consequently, the prediction intervals conditioned on the calibration data become random variables. In Vovk (2012), it is shown that, given a training dataset $(\boldsymbol{X}_i, Y_i)_{i=1}^n$ and random calibration samples

$$\Pr\left(Y_{n+1} \in \mathcal{C}(\boldsymbol{X}_{n+1}) | \{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n\right) \sim \text{Beta}(n+1-k, k), \tag{21}$$

where $k = \lfloor (n+1)\alpha \rfloor$. To address the sampling variability, we conducted the relevant calculations 100 times using different random splits of the data into training, calibration, and test sets. By averaging the results, we mitigated the impact of finite-sample variability and obtained more reliable estimates for the empirical coverage and median relative widths.

Non-conformal quantile regression		
Nominal coverage	**Empirical coverage**	**Median relative width**
0.7	0.56	0.26
0.8	0.66	0.33
0.9	0.78	0.45
Conformal quantile regression		
Nominal coverage	**Empirical coverage**	**Median relative width**
0.7	0.70	0.35
0.8	0.80	0.45
0.9	0.90	0.62

Table 4: Empirical conditional coverages and median relative widths given by non-conformal quantile regressions (top panel) and conformal quantile regressions (bottom panel). Three nominal coverage levels are considered: 0.7, 0.8 and 0.9.

Table 4 reports the empirical coverage in the test data for three nominal coverage levels: 0.7, 0.8 and 0.9. The panel at the top refers to prediction intervals obtained from quantile regressions that are not conformalized, $[\hat{q}_{\frac{\alpha}{2}}(\boldsymbol{X}), \hat{q}_{1-\frac{\alpha}{2}}(\boldsymbol{X})]$. The panel at the bottom refers to prediction intervals obtained from quantile regressions that were conformalized according to Equation 8. These results show that non-conformalized quantile regressions generate prediction intervals that severely miss the actual selling prices regardless of the nominal coverage level. On the other hand, the conformal prediction intervals maintain exact coverage regardless of the nominal coverage level. That is, the conformalized models provide reliable finite-sample coverage guarantees. Regarding the median of relative interval widths, the non-conformal models provide narrower intervals than the conformal models. This was anticipated since the former seriously undercover the observed selling prices. On the other hand, the conformal models exhibit wider intervals, reflecting their ability to capture the uncertainty in predicting real estate prices with reliable guarantees.

4.5 Conditional coverage

The marginal coverage guarantee expressed in Equation 1 should be distinguished from the conditional coverage guarantee in Equation 3, which must be satisfied for any $\boldsymbol{x}$ and joint distribution of the data, $P_{\boldsymbol{X},Y}$. However, achieving conditional coverage guarantee is impossible without making strong assumptions about $P_{\boldsymbol{X},Y}$ (Barber et al., 2021). To obtain valid conditional coverages across different regions of the regressor space, it would be necessary to train individual models on data specific to each of these regions and conformalize the intervals accordingly. Nonetheless, it remains an interesting empirical question to explore the behaviour of conditional coverages in specific regions of the regressor space when the model is trained on the entire dataset. Indeed, the data may be too limited to effectively train individual models on subsets that cover small regions of a regressor domain. This might be the case of properties that are rare, such as Victorian buildings in San Francisco. A real estate broker might want a pricing model for these specific dwellings, but may not have enough training data for this task. Indeed, in our dataset, less than %5 of the properties were built before 1900.

If a regressor X is numeric, we can divide its range into a set of K mutually exclusive partitions, $\{\mathcal{X}_k\}_{k=1}^{K}$, and test if the coverage guarantee is well-approximated locally in all partitions:

$$\Pr\left(Y_{n+1} \in \mathcal{C}(\boldsymbol{X}_{n+1}) | \boldsymbol{X}_{n+1} \in \mathcal{X}_k\right)) \geq 1-\alpha, \;\; \forall\, k = 1, \ldots, K \tag{22}$$

If a regressor X is categorical, the condition in Equation 22 can be tested across individual categories or groups of categories. To evaluate the conditional coverage for a given variable, we divide its range into three mutually exclusive partitions using the 1/3- and 2/3-quantiles. We refer to these groups as the low-, medium-, and high-values groups. Our test data consists of 1650 observations, resulting in 550 observations in each partition. Afterwards, we calculate the approximate conditional coverage in each partition, along with the respective median relative widths.

Table 5 displays the approximate conditional coverages for selected variables: house size, latitude, lot size, days on market, median income, and age. Again, we consider three nominal coverage levels: 0.7, 0.8 and 0.9. The results show that the conditional coverage guarantee in Equation 22 is satisfied in most scenarios. However, in certain cases, the model is no longer calibrated. In other words, the intervals contain the actual prices more frequently than expected, making the empirical coverage exceed the given nominal target coverage. For high values of house size, latitude, and lot size, we observe some undercoverage. However, the distortion in the empirical coverage observed in these cases is not as serious as that observed in the non-conformalized models. As for the median widths, we observe relatively small variations across the partitions, in that the factor that influences the empirical widths the most is the nominal coverage. Still, we observe some adaptability of the interval widths across the different partitions. For instance, the predictions for large houses have higher median relative widths compared to medium-

Nominal coverage	**Empirical coverage**			**Median relative width**		
	low	medium	high	low	medium	high
			House size			
0.7	0.74	0.73	0.64	0.35	0.34	0.36
0.8	0.84	0.83	0.74	0.45	0.44	0.46
0.9	0.93	0.92	0.85	0.64	0.59	0.64
			Latitude			
0.7	0.74	0.72	0.65	0.34	0.35	0.36
0.8	0.84	0.81	0.75	0.44	0.45	0.46
0.9	0.93	0.91	0.86	0.62	0.62	0.64
			Lot size			
0.7	0.71	0.72	0.67	0.35	0.35	0.34
0.8	0.80	0.82	0.78	0.45	0.45	0.44
0.9	0.91	0.91	0.88	0.63	0.63	0.62
			Days on market			
0.7	0.70	0.71	0.69	0.38	0.35	0.32
0.8	0.80	0.80	0.79	0.49	0.45	0.41
0.9	0.90	0.90	0.90	0.68	0.62	0.57
			Median income			
0.7	0.73	0.70	0.68	0.35	0.33	0.36
0.8	0.83	0.80	0.78	0.45	0.43	0.47
0.9	0.92	0.90	0.88	0.63	0.59	0.64
			Age			
0.7	0.70	0.74	0.67	0.34	0.34	0.38
0.8	0.80	0.83	0.78	0.43	0.44	0.49
0.9	0.90	0.92	0.88	0.59	0.60	0.67

Table 5: Approximate conditional coverages and median relative widths measured in three different partitions of explanatory variables corresponding to low, medium and high values of these variables. Three nominal coverage levels are considered: 0.7, 0.8 and 0.9.

sized houses. This increased uncertainty is not due to overcoverage, as the empirical coverage for large houses is actually smaller than the nominal level. Interestingly, pricing houses that are in their first year on the market (the 'low' category) proves to be more challenging. Once more, this difficulty does not arise from miscoverage, as the conditional coverages closely match the nominal coverages. Also, predictions for property prices in low and high income neighborhoods exhibit greater uncertainty compared to those for medium income neighborhoods. The increased uncertainty at both ends of the income distribution can be attributed to the significantly different variables influencing property prices at the opposite ends of San Francisco's real estate market. Furthermore, we observe that predicting the prices of older properties, typically those with an age greater than 100

years, present additional challenges. The non-trivial relationship between a property's age and its market value introduces uncertainty. Indeed, older properties tend to undergo renovations over the years, a factor not readily accounted for in our dataset. This lack of information increases the difficulty in accurately evaluating older properties.

5 Conclusion

This study presented some results regarding the uncertainty of predictions given by a real estate appraisal model. Our tool to achieve this goal was conformal quantile prediction, a model-agnostic procedure that constructs prediction intervals, ensuring valid coverage in finite samples without relying on distributional assumptions. Through an empirical application to property prices of the San Francisco Bay Area, we observed that conformal quantile prediction, consistently delivers coverage guarantees close to the desired nominal level. On the other hand, intervals based on non-conformal quantiles severely undercover the data. Also, we showed that the intervals are adaptive to the variations in the data, which is crucial considering the heterogeneous nature of real estate data. Indeed, we found that larger properties, older properties, those in both low and high income neighborhoods, and those on the market for less than one year are more challenging to evaluate. The ability to flexibly adjust the width of prediction intervals to accommodate varying levels of uncertainty and complexity within the dataset is a significant advantage. It enables real estate professionals to make informed decisions across a wide spectrum of property types, locations, and market conditions, thereby enhancing the practical utility of automated valuation models.

Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia [grant number UIDB/05069/2020].

References

Angelopoulos, A.N., Bates, S. (2023). Conformal prediction: A gentle introduction. Foundations and Trends in Machine Learning 16(4), 494–591.

Barber, R.F., Candès, E. J., Ramdas, A., and Tibshirani, R.J. (2021). The limits of distribution-free conditional predictive inference. Information and Inference: A Journal of the IMA, 10(2), 455–482.

Bellotti, A. 2017. Reliable region predictions for automated valuation models. Annals of Mathematics and Artificial Intelligence, 81, 71–84.

Breiman, L. (2001). Random Forests. Machine Learning, 45, 5–32.

Chen, T., Guestrin, E. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. San Francisco, USA.

Curth, A., Jeffares A., van der Schaar, M. (2024). Why do random forests work? Understanding tree ensembles as self-regularizing adaptive smoothers. arXiv:2402.01502v1.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of Statistics 29, 1189-1232.

Goodman, A. C. (1978) Hedonic prices, price indices and housing markets. Journal of Urban Economics, 5, 471–484.

Grinsztajn, L., Oyallon, E., Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? In Proceedings of NeurIPS 2022 – Neural Information Processing Systems. New Orleans, USA.

Guan, J., Zurada, J., Levitan, A.S. (2011). A comparison of regression and artificial intelligence methods in a mass appraisal context. Journal of Real Estate Research, 33(3), 349–387.

Hjort, A., Pensar, J., Scheel, I., Sommervoll, D.E. (2022). House price prediction with gradient boosted trees under loss functions. Journal of Property Research, 39(4), 338–364.

Ho, W. K., Tang, B.-S., Wong, S. W. (2021). Predicting property prices with machine learning algorithms. Journal of Property Research, 38(1), pp. 48-70.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the NeurIPS 2017 – Neural Information Processing Systems. Long Beach, USA.

Koenker, R., Bassett Jr., G, (1978). Regression Quantiles. Econometrica 46(1), 33-50.

Lei, J., Robins, J., Wasserman, L., (2013). Distribution-free prediction sets. Journal of the American Statistical Association, 108, 278–287.

Lei, J., Wasserman, L. (2014). Distribution-free prediction bands for non-parametric regression. Journal of the Royal Statistical Society B, 76, 71–96.

Lim, Z., Bellotti, A. (2021). Normalized nonconformity measures for automated valuation models. Expert Systems with Applications, 180, 115–165.

Lundberg, S.M., Lee, S.-I. (2017). A unified approach to interpreting model predictions, NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 4768–4777.

Lundberg, S.M., Erion, G.G., Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles, arXiv:1802.03888.

Meinshausen, N. (2006). Quantile regression forests. Journal of Machine Learning Research 7, 983–999.

Molnar, C. (2021). Interpretable Machine Learning. A Guide for Making Black Box Models Explainable.

Papadopoulos, G., Edwards, P. J., Murray, A. F., Confidence estimation methods for neural networks: a practical comparison. IEEE Transactions in Neural Networks, 12, 1278–87.

Papadopoulos, H., Proedrou, K., Vovk, V., and Gammerman, A. (2002). Inductive confidence machines for regression. In Proceedings of Machine Learning: European Conference on Machine Learning, 2002, 345–356. Helsinki, Finland.

Papadopoulos, H., Gammerman, A., Vovk, V. (2008) Normalized nonconformity measures for regression conformal prediction. In Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications, 64–69. Innsbruck, Austria.

Peterson, S., Flanagan, A.B., 2009. Neural network hedonic pricing models in mass real estate appraisal. Journal of Real Estate Research, 31(2), 147–64.

Romano, Y., Patterson, E., Candès, E. (2019) Conformalized quantile regression. In Advances in Neural Information Processing Systems, vol. 32, 3543–3553. Vancouver, Canada.

Taylor, J. W. (2000) A quantile regression neural network approach to estimating the conditional density of multiperiod returns 19, 299–311.

Vovk, V., Gammerman, A., Shafer, and G. (2005). Algorithmic Learning in a Random World. Springer.

Vovk, V., Nouretdinov, I., Gammerman, A. (2009) On-line predictive linear regression. Annals of Statistcs 37(3), 1566–1590.

Vovk, V. (2012). Conditional validity of inductive conformal predictors. In Proceedings of the Asian Conference on Machine Learning, 25, pp. 475–490. Singapore.

Wager, S., Hastie, T., Efron B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. Journal of Machine Learning Research, 15, 1625–1651.

Zhou. K.Q., Portnoy, S.L. (1996) Direct use of regression quantiles to construct confidence sets in linear models. Annals of Statistics 24, 287–306.