An Interactive Rent vs. Buy Calculator

Team Tufte Love – *Progress Report*

I. INTRODUCTION

One of the many decisions that the 40 million Americans that move each year must make is whether to buy or rent [1]. To some, the question hardly seems worth asking because few ideals are as ingrained in the American mentality as the dream of homeownership [2]. Homeownership has been found to enhance civic pride [3], contribute to less crime, and to create better familial environments and higher educational outcomes [4]. From a financial perspective, buying a house is often seen as a wise investment that builds wealth over time through home appreciation and generous tax breaks [5].

Recent research, however, is mixed. Statistical studies suggest that homeownership has a causal effect (at least Granger causal) on unemployment [6]. Financially, renting can be a better investment than homeownership [7]–[10]. Although increases in holding periods, inflation, and the spread between imputed rent and the opportunity cost of household savings can favor ownership, renters save on mortgage payments, maintenance, and improvement costs. Further, whether buying or renting is more favorable depends, in part, on interest rates and the actual holding period [8].

The many factors necessary to determine whether buying or renting is the better financial choice can be overwhelming. This has motivated our team to design an interactive application that can help movers decide which zip codes to target and whether buying or renting is the better choice in those specific zip codes.

II. PROBLEM DEFINITION

We aim to build an interactive visual analytics tool to assist potential movers by predicting and contrasting wealth accumulation over time in both rent and buy scenarios for various zip codes in the state of California. We aim to use data and machine learning techniques to supply users with starting points for choosing the required input parameters. The information generated by the tool should help users determine which zip codes better fit their budgets and whether buying or renting is a better financial decision in those zip codes.

III. SURVEY

Various frameworks have been proposed to determine the optimal housing decision. Beracha and Johnson [11] presented the first "horse race" between renting or buying and compared both scenarios using internal rate of return (IRR). Taubner proposed the use of the net present value (NPV) and found that households often need holding periods of between 5 and 10 years to achieve a break-even NPV [7]. The rent or buy decision problem is noted as equivalent to a disk spindown problem [12]. Krishnan, Long, and Vitter [12] have optimally solved that problem, but only using probability distributions under the assumption that resource use times are a priori unknown. Since our product will make recommendations based on anticipated durations at a property, this solution is not applicable.

The number of frequent movers has inspired the development of web applications to inform housing decisions. *The New York Times* published an interactive web application that keeps a running tally of the most common expenses, taking into account initial costs, recurring costs, and opportunity costs; but the application requires users to enter home price, holding period, mortgage details, tax rates, closing costs, and estimated interest rates [13]. That application does not make use of any market data to help users estimate prices or interest rates. Real estate applications such as Zillow.com [14] and Realtor.com [15] provide similar services and also rely on users to estimate interest and equity appreciation rates. Realtor.com provides initial estimates for costs and rates, but these values reflect a nationwide average, despite the fact that they significantly vary by state, city, and neighborhood [11].

IV. PROPOSED METHOD

Our approach will address the shortcomings identified in the previous section. We understand that any benefit derived from a financial model is highly dependent on its inputs. Many movers may not be informed enough to make educated estimates of sales and rental prices, real estate appreciation rates, and mortgage interest rates. If inadequate estimates are given to financial calculators, predictions are likely to be incorrect, and the user may be given a wrong recommendation. Our goal is to design an interactive application that is intuitive and less prone to error than the currently available financial calculators.

Another key innovation of our product is that we walk the user through every step of the home search and rent-versus-buy decision. Alternatives like Realtor.com force the user to get price estimates and run financial analysis in separate views, with no means of piping results from one view to the other.

Our approach will be prototyped for the state of California, the most populous state in the U.S. and the state with the third highest out-of-state movers [16]. We will utilize publicly available data (from Kaggle [17]) of historical housing prices in the U.S. from 1996 to 2017 to estimate median sales and rental prices for housing of desired characteristics in the zip codes that surround the area of interest. The same data set will be used to estimate the rate of appreciation of the real estate in that specific zip code. There is typically less inventory for house rentals than multifamily/apartment rentals.Due to such limitations, our current project will focus narrowly on two options: (1) Buy a house with the desired size and number of bedrooms or (2) Rent an apartment or multifamily dwelling in the same area with similar characteristics.

Current home listing and rental prices (as of March 2020) for each zip codes will be forecasted from historic data using the Prophet library [18]. Prophet is a nonlinear time series library that builds generalized additive models (GAMs) with time as a regression variable to account for trends (non-periodic changes), seasonality (periodic changes), and the effect of non-periodic holidays. Prophet was selected instead of traditional time series analysis techniques (e.g., ARIMA) due to Prophet's superior handling of outliers and changepoints, which are representative of external shocks in the real estate market (e.g., the Great Recession of 2008). The training set will be defined as the median home listing and rental price for each California zip code in the historical data set.

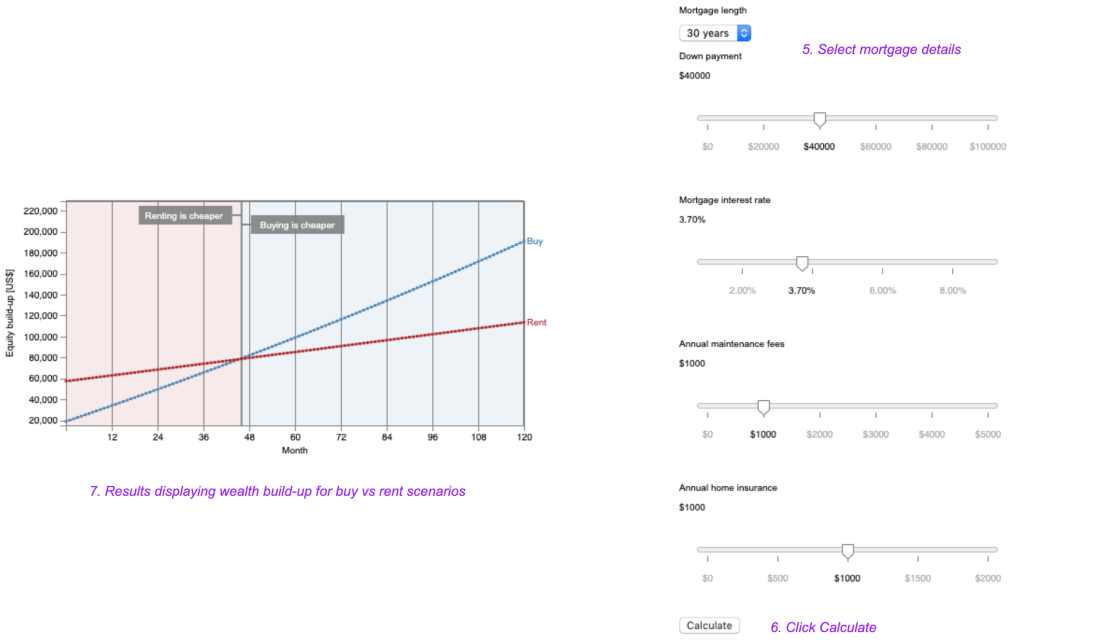
The same data set will be used to estimate the rate of appreciation of real estate in each zip code, which will be extracted from the slope of the forecast returned by the time series model.

Rent and buy scenarios are evaluated based on the wealth accumulated, using net present value(NPV), in each case. Our model is based in part on the financial calculations proposed by Khan [19]. For the current iteration of our project we have built in the following assumptions, but these might change or be made variable in the future:

- The spread (i.e., the net monthly outflow difference between the buy and rent scenarios) is assumed to be invested with an annual return rate of 4.0% (somewhat modeling what might be obtained, for example, in 401(k) account investments).
- All payments in both scenarios are modeled as processed monthly.
- Interest is compounded monthly.
- Closing and selling costs (in the buy scenario) are assumed to be 5.0% of the sales price.
- Selling costs are calculated at the end of each compared duration (modeling total costs as if one were to move after a given duration).
- Inflation is assumed to be constant at 2.0%.
- Marginal income rate is assumed at 30.0%.

Our user interface (UI) was designed to be intuitive and to prevent the user from entering inadequate estimates of real estate information, relying instead on the aforementioned estimates. The UI is composed of various blocks that will become visible only after the user has completed the previous steps and entered the required information. Figure 1 shows the two main views of the application. In the first view, the user starts the search by choosing a target area (e.g., Los Angeles, CA) and selecting a target zip code. The user will then enter the desired features for the house, such as the number of bedrooms and the square footage. Once the "Estimate" button is clicked, our machine learning algorithms will return sales

(a) Estimation of sale and rent prices for zip codes in the state of California



(b) Estimation of wealth accumulation in rent and buy scenarios

Fig. 1: The UI has two main views: the first view estimates median sales and rental prices for a house of desired features, while the second view calculates wealth accumulation in the rent and buy scenarios.

and rental price estimates for the desired houses in the selected zip code and the adjacent area. Price estimates for homes of similar characteristics will be shown through tool tips on the choropleth to allow the user to update their selection of target zip code based on the estimated prices. In the second view, the user enters mortgaged information (e.g., down payment, mortgage interest rate) and clicks "Calculate". A figure illustrating the accumulation of wealth in the rent and buy scenarios will appear on the side showing when it is a better decision to buy or rent. The user may then adjust any of the entries or selections and repeat the process.

V. EVALUATION

Financial predictions from our application will be compared to those obtained with other financial calculators (Realtor.com, *The New York Times*) to ensure that the predictions of wealth accumulation are

reasonably close. A similar comparison will be performed for the predicted decision margin, i.e., the holding horizon after which it is preferable to buy instead of rent.

The ease of use of our web application will be measured by the number of inputs that the user must enter manually. This approach resembles the DPU (defect opportunity per unit) metric used in Six Sigma [20]. The rationale of this choice is that every parameter that must be entered manually by the user is a defect opportunity. This metric will allow for comparison with other financial calculators such as Realtor.com and *The New York Times* application.

The machine learning algorithms used to predict sales price were evaluated on a testing set that consists of five zip codes (90292, 96146, 96161, 90044, and 90003). Time series data sets were split into training and test sets on November 2015. The sequence prior to the split date was taken as the training set to predict the median home listing and rental price for a 2 bedroom home. The mean absolute percentage error (MAPE) was adopted as the metric to measure the predictive power of the model regardless of the variability in the magnitude of the predicted variables (see Table I).

Zip code	Sale MAPE (%)	Rent MAPE (%)
90292	13.47	5.19
96146	14.73	NA
96161	5.93	NA
90044	18.68	NA
90003	3.09	NA

TABLE I: Mean absolute percentage error for time series forecasts of sale and rent price in five California zip codes. NA indicates that there was no data on rental prices to support the analysis.

VI. CONCLUSIONS AND DISCUSSION

Time providing, our team will expand the scope of the rent-versus-buy option to include the ability to delineate their selection of a home rental versus an apartment rental.

Due to data sparsity on a per zip code basis, comparable multifamily rental bedroom options will be derived by leveraging existing data points from the Zillow datset. More specifically, the ratio between the aggregate level housing value versus aggregate level rental value will be applied to the housing bedroom data available in order to calculate the equivalent rental metrics for each of the bedroom data points.

VII. CONTRIBUTIONS

O.A., R.L., F.L., S.S., and S.T. conceived and planned the project. S.S. and S.T. analyzed the data and built the machine learning algorithms to estimate rent and sale price, and the rate of appreciation of real estate. R.L. and F.L. programmed the financial calculations for wealth build-up in the buy and rent scenarios. O.A., R.L., and F.L. built the front end. O.A. added the choropleth with search capabilities for all zip codes in California. F.L. designed the wealth build-up plot and connected the financial calculations to the front end. The architecture and CI/CD pipeline were managed by O.A. All team members have contributed similar amount of effort.

VIII. PLAN OF ACTIVITIES

O.A., F.L., and R.L. will continue to work on the front end where they will develop an auto-zoom feature to the choropleth, add tool tips indicating estimated sale and rent prices for various zip codes, and animate the flow of the views. This front end team will also integrate the selections from the choropleth and house criteria to pass data to and from the back end models.

S.S. and S.T. will continue working on the machine learning models for price and appreciation rate. Time providing, our team will expand the scope of the rent-versus-buy option to allow users the ability to delineate their selection of a home rental versus an apartment rental. Due to data sparsity on a per zip code basis, comparable multifamily rental options will be derived by leveraging existing sales data points from the Zillow data set.

REFERENCES

[1] M. Sauter, "Population migration patterns: US cities we are flocking to," *USA Today*, 2018.

[2] C. C. Tu and M. J. Eppli, "Valuing new urbanism: The case of kentlands," *Real Estate Economics*, vol. 27, no. 3, pp. 425–451, 1999.

[3] W. M. Rohe, S. Van Zandt, and G. McCarthy, "Home ownership and access to opportunity," *Housing Studies*, vol. 17, no. 1, pp. 51–61, 2002.

[4] D. R. Haurin, T. L. Parcel, and R. J. Haurin, "Does homeownership affect child outcomes?," *Real Estate Economics*, vol. 30, no. 4, pp. 635–666, 2002.

[5] Z. X. Di, E. Belsky, and X. Liu, "Do homeowners achieve more household wealth in the long run?," *Journal of Housing Economics*, vol. 16, no. 3-4, pp. 274–290, 2007.

[6] D. G. Blanchflower and A. J. Oswald, "Does high home-ownership impair the labor market?," tech. rep., National Bureau of Economic Research, 2013.

[7] I. T. Tabner, "Buying versus renting–determinants of the net present value of home ownership for individual households," *International Review of Financial Analysis*, vol. 48, pp. 233–246, 2016.

[8] "Better off renting? owner-occupation is not always a better deal than renting," *The Economist*, Jan 2020.

[9] E. Beracha and K. H. Johnson, "Lessons from over 30 years of buy versus rent decisions: is the American dream always wise?," *Real Estate Economics*, vol. 40, no. 2, pp. 217–247, 2012.

[10] C. E. Herbert, D. T. McCue, and R. Sanchez-Moyano, "Is homeownership still an effective means of building wealth for low-income and minority households? (Was it ever?)," *Homeownership Built to Last*, 2013.

[11] E. Beracha, M. Seiler, and K. Johnson, "The rent versus buy decision: Investigating the needed property appreciation rates to be indifferent between renting and buying property," *Journal of Real Estate Practice and Education*, vol. 15, no. 2, pp. 71–87, 2012.

[12] P. Krishnan, M. P. Long, and S. J. Vitter, "Adaptive disk spindown via optimal rent-to-buy in probabilistic environments," tech. rep., USA, 1995.

[13] "Is it better to rent or buy?." https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html. Accessed: 2020-02-27.

[14] "Zillow." https://www.zillow.com. Accessed: 2020-02-27.

[15] "Rent vs. buy calculator." https://www.realtor.com/mortgage/tools/rent-or-buy-calculator/. Accessed: 2020-02-27.

[16] "Where americans are moving." https://www.usnews.com/news/best-states/slideshows/americans-are-moving-to-these-state. Accessed: 2020-03-26.

[17] "Zillow economics data." https://www.kaggle.com/zillow/zecon. Accessed: 2020-02-27.

[18] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[19] "Renting versus buying a home." https://www.khanacademy.org/economics-finance-domain/core-finance/housing/renting-v-buying/v/renting-versus-buying-a-home. Accessed: 2020-03-26.

[20] M. J. Harry, "Six sigma: a breakthrough strategy for profitability," *Quality progress*, vol. 31, no. 5, p. 60, 1998.