Affine Trees as Data Generating Processes*

Gabrielle Gayer†and Itzhak Gilboa‡

June 28, 2023

Abstract

An affine tree is a binary decision tree whose leaves contain linear regression equations. It is suggested as a model of the way economic agents reason by rules, combining logical conditions with algebraic formulas. Finding optimal affine trees is computationally hard, and we, therefore, focus on trees with a single bifurcation. We estimate them on real estate data and show that they perform better than hedonic regression. A by-product of the analysis is the finding that rule-based models, (including standard OLS), fit sales prices better than rental prices, indicating a stronger influence of rule-based thinking in speculative trade compared to consumption decisions.

1 Introduction

Hedonic regression is a basic tool for the evaluation of real estate prices. It suggests a simple algebraic rule for the assessment of a property's price based on a set of observable variables, and it lends itself to a natural interpretation as a model of the way agents think about prices. However, the rules economic agents employ in order to evaluate assets are not restricted to algebraic formulas. People often reason according to logical conditions ("if-then-else") as in association rules and other machine learning techniques. We suggest that logical conditions and linear formulas can be combined to generate a useful model of the way economic agents

*We are very grateful to Offer Lieberman for his invaluable help. We gratefully acknowledge support from the ISF Grant 1443/20, as well as from Investissements d'Avenir ANR -11- IDEX-0003 / Labex ECODEC No. ANR - 11-LABX-0047 as well as the AXA Chair for Decision Sciences at HEC, and the Foerder Institute at Tel-Aviv University (Gilboa)

†Department of Economics, Bar-Ilan University. email: gabi.gayer@gmail.com

‡Tel Aviv University and HEC, Paris tzachigilboa@gmail.com

assess asset prices. For example, people can draw a qualitative distinction between geographical areas such as a city's center and its suburbs, and assess prices per square foot separately for these two. Alternatively, they may think differently about "new" and "old" buildings, using a cutoff (such as WWII) to separate them, and then apply different hedonic regressions to the two subsets.

In this paper we assume that the data generating process is of this type. More explicitly, there are m real-valued predictors $x^1, ..., x^m$ used for predicting a real-valued y. For an index $j' \leq m$ and a cutoff $c \in (-\infty, \infty]$, the data generating process is defined by

$$y_i = \begin{cases} \beta_0 + \sum_{j=1}^{m} \beta_j x_i^j + \varepsilon_i & x^{j'} \leq c \\ \gamma_0 + \sum_{j=1}^{m} \gamma_j x_i^j + \varepsilon_i & x^{j'} > c \end{cases} \tag{1}$$

where $(\beta_0, \beta_1, ..., \beta_m), (\gamma_0, \gamma_1, ..., \gamma_m) \in \mathbb{R}^{m+1}$ are regression coefficients and ε_i are i.i.d. random variables with zero expectation. (The case $c = \infty$ is included to allow for a simple regression model, with no bifurcation, as a special case.) Such models have been discussed in the literature where the identity of the branching variable, j', is known, and the cutoff c is estimated from the data (alongside $(\beta_0, \beta_1, ..., \beta_m), (\gamma_0, \gamma_1, ..., \gamma_m)$ and the variance of ε_i).[1] For example, Andrews (1993) assumes that the branching variable is time, and estimates the cutoff c. By contrast, in our model j' also needs to be estimated from the data, in conjunction with the aforementioned parameters (see also Hansen, 1999, 2000, and, 2017)

One may view our model as a rather special case of a "decision tree", where multiple binary bifurcations are allowed, each defined by the condition $x^{j'} \leq c$ for some $j' \leq m$ and $c \in \mathbb{R}$. We refer to such a tree as *an affine tree*.[2]. Decision trees have been used in machine learning for several decades by now (for a recent survey, see Sharma and Kumar, 2016). In recent years they have also been adopted by empirical research in economics. For a recent survey of the literature the reader is referred to Varian (2014) and Mullainathan and Spiess (2017). However, to the best of our knowledge, they have not been used as a model of the way economic agents think or evaluate numerical magnitudes.

Thinking of an affine tree as a psychological model of the way people think, one may wonder how complex they can be. Indeed, we prove that finding the best

[1] One may consider a branching variable that is not one of the m variables used in the regression equations. Alternatively, if we allow the model to have some of the coefficients $(\beta_j, \gamma_j)_j$ to be restricted to zero, the branching variable may be included in the m variables we start out with.

[2] The term "decision tree" is most often used for a classification problem, where y is binary. A continuous y is often referred to as a "regression tree". By contrast, we allow for a linear regression function at each leaf, and thus opt for the name "affine tree".

affine tree (in terms of best-fit and tree size) is an NP-Complete problem. Thus, our focus on simple (single-bifurcation) trees as in (1) follows from our psychological interpretation of the model. We applied our model to real estate data from various cities in South America, including Sao Paulo, Buenos Aires, Bogota, Lima, Quito, and Montevideo, which are available online. We estimated different affine trees that correspond to different bifurcation variables, and demonstrated that these estimated affine trees outperformed the simple hedonic regression based on AIC and other measures. Moreover, we conducted a bootstrap analysis, which revealed that the hypothesis that the data are generated by a (no-bifurcation) hedonic regression can be rejected in favor of the alternative of a (single-bifurcation) tree, which holds for all bifurcation variables. We take these findings as evidence that the introduction of bifurcations into hedonic regression models may result in better models despite the added complexity. Finally, we contrast the analysis for sales data with a comparable analysis of rental prices. We find that, with bifurcation trees as well as with simple hedonic regression, the model performs better on sales than on rent data, and we offer an interpretation of this qualitative result.

The rest of this paper is organized as follows. Section 2 defines the generalization of (1) to affine trees and provides the complexity result regarding optimal trees. Section 3 applies the method to the real estate data, estimating single-bifurcation trees and contrasting them with hedonic regression. Section 4 compares the analyses for sales and rental prices. Section 5 concludes with a discussion.

2 General Model

We consider real-valued functions $y = f(x^1, ..., x^m)$ of real-valued variables $(x^1, ..., x^m)$ defined as follows. An *affine tree* is a tuple $Tr \equiv (V, E, r, L, b, c, p, d)$, where

– V is a non-empty and finite set of nodes;

– $E \subset V \times V$ is a set of directed edges defining a graph $G(V, E)$ that is a tree;

– $r \in V$ is the root and $L \subset V$ is the set of leaves of $G(V, E)$;

– $b : V \backslash L \to \{1, ..., m\}$ determines according to which variable a bifurcation is made at each non-terminal node ($v \in V \backslash L$);

– $c : V \backslash L \to \mathbb{R}$ determines the cutoff value at such a node ($v \in V \backslash L$);

– $p : V \backslash L \times \{0, 1\} \to V$ is the bifurcation function, such that for every $v \in V \backslash L$ there are exactly two edges $(v, p(v, 0)), (v, p(v, 1)) \in E$;

– $d : L \to \mathbb{R}^{m+1}$ assigns to each leaf $v \in L$ a vector $\theta = (\beta_0, \beta_1, ..., \beta_m) \in \mathbb{R}^{m+1}$

which is used to define an affine function of $x^1, ..., x^m$,

$$y\left(x^1, ..., x^m\right) = \beta_0 + \sum_{j=1}^{m} \beta_j x^j.$$

A *computation* of an affine tree (V, E, r, L, b, c, p, d) for input $x = (x^1, ..., x^m) \in \mathbb{R}^m$ is a pair $((v_1, ..., v_k), y)$ such that

(i) $(v_1, ..., v_k) \in V^k$ is a path from the root to a leaf: $v_1 = r$; $(v_l, v_{l+1}) \in E$ for every $l < k$, and $v_k \in L$;

(ii) the path follows the bifurcations dictated by $(x^1, ..., x^m)$: for every $l < k$, $v_{l+1} = p(v_l, 0)$ if $x^{b(v_l)} \leq c(v_l)$ and $v_{l+1} = p(v_l, 1)$ if $x^{b(v_l)} > c(v_l)$;

(iii) y is the value defined by the function at node v_k, that is $y = \beta_0 + \sum_{j=1}^{m} \beta_j x^j$ where $(\beta_0, \beta_1, ..., \beta_m) = d(v_k)$.

That is, the computation path begins at the root, and considers the variable x^j for $j = b(r)$. If $x^j \leq c(v_1)$, the path continues to $v_2 = p(v_1, 0)$, and if $x_j^j > c(v_1)$ – to $v_2 = p(v_1, 1)$. It then continues to consider $j' = b(v_2)$ and proceeds inductively until it reaches a node $v_k \in L$. Let $f_{Tr} = f(V, E, r, L, b, c, p, d)$ be the function (from $\mathbb{R}^m$ to $\mathbb{R}$) defined by the computation of the tree $Tr \equiv (V, E, r, L, b, c, p, d)$.

A *database* consists of observations of $x = (x^1, ..., x^m)$ and y. More formally, a *database* is a tuple $B = ((x_i, y_i))_{1 \leq i \leq t}$ (where $x_i = (x_i^1, ..., x_i^m) \in \mathbb{R}^m$, and $y_i \in \mathbb{R}$). The statistical model we are interested in involves data generating processes that are based on affine trees. Specifically, we will assume that there exists a true underlying function tree Tr, such that, for each $t \geq 1$,

$$y_t = f_{Tr}(x_t) + \varepsilon_t$$

where $(\varepsilon_t)_t$ are i.i.d. random variables with zero expectation. We assume that economic agents attempt to understand the reality they are faced with by trying to fit the "best" affine tree to a given database. The degree to which a given tree $Tr = Tr(V, E, r, L, b, c, p, d)$ fits the database $B = ((x_i, y_i))_{1 \leq i \leq t}$ is measures by the sum of squared errors:

$$SSE(B, Tr) = \sum_{i=1}^{t} \left(y_i - f_{Tr}(x_i)\right)^2$$

We will also assume that agents trade off goodness of fit with complexity. That is, they tend to prefer simpler trees over more complex ones. We consider the following two ways to measure this complexity:

(i) The number of variables used is the tree, $NV(Tr) \equiv |Im(b)|$

(ii) The number of decision nodes, $ND(Tr) \equiv |V \backslash L|$.

Note that, for every tree $Tr(V, E, r, L, b, c, p, d)$ we have

$$\begin{aligned} |V| &= 2ND(Tr) + 1 \\ NV(Tr) &\leq ND(Tr) \end{aligned}$$

A low $NV(Tr)$ would be desirable, other things being equal, because it implies that the size of the database one needs to recall and process is small. By contrast, a low $ND(Tr)$ means that the entire tree is relatively easy to recall and describe. Thus, there are reasons to assume that, other things being equal, people tend to prefer simpler trees according to each of these measures. It stands to reason that, in reality, people tend to trade off these three notions of complexity, as well as to trade off simplicity with goodness of fit. To capture this tradeoff we formulate the following problems.

Problem 1 *MIN-NV Given a database with rational values, $B = ((x_i, y_i))_{i \leq t}$, an integer $k \geq 1$ and a rational $a \geq 0$, is there a function tree $Tr(V, E, r, L, b, c, p, d)$, with $|Im(b)| \leq k$ and $SSE(B, Tr) \leq a$?*

Problem 2 *MIN-ND Given a database with rational values, $B = ((x_i, y_i))_{i \leq t}$, an integer $k \geq 1$ and a rational $a \geq 0$, is there a function tree $Tr(V, E, r, L, b, c, p, d)$, with $|V \backslash L| \leq k$ and $SSE(B, Tr) \leq a$?*

We can state[3]

Theorem 1 *MIN-NV and MIN-ND are NPC.*

This result suggests that we wish to focus on small trees. A statistician who needs to compute maximum-likelihood trees within a class (defined by NV or ND), faces a computationally complex problem. Should the tree be too large, it will be impractical to find the best one. But, more importantly, one might argue that, for the same reason, people are unlikely to be using large trees. Apart from the cognitive costs associated with remembering the tree and implementing its computation, larger trees are also less likely to be the optimal ones in their respective classes.

[3]Many problems related to minimal decision trees are known to be NPC. However, we are unaware of a proof of the result regarding affine trees. Note that, while affine trees are much richer than decision or regression trees, it does not automatically follow that finding the optimal tree in a given class is more complex than finding the optimal tree in a subclass thereof. Specifically, the continuous and algebraic structure might simplify the optimization problem (as in the famous case of linear programming).

3 Application

3.1 The Data

We downloaded a database of real estate properties from Kaggle, an open-source platform that allows users to freely access various databases. It was comprised of listings of apartments for rent and sale in Sao Paolo, Brazil that were gathered from multiple real estate websites during April 2019.[4]

The Sao Paulo data included information about the advertised price, condominium fees, exact location (latitude and longitude), size, number of rooms, and other characteristics such as number of parking spots, and number of toilets. We created an additional variable that measured the distance of the apartment from Ana Rosa terminal, which is located in the city center. There were 13,640 observations in the dataset, of which approximately 4,000 were omitted due to duplication, missing information regarding their precise location or condominium fees, or because they were outliers. The complete list of variables and their descriptive statistics appear in Appendix B in Table 4.

We also obtained 5 more databases from Kaggle on real estate property listings in Argentina, Colombia, Ecuador, Peru, and Uruguay for 2019.[5,6] We focused on apartments for rent and sale in the capital cities, Buenos Aires, Bogota, Lima, Quito, and Montevideo. The databases included information on prices, size, location (latitude and longitude), and number of rooms and bathrooms. Many observations had missing or extreme values and therefore were omitted. Nevertheless, there still remained a significant number of observations in each database (as seen in Appendix B, Table 3 and the descriptive statisticsc in Tables 5-9).

3.2 Estimation

The empirical application of apartments in Sao Paulo was aimed at understanding if the data were generated by a hedonic regression or rather could they be created by a decision tree. For this purpose, we ran five bifurcation regressions, each corresponding to a different continuous variable that acted as the bifurcation variable. These variables included latitude, longitude, distance from Ana Rosa station, size and condominium fees. The linear regressions contained the entire set of attributes, which were the same for both the decision tree models and the hedonic regression.

[4] https://www.kaggle.com/datasets/argonalyst/sao-paulo-real-estate-sale-rent-april-2019

[5] https://www.kaggle.com/datasets/rmjacobsen/property-listings-for-5-south-american-countries

[6] We thank Properati Data for allowing free access to their databases through Kaggle.

We divided the data into apartments for sale (4,297 observations) and apartments for rent (5,458 observations). These databases were randomly split into training data containing 80% of the observations (3,437 apartments for sale and 4,366 rental apartments) and test data (the remaining 20% of the observations containing 867 apartments for sale and 1091 rental apartments). The parameters of the single bifurcation model in Equation (1), which included the coefficients of the two linear regressions in the branches of the tree, σ^2 (which was assumed to be the same in both branches), as well as the cutoff value of the bifurcation variable, were estimated on the training data. These estimators were then used to predict the price/rent of the apartments in the test data in order to determine how well the model fits the data. The estimation of the coefficients in the leaves of the single bifurcation model were computed using the least squares method for each specified cutoff value of the bifurcation variable. Given these estimators, the cutoff value was chosen to minimize the total sum of squared residuals (SSR) in the leaves of the tree ($SSR = \sum_{i=1}^{t}(y_i - \hat{y}_i)^2$, where $\hat{y}_i$ is the predicted value of apartment i conditional on the cutoff value of the specified bifurcation variable). In addition, we ran a hedonic regression with no bifurcation using the same set of independent variables as we did for the estimations of the single bifurcation regressions. The estimators of the single bifurcation models, as well as those of the no bifurcation model can be found in Appendix B in Tables 10 and 11.

The performance of the hedonic regression was compared to that of the single bifurcation regressions using various measures. The results are reported in Table 1, which contains the mean squared error (MSE), the adjusted R^2, and the values of the Akaike (AIC) and Schwartz (BIC) criteria for the training database. We also report the MSE and the adjusted R^2 that were computed on the test database. As shown in Table 1, distance from Ana Rosa is the best bifurcation variable according to all measures, both for apartments for rent and apartments for sale. It achieves the highest adjusted R^2 and the lowest MSE, AIC, and BIC criteria in the training database, as well as the lowest MSE in the test data compared to any other bifurcation variable. The opposite is true for latitude, which is the worst bifurcation variable according to these same measures. However, according to all training and test measures, even the worst single bifurcation variable performs better than the hedonic regression. It reaches a higher adjusted R^2 than the hedonic regression while maintaining a lower MSE for both training and test data, as well as lower AC and BIC values.

Table 1: Sao Paulo– Measures of goodness of fit of different models

	Model	Cutoff value	Train						Test			
			MSE	Adj. R^2	# obs. branch 1	% obs. branch 1	AIC	BIC	MSE	Adj. R^2	# obs. branch 1	% obs. in branch 1
Sale	Ana Rosa dist	0.07	1.32E+10	0.85	1,062	30.9	89,897	90,056	1.46E+10	0.84	270	31.4
	Size	70.01	1.51E+10	0.83	2,288	66.6	90,363	90,523	1.63E+10	0.83	542	63.0
	Condo	797.24	1.48E+10	0.83	2,730	79.4	90,290	90,450	1.56E+10	0.83	662	77.0
	Latitude	-23.53	1.69E+10	0.81	2,331	67.8	90,743	90,903	1.84E+10	0.80	596	69.3
	Longitude	-46.64	1.53E+10	0.83	1,578	45.9	90,398	90,557	1.71E+10	0.82	421	49.0
	No cutoff		1.81E+10	0.80	3,437	100.0	90,949	91,029	1.92E+10	0.80	860	100.0
Rent	Ana Rosa dist	0.05	7.64E+05	0.70	877	20.1	71,583	71,743	7.32E+05	0.71	212	19.4
	Size	88.03	9.07E+05	0.65	3,183	72.9	72,333	72,493	8.73E+05	0.66	817	74.8
	Condo	542.18	8.86E+05	0.66	1,750	40.1	72,230	72,396	8.66E+05	0.66	466	42.7
	Latitude	-23.55	9.39E+05	0.64	2,626	60.1	72,484	72,650	9.08E+05	0.65	633	58.0
	Longitude	-46.65	8.60E+05	0.67	2,215	50.7	72,103	72,269	8.22E+05	0.68	519	47.5
	No cutoff		1.00E+06	0.61	4,366	100.0	72,744	72,827	9.75E+05	0.62	1,092	100.0

We performed the same analyses on the five additional databases. The results for the training data of these five databases showed similar trends as the main analysis, with a single bifurcation model fitting the data better than a model with no bifurcation (except for the BIC measure of Size which is slightly greater than the BIC value of no bifurcation for rental apartments in Quito). However, the results for test data were slightly more mixed, with the adjusted R^2 of the no bifurcation model being higher than that of some of the single bifurcation models for rental databases in Lima, and Quito. The results are summarized in Appendix B (Tables 22-26).

We use the bootstrap procedure to test whether the results were statistically significant.The null hypothesis, under which the model with no bifurcation is true, is contrasted with an alternative hypothesis that supposes that a single bifurcation model is correct. To test the hypotheses we examined the normalized log likelihood ratio, which under Gaussianity equals $LR = n^{0.5} \log \left(\frac{SSR(H_1)}{SSR(H_0)} \right)$. A high LR value indicates that H_0 is the correct model since the residuals of the null hypothesis are notably smaller than those of the alternative. We used the residual bootstrapping procedure to test the null hypothesis H_0 against H_1. We created 1,000 bootstrap LR estimates by resampling the residuals that were calculated under the hypothesis that H_0 is true. The LR in each resample (LR^b) is computed by re-estimating the parameters of models H_0 and H_1 on the bootstrap sample.

We reject H_0 for a small bootstrap p-value, where the p-value equals the frequency of resamples wherein $LR^b < LR$. In Sao Paulo all five p-values, corresponding to five alternative bifurcation models (Ana Rosa, Size, Condo, Latitude, and Longitude), equal zero, showing that the model with no bifurcation is refuted when compared to any of the single bifurcation models. These findings hold true for the five additional databases as well, where the no bifurcation model is uniformly rejected when contrasted with to each of the single bifurcation models (Latitude, Longitude, and

Size)[7] These results are consistent with the conjecture that people use decision trees to evaluate the price and rent of apartments.

4 Sales vs. Rents: Coordination on a Rule

Hedonic regression is a popular model for assessing not only sale but also rental prices. How well does it predict values in the two different types of markets? We conjecture that, other things being equal, hedonic regression would explain a larger portion of the variance in the sales market as compared to the rental one. That is, consider two regressions

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_i^j + \varepsilon_i \qquad i \leq n \qquad \text{(I)} \tag{2}$$

$$z_k = \beta_0 + \sum_{j=1}^{m} \beta_j x_k^j + \delta_k \qquad k \leq r \qquad \text{(II)} \tag{3}$$

with the same set of predictors $x = (x^1, ...x^m)$, where y are prices of properties and z are rents on different properties in the same locality and the same period. We predict that the R^2 of (I) would be higher than that of (II).

The logic behind this conjecture is the following: renting a property is, by and large, a pure consumption decision: renters should ask themselves how much housing rights are worth to them, and they do not need to bother about other agents' valuation of the rights they buy. By contrast, buyers of a property typically consider it as an investment, and would therefore need to think how much others in the market would be willing to pay for it.[8] In other words, when a rational agent asks herself what her reservation price for renting an apartment is, she faces a single-person decision problem, and she can allow her idiosyncratic taste to affect her answer. But when she asks the same question for buying an apartment, she realizes that she is a player in a game that involves speculative trade. In this case the market coordinates on a price for each asset, which becomes the equilibrium price. We suggest that, when there is an aspect of speculative trade, simple rules would explain more of the

[7] We also checked the converse premise where the roles of the no bifurcation model and the single bifurcation model were reversed. None of the single bifurcation models are refuted when supposing that the alternative model is the no bifurcation model.

[8] Renters might be thinking of subletting the property they rented, in which case the difference between renting and buying would be less pronounced.

variance than when such trade is absent. Simple rules are easy to coordinate on, and they can therefore serve as focal points (Schelling, 1960). In the absence of resale opportunities, a rational agent need not worry about others' valuations, and may deviate from the general rule at no economic cost.[9]

Following this intuition, we analyzed the six databases mentioned above of asking prices for both sales and rental prices that were obtained from the same localities at the same time, and we found

Table 2: Adjusted R^2– Sales vs. Rent

	Sales	**Rent**	**Ratio**
Sao Paulo	0.80	0.61	1.30
Buenos Aires	0.62	0.46	1.34
Bogota	0.73	0.73	1.00
Lima	0.71	0.58	1.22
Quito	0.51	0.14	3.77
Montevideo	0.68	0.52	1.30

Thus, our conjecture seems to be supported by the data: in 5 out of 6 pairs of databases the adjusted R^2 for the sales regression is greater than in the rental regression, and is equal for one pair of databases. Importantly, the set of variables used for prediction is precisely the same in each pair of regressions.[10,11] Clearly, there could be many other reasons for the findings. For example, it is possible that variables that are not reported in the databases are more influential in the rental markets than in the corresponding sales markets. Yet, our conjecture about the way economic agents reason about prices is in line with these data.

Next, consider affine trees as the basic model of the rules agents use. When restricting attention to simple trees, as we do in this paper, the same intuition suggests that the best-fit simple tree would obtain a better fit (a higher adjusted R^2) on a sales database than on a comparable rentals database: a single-bifurcation tree is a simple enough formula that can be thought of as the rule that the sales markets coordinate on. Hence, it can serve as a focal point. By contrast, in the rental market, where there is no resale value and no need to coordinate on prices, such a simple rule is likely to provide a lower fit.

[9] This intuition was also the basis of a previous paper by Gayer, Gilboa, and Lieberman (2007). See the Discussion section for a detailed comparison of the two.

[10] Note that there were some cities with a considerable difference between the number of apartments for rent and for sale

[11] The conclusions regarding the adjusted R^2 on the test databases are the same.

The adjusted R^2 of each bifurcation model for both sales and rental apartments for our six databases appear in Appendix B in Table 27. As can be seen from the table, apart from Bogota, each single-bifurcation tree obtains a better fit on the sales databases than it does on the rental ones, moreover for more than 50% of that pairs the ratio of the sales to rent of the of adjusted R^2s is greater than 20%.

5 Discussion

In Gayer, Gilboa, and Lieberman (2007) we compared rule-based and case-based reasoning in both sales and rental markets. The former was modeled by hedonic regressions, and the latter – by the optimal empirical similarity (see also Gilboa, Lieberman, and Schmeidler, 2006). We predicted that rule-based reasoning would perform better, as compared to case-based reasoning, in the sales than in the rentals market. The reasoning behind this prediction is the same as described in Section 4 above. The main differences between the two are: (i) in this paper we focus on rule-based reasoning, and remain silent on alternative ways of reasoning agents might employ; and (ii) here we consider a different class of rules, namely, we augment the rule-based model to include affine trees and not only hedonic regression.

In both of these, the comparison of sales and rental data assumes that in the former there is a tendency to coordinate on rules, and that the need for coordination drives people to pick simple rules. Following this logic, we conjecture that, should we consider deeper affine trees (with additional bifurcations) the difference between sales and rental markets would diminish with the complexity of the tree. Due to the computational costs, we leave this conjecture for future research.

6 Appendix A: Proofs

6.1 Proof of Theorem 1

We prove NP-Completeness of the three problems using the same reduction, from SET COVER:

Problem 3 ***SET COVER:*** *Given a natural number r, a set of u subsets of $S \equiv \{1, ..., r\}$, $\mathfrak{S} = \{S_1, ..., S_u\}$, and a natural number $t \leq u$, are there t subsets in $\mathfrak{S}$ whose union contains S? (That is, are there indices $1 \leq j_1 < ... < j_t \leq u$ such that $\bigcup_{l \leq t} S_{j_l} = S$?)*

Given an instance of SET COVER construct a database as follows.

For simplicity, assume, without loss of generality, that $S_1, ..., S_u$ are all nonempty and distinct. Let $m = u$ and $q = 0$. To each subset S_j ($j \leq u$) define a binary variable x_j. For an element of S, $l \leq r$, let $\mathfrak{S}_l$ be the set of sets containing l,

$$\mathfrak{S}_l = \{ S_k \in \mathfrak{S} \mid l \in S_k \}$$

Define a database with n observations, $B = ((x_i, y_i))_{i \leq n}$, for $n = 1 + \sum_{l \leq r} (|\mathfrak{S}_l| + 1)$ as follows. For $l \leq r$, let there be one observation i with

$$x_i^j = \mathbf{1}_{l \in S_j} \qquad y_i = 1$$

and additional $|\mathfrak{S}_l|$ observations defined as follows: for each $k \leq r$, if $S_k \in \mathfrak{S}_l$ (that is, $l \in S_k$) there is an additional observation i with

$$x_i^j = \mathbf{1}_{l \in S_j} + \mathbf{1}_{k=j} \qquad y_i = 1$$

finally, there is one observation i with

$$x_i^j = 0 \qquad \forall j \qquad y_i = 0.$$

For example, if $r = 3$ and $\mathfrak{S} = \{\{1, 2\}, \{2, 3\}\}$ we obtain the following database, with 8 observations of 2 variables x and y:

	$x^{\{1,2\}}$	$x^{\{2,3\}}$	y
$(1 \in \{1, 2, 3\})$	1	0	1
	2	0	1
$(2 \in \{1, 2, 3\})$	1	1	1
	2	1	1
	1	2	1
$(3 \in \{1, 2, 3\})$	0	1	1
	0	2	1
*	0	0	0

Clearly, the construction of the database can be done in linear time. We claim that the set S has a cover of size t iff there is an affine tree that perfectly fits the data, and whose size is, roughly, t according to each of the measures. More precisely, we claim that the following are equivalent:

(I) There is a cover of S consisting of no more than $t\,(\leq u)$ subsets from $\mathfrak{S} = \{S_1, ..., S_u\}$;

(II) There is an affine tree (V, E, r, L, b, p, d) such that $SSE\,(B, (V, E, r, L, b, c, p, d)) = 0$ and $|image\,(b)| \leq t$;

(III) There is an affine tree (V,E,r,L,b,p,d) such that $SSE\left(B,(V,E,r,L,b,c,p,d)\right)=0$ and $|V\backslash L|\leq t$;

(IV) There is an affine tree (V,E,r,L,b,p,d) such that $SSE\left(B,(V,E,r,L,b,c,p,d)\right)=0$ and $Depth\leq t+1$.

To see that, assume first that (I) holds. Assume that $1\leq j_1\leq ...\leq j_t\leq u$ are the indices of the sets that cover S. Construct a tree that has t decision nodes, ordered sequentially as follows. The root branches on x^{j_1} at the cutoff level 0. If $x^{j_1}>0$ the tree ends at a leaf whose d value is the constant affine function 1 (that is, $\beta_0=1$ and $\beta_j=0$ for $1\leq j\leq m$). If $x^{j_1}\leq 0$, the tree leads to a node that branches on x^{j_2} at 0. If $x^{j_2}>0$, the tree ends at a leaf whose d value is 1 again. If $x^{j_2}\leq 0$, the tree continues to examine x^{j_3} and so forth. Only if all the t variables are non-positive does the tree end up with the value $d=0$ (that is, $\beta_j=0$ for $0\leq j\leq m$). Clearly, this tree satisfies $|Im\,(b)|=|V\backslash L|=t$ and $Depth=t+1$. We claim that it also has $SSE\left(B,(V,E,r,L,b,c,p,d)\right)=0$. Consider first an observation i in the database B that corresponds to an element k of S (that is, $1\leq k\leq r$). Recall that there are $|\mathfrak{S}_k|+1$ such observations in B, one consisting of 1's and 0's, and describing the incidence matrix of $k\in S$, and an additional $|\mathfrak{S}_k|$ observations in each of which exactly one of the 1's is replaced by 2. However, all these observations have exactly the same set of variables that are equal to 0 (corresponding to sets that do not contain k) and they have positive values (1 or 2) in the other variables. Because $1\leq j_1\leq ...\leq j_t\leq u$ define a cover of S, there is at least one j_l such that $x_i^{j_l}=1$ and this means that, for such x_i, the tree would branch into a leaf with $y_i=1$. The final observation, however, consists of all 0 values and would result in $y_i=0$. Thus, (I) implies (II), (III), and (IV).

We now turn to show that each of (II), (III), and (IV) implies (I). Let us first consider (II). Assume that there is an affine tree (V,E,r,L,b,p,d) such that $SSE\left(B,(V,E,r,L,b,c,p,d)\right)=0$ and $|Im\,(b)|\leq t$. Thus, the tree uses up to t variables and obtains a perfect fit of the database B. We claim that the subsets S_{j_l} corresponding to these variables have to constitute a cover of S. To see this, assume that $1\leq j_1<...<j_{t'}\leq m$ are the t' variables used in the tree (with $t'\leq t$). Assume, by way of negation, that $\bigcup_{l\leq t'}S_{j_l}\subsetneq S$ and let $k\in S$ be such that $k\notin S_{j_l}$ for each $l\leq t'$. Consider now the $|\mathfrak{S}_k|+1$ observations in B that are defined by k. For each of these observations i, and each variable j_l, $x_i^{j_l}=0$ and thus the computation of the tree for each of these ends up in the same leaf. Furthermore, $x_i^{j_l}=0$ obviously holds also for the last observation, so that it, too, ends up in the same leaf. It follows that for some coefficients β_j, $0\leq j\leq m$, all these $|\mathfrak{S}_k|+2$ observations have computations

that result in

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_i^j$$

Next, we argue that, for each of these $|\mathfrak{S}_k| + 2$ observations, $\beta_j x_i^j = 0$ for every $j > 0$. For the last one, $x_i^j = 0$ for every $j > 0$ by construction. For each of the other $|\mathfrak{S}_k| + 1$ observations, either we have $x_i^j = 0$ (in case $k \notin S_j$), in which case $\beta_j x_i^j = 0$ obviously follows, or else we have two observations that are identical in the values of all variables apart from one of them, j, with one observation having $x_i^j = 1$ and the other $x_i^j = 2$. As both result in the same value $y_i = 1$, it follows that $\beta_j = 0$ and $\beta_j x_i^j = 0$. Thus, all the $|\mathfrak{S}_k| + 2$ observations have computations that result in $y_i = \beta_0$. However, for the first $|\mathfrak{S}_k| + 1$ we have $y_i = 1$, while for the last – $y_i = 0$, a contradiction. Hence, such a tree that obtains a perfect fit has to define a cover of S.

Next, assume that (III) holds. Observing that, for any tree, $|V \backslash L| \leq |Im\,(b)|$ implies that (II) holds, and therefore also (I). $\square$

7 Appendix B : Tables

Table 3: Number of observations in the different databases

	Sao Paulo	**Buenos Aires**	**Bogota**	**Lima**	**Quito**	**Montevideo**
Apartments for rent and for sale in city	13,640	58,439	150,052	54,980	28,529	20,635
Obs. with missing values	2,683	44,508	94,905	40,903	10,722	10,801
Obs. will non- missing values	10,957	13,931	55,147	14,077	17,807	9,834
% of obs. with non-missing values	80%	24%	37%	26%	62%	48%
Duplicates	319	141	1,655	235	233	2
Outliers	883	726	3,729	1,180	1,376	673
Obs. in clean data	9,755	13,064	49,763	12,662	16,198	9,159
% of clean obs. in original database	72%	22%	33%	23%	57%	44%
Sales	4,297	10,199	36,778	10,878	10,352	6,048
Rent	5,458	2,865	12,985	1,784	5,846	3,111

Table 4: Sao Paulo– Descriptive statistics

		Mean	**STD**	**Median**	**Min**	**Max**
Sales	Price (BRL)	482,018	301,352	380,000	45,000	1,500,000
	Condo (BRL)	577	367	490	1	2800
	Size (m^2)	70.29	29.66	62	30	200
	Rooms	2.26	0.66	2	1	5
	Toilets	1.96	0.75	2	1	6
	Suites	0.86	0.61	1	0	4
	Parking	1.23	0.56	1	0	4
	Elevator	0.42	0.49	0	0	1
	Swimming Pool	0.55	0.50	1	0	1
	New	0.02	0.13	0	0	1
	Latitude	-23.56	0.06	-23.55	-23.77	-23.40
	Longitude	-46.61	0.09	-46.63	-46.81	-46.37
	Ana Rosa	0.11	0.07	0.09	0.01	0.33
Rent	Price (BRL)	2,451	1,613	1,900	480	10,000
	Condo (BRL)	742	466	600	1	3000
	Size (m^2)	77.74	36.68	65	30	200
	Rooms	2.24	0.76	2	1	5
	Toilets	2.00	0.82	2	1	7
	Suites	0.92	0.70	1	0	4
	Parking	1.34	0.68	1	0	5
	Elevator	0.32	0.47	0	0	1
	Swimming Pool	0.50	0.50	0	0	1
	New	0.00	0.02	0	0	1
	Latitude	-23.56	0.05	-23.56	-23.74	-23.39
	Longitude	-46.64	0.07	-46.65	-46.94	-46.38
	Ana Rosa	0.09	0.05	0.08	0.01	0.33

Table 5: Buenos Aires– Descriptive statistics

		Mean	**STD**	**Median**	**Min**	**Max**
Sales	Latitude	-37.88	0.32	-38.00	-38.35	-36.67
	Longitude	-57.46	0.24	-57.55	-58.01	-56.67
	Size (m^2)	82.99	72.00	59.00	16.00	1,050.00
	Rooms	2.65	1.17	3	1	6
	Bathrooms	1.57	0.75	1	1	4
	House	0.26	0.44	0	0	1
	Price (USD)	131,831	94,709	98,000	5,000	680,000
Rent	Latitude	-37.89	0.30	-38.00	-38.35	-37.02
	Longitude	-57.46	0.23	-57.55	-58.00	-56.80
	Size (m^2)	72.35	53.64	56	17	570
	Rooms	2.59	1.08	3	1	6
	Bathrooms	1.45	0.65	1	1	3
	House	0.21	0.41	0	0	1
	Price (URS)	19,518	20,467	13,000	2,200	185,000

Table 6: Bogota– Descriptive statistics

		Mean	**STD**	**Median**	**Min**	**Max**
Sales	Latitude	4.70	0.04	4.70	4.54	4.82
	Longitude	-74.06	0.03	-74.05	-74.17	-74.01
	Size (m^2)	136.74	88.12	109.00	10.00	800.00
	Rooms	2.87	0.96	3	1	7
	Bathrooms	2.82	1.10	3	1	6
	House	0.18	0.39	0	0	1
	Price (COP)	738,628,475	562,765,653	550,000,000	20,060,000	3,182,998,000
Rent	Latitude	4.69	0.04	4.69	4.57	4.81
	Longitude	-74.06	0.02	-74.05	-74.14	-74.01
	Size (m^2)	122.55	83.87	91	11	520
	Rooms	2.48	0.95	3	1	6
	Bathrooms	2.56	1.09	2	1	6
	House	0.09	0.29	0	0	1
	Price (COP)	3,627,813	2,955,958	2,600,000	295,000	16,500,000

Table 7: Lima– Descriptive statistics

		Mean	**STD**	**Median**	**Min**	**Max**
Sales	Latitude	-12.09	0.05	-12.10	-12.30	-11.89
	Longitude	-77.01	0.05	-77.01	-77.12	-76.85
	Size (m^2)	177.72	143.54	132.00	10.00	1,104.00
	Rooms	3.20	1.13	3	1	8
	Bathrooms	2.72	1.15	3	1	6
	House	0.23	0.42	0	0	1
	Price (USD)	305,269	263,630	224,100	13,500	1,950,000
Rent	Latitude	-12.11	0.02	-12.11	-12.20	-12.00
	Longitude	-77.02	0.03	-77.03	-77.11	-76.89
	Size (m^2)	160.79	127.05	120	16	1,008
	Rooms	2.68	1.07	3	1	7
	Bathrooms	2.57	1.05	2	1	5
	House	0.13	0.33	0	0	1
	Price (USD)	1,407	926	1,100	380	10,000

Table 8: Quito– Descriptive statistics

		Mean	**STD**	**Median**	**Min**	**Max**
Sales	Latitude	-0.18	0.06	-0.18	-0.37	0.01
	Longitude	-78.47	0.03	-78.48	-78.58	-78.37
	Size (m^2)	188.21	200.84	130.00	10.00	2,730.00
	Rooms	2.88	1.05	3	1	7
	Bathrooms	2.63	1.05	3	1	6
	House	0.44	0.50	0	0	1
	Price (USD)	162,800	106,497	130,000	5,500	685,000
Rent	Latitude	-0.18	0.03	-0.18	-0.29	-0.06
	Longitude	-78.48	0.02	-78.48	-78.54	-78.41
	Size (m^2)	120.25	80.64	94	11	623
	Rooms	2.19	0.99	2	1	5
	Bathrooms	2.20	0.95	2	1	5
	House	0.09	0.29	0	0	1
	Price (USD)	739	797	600	50	22,000

Table 9: Montevideo– Descriptive statistics

		Mean	**STD**	**Median**	**Min**	**Max**
Sales	Latitude	-34.89	0.02	-34.90	-34.93	-34.75
	Longitude	-56.15	0.04	-56.15	-56.28	-56.02
	Size (m^2)	147.85	220.92	80.00	10.00	4,774.00
	Rooms	2.27	0.95	2	1	5
	Bathrooms	1.65	0.82	1	1	4
	House	0.25	0.43	0	0	1
	Price (USD)	234,833	153,933	178,000	10,000	887,314
Rent	Latitude	-34.89	0.02	-34.90	-34.93	-34.76
	Longitude	-56.16	0.03	-56.16	-56.26	-56.05
	Size (m^2)	72.55	118.75	55	10	3,000
	Rooms	1.72	0.73	2	1	4
	Bathrooms	1.18	0.38	1	1	2
	House	0.13	0.33	0	0	1
	Price (UYU)	21,718	6,723	21,000	1,800	48,000

Table 10: Estimation of price of apartments for sale in Sao Paulo

Bifurcation variable	No cutoff	Ana Rosa	Size	Condo	Latitude	Longitude
C UC	252,032.478* (17,080.896)	68,132,920.074* (6,344,255.065)	26,912,344.329* (2,679,662.649)	30,915,099.486* (2,410,541.575)	46,156,280.248* (2,635,640.877)	33,759,669.470* (4,388,459.280)
C OC		24,691,242.001* (2,352,174.900)	72,280,472.099* (4,337,730.499)	87,818,115.784* (6,181,209.241)	-42,152,111.550* (12,073,660.884)	49,746,133.651* (14,133,717.641)
Condo UC	0.309* (.058)	85.836* (13.074)	236.837* (17.458)	188.871* (19.224)	179.449* (11.870)	151.165* (12.872)
Condo OC		42.845* (14.986)	94.597* (12.356)	-6.539 (15.863)	66.498* (24.220)	65.814* (17.276)
Size UC	17.754* (0.857)	5,440.788* (211.472)	4,660.219* (363.891)	4,533.095* (197.116)	4,522.176* (174.562)	4,688.437* (191.707)
Size OC		4,130.977* (173.780)	3,431.520* (181.330)	4,513.660* (214.132)	3,749.789* (282.882)	4,088.376* (209.965)
Rooms UC	-279.365* (29.297)	-1,648.305 (7,255.506)	-28,264.367* (6,568.940)	-22,608.907* (5,503.738)	-16,958.932* (5,784.853)	-19,492.340* (6,517.460)
Rooms OC		-2,406.014 (5,602.912)	-23,371.015* (7,487.496)	-20,760.708* (8,955.715)	-1,096.746 (9,211.195)	-3,014.319 (6,651.973)
Bathrooms UC	65.719 (45.032)	18,996.048* (9,159.302)	-55,998.898 (30,431.371)	32,135.709* (12,951.172)	31,392.091* (8,411.992)	41,685.886* (8,709.067)
Bathrooms OC		39,761.204* (8,641.582)	44,321.212* (7,135.598)	42,021.188* (8,094.039)	62,159.025* (13,185.756)	26,215.537* (10,657.847)

Continued on next page

Table 10 – continued from previous page

Bifurcation variable	No cutoff	Ana Rosa	Size	Condo	Latitude	Longitude
Suites UC	315.663*	30,343.198*	64,320.382*	-9,220.808	2,364.771	4,360.011
	(50.796)	(11,057.008)	(30,239.151)	(13,470.356)	(9,454.812)	(10,249.350)
Suites OC		-3,456.426	38,657.188*	46,610.601*	-11,292.823	9,663.238
		(9,324.086)	(8,718.432)	(10,008.793)	(14,644.823)	(11,429.474)
Parking UC	285.167*	111,410.108*	47,085.929*	81,979.537*	88,425.474*	93,775.400*
	(33.422)	(8,359.536)	(8,348.852)	(6,423.057)	(6,568.847)	(7,611.468)
Parking OC		113,204.019*	113,429.711*	119,105.964*	107,653.818*	109,692.755*
		(6,271.620)	(7,212.483)	(9,306.631)	(10,079.363)	(7,351.451)
Elevator UC	-92.931*	-24,217.680*	-8,599.759	-10,258.623	-24,507.640*	-19,333.712*
	(33.394)	(7,688.667)	(5,769.119)	(5,197.098)	(5,985.055)	(6,810.976)
Elevator OC		-12,704.033*	-10,998.076	-12,394.398	-7,174.095	-13,952.412*
		(5,223.751)	(7,850.156)	(9,842.047)	(8,537.056)	(6,374.159)
Pool UC	361.996*	76,439.781*	32,649.920*	34,904.756*	18,900.610*	11,841.061
	(32.828)	(8,541.652)	(5,776.905)	(5,264.432)	(6,023.919)	(7,097.004)
Pool OC		39,611.585*	48,585.308*	54,775.529*	46,701.616*	47,748.109*
		(5,365.412)	(8,959.481)	(11,523.979)	(9,020.741)	(6,524.254)
New UC	-193.359	-37,772.301	29,568.986	24,808.415	10,980.607	18,631.401
	(580.069)	(31,522.724)	(19,207.000)	(17,134.012)	(22,600.271)	(28,340.727)
New OC		8,941.715	-5,448.100	-60,858.438	773.505	18,368.633
		(17,913.010)	(33,780.046)	(62,432.105)	(28,071.335)	(20,885.131)

Continued on next page

Table 10 – continued from previous page

Bifurcation variable	No cutoff	Ana Rosa	Size	Condo	Latitude	Longitude
Latitude UC	119.776 (305.413)	757,586.395* (123,680.008)	104,418.287* (45,712.165)	76,706.491 (41,827.537)	186,095.276* (73,480.614)	131,531.943* (50,355.047)
Latitude OC		164,559.806* (38,668.857)	267,939.429* (76,135.577)	610,679.841* (102,642.108)	-1,309,269.709* (245,384.815)	154,202.744 (101,657.184)
Longitude UC	5,302.899* (329.853)	1,070,050.094* (127,392.894)	520,529.097* (58,091.875)	622,388.613* (52,192.274)	891,988.313* (70,511.011)	652,473.285* (95,476.313)
Longitude OC		445,913.095* (50,954.637)	1,405,513.415* (90,807.801)	1,563,954.270* (128,249.328)	-242,494.759 (157,529.443)	987,318.142* (263,476.838)
Ana Rosa UC	-15,301.152* (531.878)	-7,677,572.047* (267,971.152)	-1,311,200.488* (85,337.791)	-1,526,082.873* (77,259.500)	-2,338,740.544* (98,439.919)	-3,369,059.055* (122,328.334)
Ana Rosa OC		-1,204,842.073* (80,038.846)	-4,487,862.777* (137,255.043)	-5,804,177.732* (186,127.550)	-379,933.369 (213,454.480)	-1,777,012.288* (282,989.737)

The table presents the OLS coefficients conditional on the cutoff value estimates. Standard errors appear in parentheses.
UC stands for under cutoff and OC stands for over cutoff. * absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 11: Estimation of price of apartments for rent in Sao Paulo

Bifurcation variable	No cutoff	Ana Rosa	Size	Condo	Latitude	Longitude
	42,922,389.780*	754,270.013*	193,103.632*	114,062.910*	339,952.726*	336,404.134*
	(2,433,202.630)	(69,862.736)	(19,222.928)	(25,064.183)	(19,154.194)	(30,121.971)
C OC		107,349.778*	375,315.869*	353,576.668*	-413,861.966*	374,193.022*
		(15,851.524)	(32,717.207)	(21,388.518)	(55,452.626)	(74,377.019)
Condo UC	181.560*	-0.539*	0.673*	-0.027	0.228*	-0.017
	(10.778)	(0.084)	(0.081)	(0.214)	(0.070)	(0.068)
Condo OC		-0.101	-0.252*	-0.063	0.059	-0.048
		(0.068)	(0.078)	(0.064)	(0.097)	(0.094)
Size UC	4,077.892*	28.604*	15.026*	11.295*	18.628*	23.921*
	(151.126)	(1.516)	(1.823)	(1.967)	(1.072)	(1.062)
Size OC		19.232*	20.049*	20.357*	19.077*	13.057*
		(.898)	(1.326)	(.930)	(1.357)	(1.237)
Rooms UC	-16,692.579*	-521.184*	-296.857*	-97.211	-296.442*	-390.617*
	(4,991.125)	(55.663)	(37.466)	(52.188)	(35.997)	(36.957)
Rooms OC		-127.647*	-305.645*	-283.128*	-196.884*	-34.010
		(29.317)	(53.328)	(34.228)	(46.828)	(41.164)
Bathrooms UC	44,944.194*	-99.102	-71.957	-687.276*	72.405	-77.909
	(7,294.106)	(71.560)	(131.720)	(254.385)	(53.774)	(50.408)
Bathrooms OC		67.359	-24.744	30.502	12.964	42.947
		(47.693)	(47.401)	(43.656)	(75.513)	(78.502)

Continued on next page

Table 11 – continued from previous page

Bifurcation variable	No cutoff	Ana Rosa	Size	Condo	Latitude	Longitude
Suites UC	-2,317.260	564.638*	300.375*	711.384*	351.884*	452.682*
	(8,175.714)	(85.943)	(133.212)	(255.994)	(61.238)	(58.448)
Suites OC		203.967*	544.197*	425.972*	241.329*	200.033*
		(52.887)	(59.070)	(50.773)	(83.360)	(84.097)
Parking UC	89,301.199*	779.429*	317.992*	294.420*	280.765*	334.486*
	(5,645.862)	(62.381)	(45.476)	(74.873)	(40.760)	(43.333)
Parking OC		238.694*	298.537*	291.282*	316.202*	313.844*
		(33.570)	(46.493)	(35.324)	(54.114)	(45.120)
Elevator UC	-17,003.049*	-116.192	-99.697*	-58.446	-130.819*	-45.670
	(5,030.983)	(64.304)	(37.715)	(50.856)	(41.622)	(44.002)
Elevator OC		-67.080*	22.812	-77.386	-55.415	-88.866*
		(32.877)	(61.973)	(40.448)	(51.844)	(43.803)
Pool UC	23,026.301*	640.106*	336.491*	433.857*	307.952*	310.021*
	(5,135.148)	(68.529)	(36.334)	(50.646)	(41.096)	(42.792)
Pool OC		348.957*	623.598*	409.957*	433.048*	380.850*
		(32.043)	(64.849)	(40.360)	(50.578)	(43.960)
New UC	17,780.093	0.000*	-189.534	-122.358	-567.539	-38.382
	(18,127.447)	(0.000)	(553.132)	(947.009)	(972.685)	(932.493)
New OC		138.741	0.000*	-284.649	241.758	51.292
		(507.610)	(0.000)	(668.993)	(691.068)	(659.687)

Continued on next page

Table 11 – continued from previous page

Bifurcation variable	No cutoff	Ana Rosa	Size	Condo	Latitude	Longitude
Latitude UC	93,119.709* (42,390.857)	-1,523.111 (1,156.758)	183.668 (332.407)	-88.080 (430.283)	-844.090 (580.747)	601.780 (356.877)
Latitude OC		433.124 (275.328)	457.621 (620.458)	852.431* (392.629)	-10,999.768* (1,225.864)	559.206 (591.960)
Longitude UC	870,160.584* (52,497.679)	16,850.422* (1,419.155)	4,011.597* (366.757)	2,444.406* (473.323)	7,665.675* (443.746)	6,848.857* (639.512)
Longitude OC		2,062.309* (308.084)	7,743.072* (662.448)	7,091.145* (424.796)	-3,334.821* (708.761)	7,709.190* (1,399.803)
Ana Rosa UC	-2,143,856.156* (74,234.067)	-65,777.011* (2,668.128)	-10,860.287* (608.114)	-7,201.819* (777.331)	-20,385.473* (719.694)	-23,400.208* (863.921)
Ana Rosa OC		-7,634.520* (536.527)	-33,093.704* (1,094.008)	-26,147.208* (724.519)	-523.426 (1,086.622)	-13,447.067* (1,558.652)

The table presents the OLS coefficients conditional on the cutoff value estimates. Standard errors appear in parentheses
UC stands for under cutoff and OC stands for over cutoff. * absolute t-values of 2 or above (conditional on the cutoff parameter estimates)

Table 12: Estimation of price of apartments for sale in Buenos Aires

Bifurcation variable	No cutoff	Latitude	Longitude	Size
C UC	10,632,053.640* (608,483.084)	2,087,523.663 (1,201,370.401)	6,810,906.605* (1,223,997.062)	8,585,941.756* (676,587.898)
Constant OC		15,857,387.643* (715,617.622)	10,761,080.288* (856,680.665)	15,506,340.248* (1,078,874.046)
Latitude UC	-262,780.526* (15,622.441)	-14,245.219 (108,999.585)	-94,890.596* (30,759.390)	-209,978.948* (17,093.718)
Latitude OC		-410,306.673* (18,511.330)	-276,319.178* (21,402.369)	-366,421.554* (29,114.824)
Longitude UC	358,634.460* (20,721.258)	46,118.185 (81,017.161)	180,974.768* (35,323.903)	288,057.090* (22,868.164)
Longitude OC		546,810.886* (24,510.674)	369,871.748* (28,887.559)	510,236.967* (37,635.298)
Size UC	505.153* (15.173)	296.222* (27.609)	326.493* (23.592)	2,077.325* (54.508)
Size OC		585.199* (17.598)	646.041* (19.252)	315.829* (16.935)
Rooms UC	15,761.943* (945.865)	13,552.604* (2,814.302)	10,318.930* (2,025.461)	-11,076.030* (1,326.786)
Rooms OC		15,919.044* (991.895)	15,205.378* (1,069.115)	14,451.061* (1,868.814)
Bathrooms UC	50,177.239* (1,293.999)	47,129.729* (3,687.471)	42,579.175* (2,633.139)	25,011.923* (1,660.032)
Bathrooms OC		46,170.865* (1,362.985)	48,586.921* (1,467.120)	51,997.105* (2,056.948)
House UC	-36,933.660* (2,123.235)	-6,646.633 (5,557.124)	-10,330.537* (3,929.997)	-60,067.117* (2,528.664)
House OC		-31,269.482* (2,414.872)	-42,115.646* (2,642.206)	-87,728.237* (4,784.423)

The table presents the OLS coefficients conditional on the cutoff value estimates.
Standard errors appear in parentheses. UC stands for under cutoff and OC stands for over cutoff.
* absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 13: Estimation of prices of apartments for rent in Buenos Aires

Bifurcation variable	No cutoff	Latitude	Longitude	Size
C UC	2,914,395.173* (352,078.693)	1,820,473.769* (324,851.400)	2,020,412.510* (323,689.951)	2,508,454.527* (371,462.140)
Constant OC		58,438,997.544* (4,708,174.731)	80,677,314.458* (8,631,405.603)	4,332,267.692* (785,077.431)
Latitude UC	-19,979.392* (9,592.364)	-21,710.563* (8,764.203)	-19,032.678* (8,682.626)	-22,568.955* (10,141.661)
Latitude OC		369,151.809* (126,066.979)	643,337.251* (162,784.708)	-2,617.679 (21,073.732)
Longitude UC	63,902.109* (12,373.112)	45,982.018* (11,202.844)	47,692.716* (11,225.069)	58,523.900* (13,067.908)
Longitude OC		786,659.721* (114,688.371)	998,904.121* (208,316.301)	76,932.037* (27,394.359)
Size UC	59.014* (10.303)	40.006* (9.880)	42.550* (9.817)	162.347* (25.703)
Size OC		158.193* (27.931)	142.786* (30.762)	-2.944 (14.561)
Rooms UC	2,520.420* (486.367)	2,372.444* (452.677)	2,170.156* (450.683)	711.322 (638.344)
Rooms OC		3,119.767 (1,733.164)	5,805.863* (2,015.842)	3,753.696* (1,271.436)
Bathrooms UC	6,740.213* (719.426)	5,795.766* (688.163)	6,177.485* (679.004)	4,605.018* (803.138)
Bathrooms OC		8,917.640* (2,088.783)	7,861.316* (2,546.846)	5,965.430* (1,722.547)
House UC	-2,062.408 (1,047.485)	-2,986.622* (997.503)	-2,538.959* (981.117)	-5,479.077* (1,117.421)
House OC		-2,675.361 (3,188.534)	-16,142.786* (4,015.485)	-5,627.805 (3,405.091)

The table presents the OLS coefficients conditional on the cutoff value estimates.
Standard errors appear in parentheses. UC stands for under cutoff and OC stands for over cutoff.
* absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 14: Estimation of prices of apartments for sale in Bogota

Bifurcation variable	No cutoff	Latitude	Longitude	Size
Constant UC	176,467,169,785.068* (5,432,406,625.229)	211,686,057,720.620* (8,316,494,775.337)	83,534,095,746.501* (8,398,969,347.172)	115,951,076,672.627* (5,609,813,228.176)
Constant OC		133,427,119,753.486* (7,060,034,012.401)	135,129,212,353.537* (15,992,494,622.315)	476,880,671,849.863* (14,158,849,502.427)
Latitude UC	-1,701,328,127.610* (46,834,775.775)	2,305,992,819.511* (121,922,493.723)	-83,805,478.249 (54,120,332.757)	-1,513,779,128.611* (49,979,916.628)
Latitude OC		-1,986,644,276.297* (85,193,479.625)	-4,135,303,764.677* (78,625,752.581)	-1,005,568,898.556* (104,294,113.948)
Longitude UC	2,274,370,800.416* (72,054,847.548)	3,001,530,792.729* (110,645,651.000)	1,121,206,662.846* (112,263,259.757)	1,469,424,079.993* (74,190,116.622)
Longitude OC		1,675,067,374.697* (95,397,917.960)	1,562,166,147.438* (214,190,933.297)	6,364,360,335.871* (188,814,911.256)
Size UC	4,885,874.400* (30,645.842)	4,643,787.982* (54,111.347)	3,522,420.839* (45,129.620)	6,013,552.168* (61,963.055)
Size OC		4,653,948.463* (34,956.325)	5,533,819.358* (37,441.226)	3,393,408.524* (53,149.992)
Rooms UC	-89,653,393.388* (2,436,473.899)	-63,283,584.418* (4,146,809.697)	-81,669,142.361* (3,394,196.723)	-84,382,325.507* (2,668,198.043)
Rooms OC		-68,020,168.827* (2,886,899.607)	-74,592,332.996* (3,154,389.976)	-147,322,843.424* (4,824,203.437)
Bathrooms UC	117,271,183.629* (2,404,199.268)	143,921,099.494* (4,296,054.065)	124,368,666.538* (3,575,244.053)	66,296,595.206* (2,889,222.280)
Bathrooms OC		93,987,891.310* (2,724,931.914)	91,189,012.508* (2,900,315.392)	119,003,781.893* (4,475,328.350)
House UC	-213,200,611.673* (5,328,367.654)	-365,763,028.287* (10,165,923.610)	-122,627,016.814* (7,351,756.661)	-133,673,284.774* (5,775,797.707)
House OC		-116,899,573.158* (5,934,458.640)	-134,145,395.774* (7,139,379.279)	-370,164,215.873* (10,327,313.523)

The table presents the OLS coefficients conditional on the cutoff value estimates. Standard errors appear in parentheses. UC stands for under cutoff and OC stands for over cutoff. * absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 15: Estimation of prices of apartments for rent in Bogota

Bifurcation variable	No cutoff	Latitude	Longitude	Size
Constant UC	1,202,596,285.984* (56,763,808.810)	1,612,433,873.190* (84,380,083.129)	621,594,784.328* (84,114,799.961)	702,032,488.756* (62,633,109.272)
Constant OC		375,739,217.587* (72,354,664.585)	1,033,946,053.864* (173,897,985.649)	3,050,285,357.233* (127,458,933.064)
Latitude UC	-11,937,814.911* (455,493.618)	-1,186,313.889 (924,155.339)	-4,353,368.885* (519,904.948)	-7,516,710.338* (494,414.690)
Latitude OC		-12,941,046.945* (1,091,148.185)	-25,280,333.568* (841,049.255)	-27,062,223.242* (1,052,178.370)
Longitude UC	15,482,406.746* (755,624.284)	21,696,969.012* (1,111,794.700)	8,117,055.962* (1,129,977.938)	8,998,943.996* (831,430.434)
Longitude OC		4,247,387.580* (981,117.880)	12,356,636.813* (2,321,823.764)	39,461,343.736* (1,707,279.849)
Size UC	27,758.471* (307.893)	29,409.755* (395.845)	17,933.057* (485.906)	29,383.908* (763.501)
Size OC		22,505.564* (438.386)	30,930.076* (372.067)	23,379.128* (471.828)
Rooms UC	-369,466.913* (23,067.059)	-326,691.237* (30,811.871)	-283,553.240* (32,749.276)	-465,976.258* (27,078.532)
Rooms OC		-306,935.395* (31,822.656)	-372,587.883* (29,917.063)	-491,408.578* (46,291.739)
Bathrooms UC	490,875.763* (22,338.575)	571,473.590* (28,919.505)	664,780.015* (35,014.685)	327,770.413* (29,696.583)
Bathrooms OC		395,179.010* (31,234.872)	321,693.534* (26,392.931)	683,864.763* (37,199.360)
House UC	-1,569,389.900* (58,724.518)	-2,041,131.986* (87,245.565)	-703,966.909* (82,092.169)	-427,958.837* (81,300.709)
House OC		-779,793.307* (72,836.442)	-1,242,214.723* (79,648.353)	-1,749,976.842* (85,353.927)

The table presents the OLS coefficients conditional on the cutoff value estimates.
Standard errors appear in parentheses. UC stands for under cutoff and OC stands for over cutoff.
* absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 16: Estimation of prices of apartments for sale in Lima

Bifurcation variable	No cutoff	Latitude	Longitude	Size
C UC	-50,132,438.731* (3,031,465.535)	-135,108,679.822* (5,836,975.427)	41,802,623.502* (7,529,155.569)	-30,484,820.800* (2,978,593.966)
C OC		-4,014,646.876 (3,159,916.407)	-84,434,061.748* (5,406,408.422)	-203,131,530.948* (8,671,113.291)
Latitude UC	-613,784.335* (33,127.752)	1,117,276.970* (78,143.435)	-811,768.907* (45,752.468)	-453,533.743* (31,841.207)
Latitude OC		-880,062.394* (47,208.070)	348,048.077* (48,762.352)	-417,491.489* (151,723.225)
Longitude UC	-554,336.976* (37,227.045)	-1,930,477.165* (70,967.929)	670,023.558* (94,537.316)	-324,326.104* (36,499.263)
Longitude OC		85,691.602* (39,400.566)	-1,151,573.181* (71,243.651)	-2,577,471.896* (112,334.896)
Size UC	1,363.636* (13.899)	1,595.409* (20.597)	1,923.182* (26.910)	1,788.412* (26.943)
Size OC		1,082.947* (16.140)	1,188.658* (14.918)	886.478* (25.554)
Rooms UC	-20,970.376* (1,839.315)	-8,230.011* (2,516.755)	-29,516.740* (2,489.658)	-23,810.792* (1,849.491)
Rooms OC		-17,743.122* (2,224.011)	-18,301.685* (2,415.916)	-63,654.591* (4,707.890)
Bathrooms UC	53,413.287* (1,804.109)	30,029.308* (2,320.979)	40,296.427* (2,586.052)	34,975.050* (1,887.607)
Bathrooms OC		44,143.027* (2,389.315)	48,607.223* (2,194.271)	74,742.896* (4,562.191)
House UC	12,572.661* (4,639.017)	42,557.053* (6,193.212)	13,889.577* (6,645.812)	35.435 (4,644.282)
House OC		57,147.204* (5,922.602)	48,235.232* (5,687.060)	30,560.564* (11,860.530)

The table presents the OLS coefficients conditional on the cutoff value estimates.
Standard errors appear in parentheses. UC stands for under cutoff and OC stands for over cutoff.
* absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 17: Estimation of prices of apartments for rent in Lima

Bifurcation variable	No cutoff	Latitude	Longitude	Size
C UC	-311,383.081* (40,018.427)	2,294,352.858* (660,801.250)	-247,551.861* (48,544.689)	-247,454.338* (39,026.507)
C OC		-320,709.204* (37,872.347)	-324,708.747 (365,720.328)	-1,302,131.999* (171,053.774)
Latitude UC	-2,076.375* (793.370)	-22,118.455* (9,631.145)	-2,196.210* (800.868)	-1,859.938* (757.907)
Latitude OC		-1,714.899* (799.050)	13,787.480* (4,169.852)	26,638.006* (8,196.645)
Longitude UC	-3,720.159* (502.072)	33,347.372* (9,022.869)	-2,872.293* (590.172)	-2,923.807* (487.665)
Longitude OC		-3,898.667* (474.901)	-6,392.060 (4,856.589)	-21,195.789* (2,701.906)
Size UC	4.989* (0.193)	-0.755 (2.534)	6.523* (0.216)	5.717* (0.251)
Size OC		4.935* (0.184)	1.537* (0.387)	2.305* (0.719)
Rooms UC	86.843* (22.847)	2,455.657* (226.547)	21.615 (22.174)	56.059* (22.241)
Rooms OC		60.366* (21.742)	493.090* (108.381)	244.731 (206.285)
Bathrooms UC	28.581 (21.826)	-160.378 (299.642)	18.549 (21.383)	31.360 (21.254)
Bathrooms OC		37.185 (20.675)	-356.864* (75.088)	-895.415* (134.643)
House UC	238.186* (64.221)	-5,697.635* (598.010)	214.566* (63.071)	192.046* (61.393)
House OC		303.020* (61.127)	966.892* (230.574)	-1,999.708* (669.254)

The table presents the OLS coefficients conditional on the cutoff value estimates. Standard errors appear in parentheses. UC stands for under cutoff and OC stands for over cutoff.
* absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 18: Estimation of prices of apartments for sale in Quito

Bifurcation variable	No cutoff	Latitude	Longitude	Size
Constant UC	35,819,601.882* (2,128,335.717)	686,061.103 (451,151.617)	17,906.228 (65,738.324)	112,957.568 (58,618.368)
Constant OC		99,823.855 (58,199.583)	-2,727,469.782* (781,032.805)	-213,495.414 (332,251.106)
Latitude UC	-138,920.021* (13,276.403)	77,167.895* (21,981.384)	-1,828.734* (367.481)	-1,581.907* (365.835)
Latitude OC		-1,476.466* (378.997)	15,102.325* (2,578.033)	-3,814.143 (3,212.065)
Longitude UC	456,414.344* (27,126.422)	8,445.868 (5,731.069)	228.982 (837.561)	1,439.415 (746.821)
Longitude OC		1,272.080 (741.474)	-34,811.134* (9,958.882)	-2,732.009 (4,231.010)
Size UC	237.773* (4.847)	-2.224 (2.775)	3.772* (0.222)	3.137* (0.292)
Size OC		3.696* (0.220)	2.979* (1.309)	1.692 (1.275)
Rooms UC	-4,530.625* (1,091.115)	733.619* (190.857)	-37.216* (17.337)	-29.385 (18.176)
Rooms OC		-40.392* (17.305)	175.009 (131.253)	103.107 (101.333)
Bathrooms UC	44,728.464* (1,053.674)	-454.166* (178.106)	55.602* (18.284)	74.160* (18.744)
Bathrooms OC		67.493* (18.175)	235.717* (116.853)	-30.170 (109.506)
House UC	-32,609.030* (1,957.641)	-1,850.650* (358.761)	-187.906* (46.216)	-142.369* (46.536)
House OC		-169.305* (45.764)	-990.169* (225.229)	-770.668* (190.162)

The table presents the OLS coefficients conditional on the cutoff value estimates.
Standard errors appear in parentheses. UC stands for under cutoff and OC stands for over cutoff.
* absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 19: Estimation of prices of apartments for rent in Quito

Bifurcation variable	No cutoff	Latitude	Longitude	Size
C UC	89,673.721 (57,835.697)	39,608,816.784* (2,267,175.732)	31,226,829.374* (3,469,753.924)	29,021,602.592* (1,924,618.258)
Constant OC		-38,927,515.633* (4,755,463.111)	-68,765,987.516* (7,510,045.565)	74,790,265.289* (5,427,639.229)
Latitude UC	-1,601.387* (364.800)	449,734.776* (24,466.298)	-63,564.532* (14,617.640)	-111,733.115* (11,950.114)
Latitude OC		-39,365.596 (42,707.452)	-713,997.285* (35,800.132)	91,025.488* (35,908.210)
Longitude UC	1,143.042 (736.847)	503,224.921* (28,903.045)	397,667.975* (44,220.781)	369,542.600* (24,529.672)
Longitude OC		-496,161.830* (60,619.922)	-875,338.812* (95,745.853)	950,101.543* (69,175.221)
Size UC	3.688* (.220)	208.459* (5.035)	227.105* (5.798)	771.976* (13.261)
Size OC		274.684* (10.722)	249.472* (8.105)	11.826 (6.506)
Rooms UC	-33.630 (17.286)	948.392 (1,214.373)	-3,582.735* (1,134.106)	-17,857.306* (1,044.575)
Rooms OC		-3,387.094 (1,944.301)	7,686.205* (3,126.797)	-11,343.357* (2,453.656)
Bathrooms UC	63.163* (18.177)	43,257.784* (1,181.948)	42,725.731* (1,134.236)	23,912.655* (1,050.700)
Bathrooms OC		29,531.401* (1,872.542)	40,806.701* (2,551.005)	43,014.356* (2,381.129)
House UC	-193.255* (45.126)	-8,830.497* (2,492.727)	-36,735.669* (2,102.551)	-44,887.241* (1,730.191)
House OC		1,198.790 (3,447.543)	1,579.017 (4,815.502)	3,640.706 (7,969.153)

The table presents the OLS coefficients conditional on the cutoff value estimates.
Standard errors appear in parentheses. UC stands for under cutoff and OC stands for over cutoff.
* absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 20: Estimation of prices of apartments for sale in Montevideo

Bifurcation variable	No cutoff	Latitude	Longitude	Size
C UC	14,419,091.225* (3,014,642.053)	-358,575,560.901* (76,116,470.068)	-6,604,658.753 (10,795,166.768)	5,217,609.222 (3,501,468.105)
Constant OC		30,243,001.425* (3,079,562.314)	-9,476,749.015* (3,428,669.871)	28,006,237.871* (5,186,094.445)
Latitude UC	-1,253,049.216* (69,186.594)	-7,263,822.378* (1,847,320.838)	-657,609.730* (108,231.323)	-1,108,188.932* (79,777.825)
Latitude OC		-922,379.672* (70,732.124)	-2,768,357.215* (111,728.990)	-1,351,334.669* (118,811.661)
Longitude UC	1,035,463.694* (33,856.036)	-1,868,407.196 (944,566.401)	290,059.799 (151,367.910)	781,156.683* (42,252.870)
Longitude OC		1,111,598.140* (33,133.677)	1,552,132.765* (53,659.646)	1,337,626.074* (50,470.712)
Size UC	84.859* (6.570)	376.275* (53.089)	75.909* (15.203)	358.463* (49.544)
Size OC		79.507* (6.402)	72.285* (6.829)	16.756* (7.422)
Rooms UC	29,923.209* (1,884.706)	57,887.973* (7,578.033)	17,770.347* (2,988.333)	15,101.474* (2,236.332)
Rooms OC		26,922.231* (1,890.560)	41,166.479* (2,213.738)	46,921.153* (3,549.532)
Bathrooms UC	91,238.722* (2,170.792)	79,020.596* (7,382.713)	44,382.654* (4,591.840)	76,634.797* (2,824.896)
Bathrooms OC		85,693.027* (2,230.645)	87,300.578* (2,390.964)	79,563.533* (3,312.466)
House UC	-8,195.520* (3,662.220)	-17,420.512 (19,588.651)	10,180.420 (6,305.297)	-14,704.121* (4,522.289)
House OC		-5,358.477 (3,617.041)	-13,444.354* (4,124.668)	-54,966.811* (6,149.700)

The table presents the OLS coefficients conditional on the cutoff value estimates.
Standard errors appear in parentheses. UC stands for under cutoff and OC stands for over cutoff.
* absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 21: Estimation of prices of apartments for rent in Montevideo

Bifurcation variable	No cutoff	Latitude	Longitude	Size
C UC	-2,326,315.249* (251,649.510)	-2,979,684.793* (316,420.109)	53,625.515 (704,313.596)	-1,934,795.445* (256,160.916)
Constant OC		-5,133,659.344* (658,138.000)	-3,677,110.443* (330,372.746)	-6,429,612.218* (901,958.563)
Latitude UC	-108,133.099* (4,334.512)	-173,670.892* (8,793.788)	-65,704.545* (6,329.565)	-101,003.020* (4,427.001)
Latitude OC		-99,355.933* (11,320.170)	-180,106.106* (7,181.448)	-152,058.486* (15,550.240)
Longitude UC	25,578.757* (2,995.065)	54,664.593* (3,619.642)	41,569.656* (9,953.488)	28,131.271* (3,102.093)
Longitude OC		-29,826.064* (6,499.205)	46,250.740* (5,306.153)	-20,211.006* (9,052.378)
Size UC	3.179* (.747)	3.659* (1.080)	2.030* (.832)	49.954* (4.279)
Size OC		2.206* (1.015)	6.007* (1.476)	1.092 (.913)
Rooms UC	3,124.756* (151.255)	3,151.599* (159.415)	3,473.149* (234.608)	2,389.635* (164.612)
Rooms OC		3,955.430* (380.002)	3,178.829* (190.318)	3,777.432* (587.084)
Bathrooms UC	4,764.340* (287.310)	4,135.247* (291.760)	2,981.608* (567.241)	4,194.462* (293.851)
Bathrooms OC		4,209.396* (1,220.214)	4,360.784* (330.255)	2,846.285* (1,139.098)
House UC	2,035.191* (312.755)	2,833.709* (372.158)	2,322.209* (501.714)	875.886* (332.324)
House OC		996.170 (515.964)	1,896.713* (383.870)	6,708.980* (985.403)

The table presents the OLS coefficients conditional on the cutoff value estimates.
Standard errors appear in parentheses. UC stands for under cutoff and OC stands for over cutoff.
* absolute t-values of 2 or above (conditional on the cutoff parameter estimates).

Table 22: Buenos Aires– Measures of goodness of fit of different models

	Model	**Cutoff value**	**Train**						**Test**			
			MSE	**Adj.** R^2	**# obs. branch 1**	**% obs. branch 1**	**AIC**	**BIC**	**MSE**	**Adj.** R^2	**# obs. branch 1**	**% obs. in branch 1**
Sales	Latitude	-38.04	3.22E+09	0.64	703	8.6	201,800	201,898	3.41E+09	0.62	177	8.7
	Longitude	-57.56	3.25E+09	0.64	1,418	17.4	201,885	201,983	3.45E+09	0.62	347	17.0
	Size	123.03	2.95E+09	0.67	6,665	81.7	201,085	201,183	3.30E+09	0.63	1,683	82.5
	No cutoff		3.39E+09	0.62	8,159	100.0	202,205	202,254	3.65E+09	0.60	2,040	100.0
Rent	Latitude	-37.13	1.77E+08	0.57	2,071	90.4	50,066	50,146	2.20E+08	0.50	519	90.6
	Longitude	-56.86	1.79E+08	0.56	2,135	93.2	50,083	50,164	2.44E+08	0.45	538	93.9
	Size	135.02	1.98E+08	0.52	2,051	89.5	50,317	50,397	2.28E+08	0.48	514	89.7
	No cutoff		2.21E+08	0.46	2,292	100.0	50,553	50,593	2.50E+08	0.44	573	100.0

Table 23: Bogota– Measures of goodness of fit of different models

	Model	**Cutoff value**	**Train**						**Test**			
			MSE	**Adj.** R^2	**# obs. branch 1**	**% obs. branch 1**	**AIC**	**BIC**	**MSE**	**Adj.** R^2	**# obs. branch 1**	**% obs. in branch 1**
Sales	Latitude	4.68	7.60E+16	0.76	8,499	28.9	1,227,148	1,227,264	7.44E+16	0.76	2,108	28.7
	Longitude	-74.06	7.32E+16	0.77	10,383	35.3	1,226,034	1,226,150	7.01E+16	0.78	2,651	36.0
	Size	215.03	7.27E+16	0.77	24,773	84.2	1,225,847	1,225,963	7.17E+16	0.77	6,252	85.0
	No cutoff		8.49E+16	0.73	29,422	100.0	1,230,400	1,230,458	8.24E+16	0.74	7,356	100.0
Rent	Latitude	4.69	2.03E+12	0.77	5,398	52.0	323,908	324,010	1.97E+12	0.78	1,341	51.6
	Longitude	-74.06	2.01E+12	0.77	3,779	36.4	323,771	323,873	1.86E+12	0.79	959	36.9
	Size	171.07	2.08E+12	0.76	8,170	78.6	324,138	324,239	1.92E+12	0.78	2,048	78.9
	No cutoff		2.32E+12	0.73	10,388	100.0	325,264	325,315	2.19E+12	0.75	2,597	100.0

Table 24: Lima– Measures of goodness of fit of different models

	Model	**Cutoff value**	**Train**						**Test**			
			MSE	**Adj.** R^2	**# obs. branch 1**	**% obs. branch 1**	**AIC**	**BIC**	**MSE**	**Adj.** R^2	**# obs. branch 1**	**% obs. in branch 1**
Sales	Latitude	-12.10	1.62E+10	0.77	4,585	52.7	229,269	229,368	1.55E+10	0.78	1,194	54.9
	Longitude	-77.02	1.71E+10	0.75	4,236	48.7	229,744	229,843	1.68E+10	0.77	1,027	47.2
	Size	340.10	1.73E+10	0.75	7,921	91.0	229,884	229,983	1.82E+10	0.75	1,971	90.6
	No cutoff		2.03E+10	0.71	8,702	100.0	231,239	231,288	2.03E+10	0.72	2,176	100.0
Rent	Latitude	-12.15	3.33E+05	0.63	20	1.4	22,224	22,298	4.03E+05	0.39	7	2.0
	Longitude	-76.95	3.25E+05	0.64	1,351	94.7	22,190	22,264	2.56E+05	0.61	338	94.7
	Size	536.10	3.34E+05	0.63	1,393	97.6	22,226	22,300	2.43E+05	0.63	352	98.6
	No cutoff		3.77E+05	0.58	1,427	100.0	22,387	22,424	2.79E+05	0.59	357	100.0

Table 25: Quito– Measures of goodness of fit of different models

	Model	**Cutoff value**	**Train**						**Test**			
			MSE	**Adj.** R^2	**# obs. branch 1**	**% obs. branch 1**	**AIC**	**BIC**	**MSE**	**Adj.** R^2	**# obs. branch 1**	**% obs. in branch 1**
Sales	Latitude	-0.15	4.76E+09	0.57	5,762	69.6	208,056	208,155	4.59E+09	0.61	1,442	69.6
	Longitude	-78.44	5.17E+09	0.54	6,842	82.6	208,736	208,834	5.22E+09	0.56	1,714	82.8
	Size	398.03	3.96E+09	0.65	7,627	92.1	206,534	206,632	3.79E+09	0.68	1,895	91.5
	No cutoff		5.49E+09	0.51	8,281	100.0	209,228	209,277	5.42E+09	0.54	2,071	100.0
Rent	Latitude	-0.27	5.83E+05	0.14	31	0.7	75,379	75,469	3.61E+05	0.17	9	0.8
	Longitude	-78.44	5.82E+05	0.15	4,568	97.7	75,367	75,457	3.61E+05	0.17	1,138	97.3
	Size	374.04	5.86E+05	0.14	4,581	98.0	75,397	75,487	3.52E+05	0.19	1,142	97.6
	No cutoff		5.91E+05	0.14	4,676	100.0	75,424	75,470	3.52E+05	0.20	1,170	100.0

Table 26: Montevideo– Measures of goodness of fit of different models

	Model	Cutoff value	Train						Test			
			MSE	Adj. R^2	# obs. branch 1	% obs. branch 1	AIC	BIC	MSE	Adj. R^2	# obs. branch 1	% obs. in branch 1
Sales	Latitude	-34.92	7.10E+09	0.70	390	8.1	123,498	123,588	7.14E+09	0.70	91	7.5
	Longitude	-56.17	6.72E+09	0.72	1,558	32.2	123,232	123,323	6.81E+09	0.71	410	33.9
	Size	150.18	6.82E+09	0.71	3,702	76.5	123,304	123,395	6.70E+09	0.72	894	73.9
	No cutoff		7.61E+09	0.68	4,838	100.0	123,820	123,866	7.56E+09	0.68	1,210	100.0
Rent	Latitude	-34.87	1.95E+07	0.56	2,097	84.3	48,855	48,936	2.04E+07	0.56	517	83.0
	Longitude	-56.17	1.97E+07	0.56	924	37.1	48,878	48,960	2.00E+07	0.57	232	37.2
	Size	145.04	1.99E+07	0.55	2,361	94.9	48,903	48,985	2.05E+07	0.56	597	95.8
	No cutoff		2.13E+07	0.52	2,488	100.0	49,053	49,094	2.19E+07	0.54	623	100.0

Table 27: Adjusted R^2– Sales vs. Rent all models

	Model	Sales	Rent	Ratio
Sao Paulo	Ana Rosa	0.85	0.70	1.21
	Size	0.83	0.65	1.28
	Condo	0.83	0.66	1.27
	Latitude	0.81	0.65	1.24
	Longitude	0.83	0.69	1.20
	No cutoff	0.80	0.61	1.30
Buenos Aires	Latitude	0.64	0.57	1.13
	Longitude	0.64	0.56	1.13
	Size	0.67	0.52	1.30
	No cutoff	0.62	0.46	1.34
Bogota	Latitude	0.76	0.77	0.99
	Longitude	0.77	0.77	1.00
	Size	0.77	0.76	1.01
	No cutoff	0.73	0.73	1.00
Lima	Latitude	0.77	0.63	1.22
	Longitude	0.75	0.64	1.18
	Size	0.75	0.63	1.19
	No cutoff	0.71	0.58	1.22
Quito	Latitude	0.57	0.14	3.97
	Longitude	0.54	0.15	3.66
	Size	0.65	0.14	4.57
	No cutoff	0.51	0.14	3.77
Montevideo	Latitude	0.70	0.56	1.25
	Longitude	0.72	0.56	1.29
	Size	0.71	0.55	1.29
	No cutoff	0.68	0.52	1.30

8 Bibliography

Andrews, Donald (1993), “Tests for Parameter Instability and Structural Change with Unknown Change Point”, *Econometrica*, 61: 821-856.

Gayer, Gabrielle, Itzhak Gilboa, and Offer Lieberman (2007), “Rule-Based and Case-Based Reasoning in Housing Prices”, *BE Journals in Theoretical Economics*, **7**.

Hansen, Bruce E. (1999), “The Grid Bootstrap and the Autoregressive Model”, *The Review of Economics and Statistics 1999*, 81 (4): 594–607.

Hansen, Bruce E. (2000), Splitting and Threshold Estimation”, *Econometrica*, 68: 575-603.

Hansen, Bruce E. (2017), series econometrics for the 21st century”, *The Journal of Economic Education*, 48:3, 137-145.

Gilboa, Itzhak, Offer Lieberman, and David Schmeidler (2006), “Empirical Similarity”, *Review of Economics and Statistics*, **88**: 433-444.

Mullainathan Sendhil and Jann Spiess (2017), “Machine Learning: An Applied Econometric Approach,” *Journal of Economic Perspectives*, **31**: 87-106.

Schelling, Thomas C. (1960), *The Strategy of Conflict*. Cambridge: Harvard University Press.://www.overleaf.com/project/636a52e126d3e96566ec4162

Sharma, Himani and Sunil Kumar (2016), “A Survey on Decision Tree Algorithms of Classification in Data Mining”, *International Journal of Science and Research* (IJSR) **5**: 2094-2097.

Varian, Hal R. (2014), “Big Data: New Tricks for Econometrics”, *Journal of Economic Perspectives,* **28**: 3-28.