Effects of Home Rental Sites on Residential Real Estate: Evidence from New Hampshire

Glenn Ellison and Sara Fisher Ellison

Massachusetts Institute of Technology, Paris School of Economics, NBER, and CESifo
Massachusetts Institute of Technology, Paris School of Economics and CESifo

March 2024

Contact Information: Department of Economics, M.I.T., 50 Memorial Drive, Cambridge, MA 02142; email gellison@mit.edu and sellison@mit.edu.

Acknowledgments: The authors are grateful to the Paris School of Economics for their hospitality while working on this project. None of this would have been possible without our outstanding (and large!) group of research assistants: Sarah Berman, Mira Chandriani, Grace Coogan, Dionne-Marie Daku-Mante, Zoe Demers, Raymundo Duran Lozano, Cole Foster, Eric Gan, Tanush Garg, Deekshita Kacham, Tammy Le, Eileen Liu, Tingyi (Tina) Lu, Zev Moore, Tommy O'Donnell, and Meg Snyder. All played important roles, but we want to especially note the contributions of Tammy and Raymundo in leading the management of the SafeGraph and Zillow data, respectively. Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group.

1. Introduction

Vacation home rental websites like VRBO and AirBNB are close-to-textbook examples of how web-enabled reductions in transaction costs could lead to substantial improvements in social welfare. If a vacation (or primary) home whose owner would have left it vacant on a particular night can provide hotel-like services to a tourist, there can be a great deal of added surplus to split between the participants on the two sides of the exchange. We have used a fixed resource much more efficiently, improving social welfare. Such websites have, however, attracted a great deal of criticism from the very start. While some of this is no doubt attributable to hotels' worries about the long-term decline in their business that could result, there are also concerns related to the much-discussed housing affordability crisis. If the ability to rent unused nights via home rental platforms draws more purchasers into the second-home market (and new construction cannot fully offset this demand), then houses will be removed from the primary-home market, driving up prices and preventing some from purchasing primary homes.

An anecdote that occurred in the midst of our work on the project may motivate the questions it explores. While driving in rural New Hampshire, the between-song DJ banter that normally fades into the background suddenly caught our attention. The DJ was talking about out-of-towners who were buying up the properties in her town to rent them to tourists. She complained that many people who had lived in her town for decades were now having trouble affording the more expensive real estate. The DJ is not alone. Her sentiment is echoed throughout the popular press as exemplified in a recent New York Times article (White (2023)) profiling an Airbnb-influencer-turned-affordable-housing-crusader, who made the abrupt pivot after realizing that her lucrative side gig was placing her on the front lines of the war for affordable housing.[1] In addition to popular press accounts, organizations such as the Economic Policy Institute have sounded alarm bells about the negative effects of short-term rentals (Morrison (2023)).

Several questions one would need to assess these arguments are both intellectually interesting and potentially practically important. How separate are the markets for vacation properties and standard owner-occupied single family homes? Are vacation home prices really surging? How important are home-rental websites in driving this increase? Is this surge spilling over to the market for standard, non-vacation properties? In this paper, we use New Hampshire as a setting to explore the effects of the rise of home rental websites

[1] Also see Glusac (2020) for more anecdotes about the effect of Airbnb on neighborhoods.

on the market for primary and vacation homes.

Several features of New Hampshire make it well-suited for study. One is simply that there are a large number of vacation homes in the state. Another is that the popularity of direct lakefront homes as summer vacation spots gives us the opportunity to compare price trends for houses that are quite close together, but which differ substantially in their exposure to vacation-rental platforms. Another is that there is substantial variation across different parts of the state in the prevalence of vacation homes: there are many more in the areas near ski resorts and lakes. Differences in how home-rental websites affect different areas can potentially be informative about spillovers. Another feature is that we were able to obtain information from multiple sources to put together a rich, granular data set on the state's housing markets.

One data source that we will exploit in this paper includes (essentially) complete transactions data going all the way back to the late 1990's when home-rental sites like VRBO.com first appeared. The data continue through the first two decades of the 2000's, a period that saw explosive growth in the sophistication and use of online platforms for vacation rentals. We also have access to detailed data on the property characteristics derived from Zillow's ZTRAX databases.

To complement these transactions and characteristics data, we have constructed from various sources an extremely rich and detailed map of housing, vacation, and other amenities in the state. We also use data from the Census Bureau and NOAA to identify which houses are (likely) in vacation-relevant waterfront locations. These measures make it feasible to examine the coevolution of prices for vacation and standard homes within a narrow geographic area. We also have access to cellphone "ping" data, which gives us month-by-month information on individual mobility aggregated to the level of geographic sites. We have a measure of how many people visited the Story Land amusement park in Glen and the Funarama Arcade in Hampton Beach in August of 2019, for instance. Given the level of geographic detail at which the cellphone ping data are reported, we are able to create an index reflecting the proximity of each New Hampshire home to vacation amenities.

Our main analysis involves estimation of hedonic-style models of housing prices, focusing both on what we can learn from within-market comparisons of the prices of vacation and standard homes, and on what we can say about how increases in vacation homes have spilled over to the submarket for standard, or non-vacation, homes. The logic behind the approach and the strategy for identifying the effects of interest are discussed in Section 4. There, we argue that we can think of many parts of New Hampshire as having two distinct submarkets interacting along a boundary: a set of vacation homes will transact at prices related to the level of vacation amenities they provide, while most houses will sell to buyers in the standard submarket. The primary interaction between these two submarkets is that demand increases in the vacation submarket (driven by, for example, the rise of home-rental websites) will reduce the total supply of housing and land in the standard submarket, raising the prices that locals must pay. The model suggests that we can put a lower bound on the effect of home-rental websites on vacation home prices via a differences-in-differences approach examining how the premium paid for nearby houses more and less suitable for vacation use changed as home-rental websites became more popular. And it suggests that we may be able to identify spillovers with a triple difference approach comparing sensitivities in markets with more and fewer vacation homes.

Among our findings are that 1) the premium paid for waterfront homes increased by over 30% in the first twenty years after the appearance of home-rental websites, 2) other properties with high values of our vacation-amenities index also increased as the usage of home-rental websites grew, 3) we can provide statistically significant evidence of spillovers to the standard, or non-vacation, submarket, but 4) the effects we can identify in that submarket are substantially smaller than those in the submarket for vacation homes.

We discuss the relevant literatures in the following section, followed by some information on the empirical setting in Section 3. Section 4 discusses how home-rental platforms might be expected to affect prices for vacation and standard homes, and uses well-known models to explain our two identification strategies and motivate our estimation equations. We offer a detailed explication of the data sources and the variables that we have constructed from them in Section 5. It includes a description of exercises we carried out in constructing our proximity-to-vacation-amenities index. In particular, we need to select weights to place on proximity to different types of amenities, and we use auxiliary regressions to do this. Section 6 presents details of the main analysis and results. Its first subsection covers results derived from analyses of variation in the premiums paid for waterfront properties. The second covers results derived from analyses of variation in the premium paid for proximity to vacation amenities. We conclude in Section 7.

2. Literature Review

Our analysis draws on many literatures, some with substantive connections to ours and others with more methodological or data-related connections. The primary literature to which we contribute is a growing area of research on the effects of digital platforms on real estate markets. Farronato and Fradkin (2018) look at the welfare effects of Airbnb entry, which happened in a staggered fashion across US cities. They find the largest effects in large cities during busy periods, where supply of Airbnbs reacts to higher demand and acts as a pressure valve on the prices that hotels can charge in high-demand periods. Similarly, Zervas et al. (2017) find a causal impact of Airbnb rentals on existing hotels' revenues to be in the 8-10% range, with lower-priced hotels bearing a larger share of the effect.[2] The focus of Cunha and Lobão (2022)'s study is somewhat different. They are interested in the effects of short-term housing availability on primary housing markets in Portugal, which makes it closer to the questions we want to answer. They exploit a liberalization of some short-term rental laws and perform a difference-in-differences estimation, finding large housing price increases in those municipalities with particularly inelastic housing supply when short-term rentals are more widely allowed and, therefore, increase.

We draw, also, on a long-standing and rich literature using hedonic and hedonic-like models to analyze housing markets. This literature dates back at least to Rosen (1974) and Epple (1987), who established the theoretical basis for the empirical hedonic framework. The assumptions underlying the framework make it particularly well-suited to housing and real estate markets, where it has been used extensively for decades. Palmquist (1984) and Witte et al. (1979) provided early methodological advances, and Sheppard (1999)'s handbook chapter provides a more recent summary of hedonic techniques used in housing markets. A partial list of interesting papers in this literature could include Bishop and Murphy (2011), Chay and Greenstone (2005), Davis (2004), Baldauf et al. (2020), and Black (1999), studying questions as diverse as the importance of school quality in housing choices, the extent to which pollution levels are capitalized into house prices, climate change's differential effect on housing costs across the political spectrum, and the impact of changes in crime rates on real estate markets.

Finally, researchers before us have used and written about our two main data sets, one on properties, characteristics, and transactions from Zillow, and one on personal mobility from SafeGraph. We found the lessons they learned using these data helpful. In particular, Nolte et al. (2021) provided numerous tips and much wisdom on managing, extracting from, and using the Zillow data. Other interesting papers using Zillow data include Boeing (2017), Kahn (2021), and Poursaeed et al. (2018), the final being an example of the data's usefulness outside of economics, in machine vision research in particular.

[2] They study Airbnb entry into Austin, Texas, and find the effect to be more pronounced during periods of exogenously high demand, such as SXSW.

Chen and Rohla (2018) was among the first academic papers to use cell phone ping data, and they take time to discuss the advantages and shortcomings of those data.[3] In particular, they discuss sample selection issues and how representative their sample is of the US population as a whole. Other notable papers in this vein include Chen et al. (2021), Brelsford et al. (2022), Chiou and Tucker (2020), Gao et al. (2020), and Juhász and Hochmair (2020).

3. Setting

For those interested in the effect of online rental platforms on real estate markets and the interplay between markets for vacation and standard properties, New Hampshire provides an excellent setting. It has a relatively large rural population—47% of its total—and none of its urban areas dominate its labor or real estate markets. For instance, its largest city, Manchester, has a population of only 115,000 out of 1.4 million total in the state. Its second largest city is Nashua at 90,000, followed by Concord, the state capitol, at 45,000. Only 20% of the state's workforce works in one of those three cities. About 15% works outside New Hampshire, primarily in Massachusetts. These commuters are an important factor boosting real estate prices in the southeastern part of the state.

The state has a wealth of vacation and leisure activities across all four seasons, including downhill and cross-country skiing, hiking, mountain biking, boating, swimming, ice skating, ice fishing, fall foliage viewing, apple picking, spectator sports, and visiting historical sites. Not surprisingly, it attracts a lot of visitors, from the day-trippers up from Boston, to long-haul RVers staying a few days at one of the campsites; from skiers staying for a week at a mountainside resort, to those summering at lakeside cabins passed down through the generations.

As a measure of the importance of the vacation sector, in the Lakes Region and White Mountain counties (Grafton, Carroll, and Belknap) more than one-quarter of recent home sales were reported to have been purchased as second homes. For comparison, in the four southeastern counties closest to Boston, that number is less than 4%.

The Great Recession had a major impact on housing markets in the state, as it did in most places throughout the country. One can see this effect clearly in a graph of average prices by year from our data. Prices declined by about 25% from the 2005 to 2011, and did not reattain the 2005 peak again until 2018. The nonmonotonicity of price trends due to the

[3] They use data supplied by Veraset, as opposed to SafeGraph, but much of their discussion will be equally relevant.

Figure 1: Maps of New Hampshire Cities and Counties (Source: GISGeography.com)

Great Recession can be seen as making this a more favorable setting to identify the effects of the growth of vacation rental platforms. In particular, use of online rental platforms had only a brief and quite mild decline in usage during the Great Recession, whereas prices had a steeper, longer downturn.

4. Economics and Empirical Framework

In this section we discuss what classic models of housing markets suggest about the effects that home-rental websites could have on housing markets and to motivate the framework that we will use to provide evidence on the effects.

4.1 Economics

We can think of regional housing markets in New Hampshire as the pasting together of two submarkets, one for vacation properties and one for standard, primary homes. The salient features of each of these submarkets are best described by different classic models from the

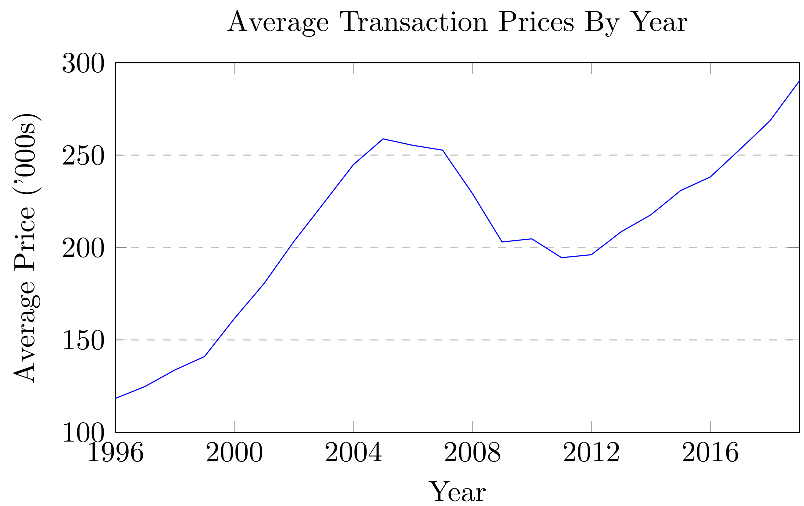

Figure 2: Average Transaction Prices in New Hampshire

real estate literature. And we think of home rental platforms has having boosted the value of vacation rentals by increasing the efficiency of home-rental markets.

4.1.1 Housing markets

First, we think of the majority of the housing market in most parts of New Hampshire as best described by what Glaeser (2008) refers to as the Rosen-Roback framework (Rosen, 1979; Roback, 1982). In such a model there are three primitives: (1) a consumer utility function that values housing, local amenities, and other consumption; (2) a production sector with output determined by labor, traded and nontraded capital, and productivity; and (3) a housing construction sector with costs increasing in housing built per unit land. As Glaeser nicely exposits, a well-chosen combination of functional form assumptions produces a tractable general equilibrium model in which the equilibrium levels of housing prices, wages, and the population are jointly determined by three primitives of the region: the productivity levels A, the amenity levels θ, and the amount L of buildable land.[4] Most relevant to our application is the relationship between housing prices p_H and the primitives of the model. It turns out to be log-linear

$$\log p_H = a_0 + a_A \log A + a_\theta \log \theta - a_L logL. \tag{1}$$

[4] The assumptions include that households have Cobb-Douglas utility functions $u(c, H, \theta) = \theta c^{1-\alpha} H^\alpha$, the production sector has profits $AN^\beta K^\gamma Z^{1-\beta-\delta} - wN - K$, where N is the number of workers and Z a fixed supply of non-traded capital, and the housing construction sector has profits $p_H hL - ch^\delta L - p_L L$ where h is the height of buildings and p_L is the price of land.

The elasticity a_L is the percent increase in housing prices from a one percent reduction in the amount of land devoted to this housing market. It can be written as a function of parameters in the utility and production functions, e.g. it is larger when there are greater decreasing returns to housing construction, and smaller when the weight on housing in the Cobb-Douglas utility function is larger.[5] In settings like New Hampshire, this elasticity would typically be fairly small.

Second, we think of the submarket for houses with high vacation amenities as best described by reinterpreting what Glaeser (2008) refers to as the Alonso-Muth-Mills (AMM) spatial equilibrium model of housing prices (Alonso, 1964; Muth, 1969; Mills, 1967; Wheaton, 1974; Brueckner, 1987). In the classic AMM model there is a fixed amount of land at each location ℓ, and the consumer utility from living at ℓ decreases with the distance from ℓ to the city center due to costs incurred in commuting to work in the city center. For example, if ℓ itself is the distance we might have

$$U(C, H; \ell) = W - c \cdot \ell - p_H(\ell) \cdot H + \alpha \log(H),$$

where c is the per unit commuting cost, $p_H(\ell)$ is the per unit housing cost, and H is the quantity of housing purchased. For markets to clear in the simplest version of this model with homogeneous households and the separable utility function above, prices must exactly offset commuting advantages:

$$\frac{dp_H(\ell)}{d\ell} = -c/H^*(\ell),$$

where $H^*(\ell)$ is the quantity of housing purchased by those choosing to live at location ℓ. As Glaeser (2008) explains, this differential equation plus the first-order condition for optimal housing choice imply a simple equation where the log of housing prices will decline linearly in distance from the city center.

$$\log p_H(\ell) = a_0 - \frac{c}{\alpha}\ell \tag{2}$$

City size in such a model is typically endogenized by assuming that a large potential population would receive utility $\underline{U}$ from living in a different city, and housing prices at the edge of the city are pinned down by an outside use for land (e.g. agriculture). The two outside options will pin down the physical (the largest distance ℓ at which land is used by those commuting to the city) and population (the number of people) size of the city.

[5] Intuition for the latter is that when the weight on housing in the utility function is greater, a house price increase would cause greater out-migration to less expensive cities, which offsets some of the loss of land.

We think of this model as relevant to the vacation-home submarket by reinterpreting $-c \cdot \ell$ as the level of vacation-amenity-derived consumer surplus received by those who are on vacation at location ℓ, rather than as a commuting cost. The two outside options that pin down the extent of the submarket will be the utility received by vacationing somewhere other than the region of New Hampshire, and the value of the property in the standard submarket, i.e. its value to those living and working in New Hampshire, which we have argued can be modeled a la Rosen-Roback. The two submarkets interact through their boundary. An increase in the value derived from owning a vacation property will lead the vacation submarket to expand, reducing the land available in the standard submarket, increasing prices in that submarket and leading to out-migration. A productivity increase in the production sector will increase prices in the standard submarket, causing it to expand, pushing up prices in the vacation submarket, and reducing the share of houses devoted to vacation use.

While urban economists using AMM as a model of city structure work with elaborations of the model in which urban land is built up to height $h(\ell)$ by a construction sector, which then feeds back into the distribution of commuting costs, the simplest version of the model seems reasonable for thinking of vacation housing markets where lot sizes will be smaller in high-priced areas, but vertical development remains unusual.

The left panel of Figure 3 illustrates the determination of equilibrium prices in such a model. The variable ℓ graphed along the x axis can be interpreted as the percentile rank of a location in the distance from desirable vacation activities, or as the negative of the value of vacation-amenities present at the location. The y axis measures the log of housing prices. The red and blue curves correspond to pair of equations (1) and (2) that determine the equilibrium allocation of properties between vacation and standard uses. The red line, corresponding to (2), is the log price $\log p^V(\ell; \underline{U}, p^{NV*})$ that clears the vacation home market at each location ℓ given the value $\underline{U}$ of the outside vacation option and the value of land in the standard submarket.[6] The blue curve, corresponding to (1), is the equilibrium log price $\log p^{NV}(L(1-\ell))$ that would prevail in the standard submarket if a fraction ℓ of the total land L was devoted to vacation homes, leaving $L(1-\ell)$ available.[7] We have graphed

[6] Under the assumptions sketched above, this curve is linear.

[7] We show this curve as slightly convex because, in the model sketched above, log prices increase linearly in $\log(1-\ell)$. While the unusually high prevalence of vacation homes in a number of New Hampshire markets is one of the motivations for studying this setting, vacation homes are still a small fraction of most geographic markets. Accordingly, one should think of the figure as zoomed in on the lowest ℓ part of the market. In this region with $1-\ell \approx 1$ we will have $\frac{d}{d\ell} \log p^{NV}(L(1-\ell)) \approx \alpha_L$ and $\frac{d^2}{d\ell^2} \log p^{NV}(L(1-\ell)) \approx \alpha_L$, so especially given that we've stretched out the x axis the curve will look fairly flat without much convexity.

this curve as much less steeply sloped than the p^V curve because suitability for vacation use will drop off quite quickly in most markets and the elasticity of prices to total available land should not be very large. The two curves intersect at $\ell = \ell^*$, which would be the equilibrium share of land in the vacation market. A graph of equilibrium house prices in the market versus ℓ would have the piecewise linear shape given by the two bold lines in the figure. On the left side prices are declining as we move further from the vacation amenities. Once we pass the cutoff ℓ^*, prices are unrelated to the vacation amenities and are constant at the level shown by the thick black line.

Now, suppose that home-rental websites enter the market and/or become popular. As in Gehrig (1998) and Ellison et al. (2004) the growth of such sites can make participants on both sides of the market better off. A standard increasing returns story for why we should expect both sides to gain is that, as the number of listed properties and the number of consumers using the platform increase, we improve the matching of consumers to homes. Homeowners are better off because prices and/or the probability of finding a renter increase, and consumers can be better off despite higher prices because of improvements in match quality. It seems plausible that benefits could have continued to accrue for quite some time as home-rental sites grew given that there is a great deal of variety in vacation homes. At any point in time it seems reasonable to assume that the benefit that vacation homeowners get from the availability of the platform would be roughly proportional to the home's price.

The right panel of Figure 3 illustrates how a market would change when an improvement in online vacation rental platforms increases the value of owning a vacation property. In the figure, which graphs log prices, an increase in values proportional to a home's value would lead to a constant upward shift in the red curve. This draws more land into the vacation submarket shifting the crossover point to the right to $\ell^{*'}$. Prices increase at every location ℓ. One the left side of the figure the increase is the gap between the dashed and bold red lines. On the right side of the figure it is the gap between the dashed and bold black lines. The increase is substantially larger in the vacation submarket. The dotted black line at the bottom of the figure graphs the log price changes in properties at each location. They are larger in the vacation submarket.

4.1.2 Home rental platforms

In the discussion above we treated the growth of home-rental platforms as providing a benefit to vacation-home owners. The literature on two-sided markets contains a number of

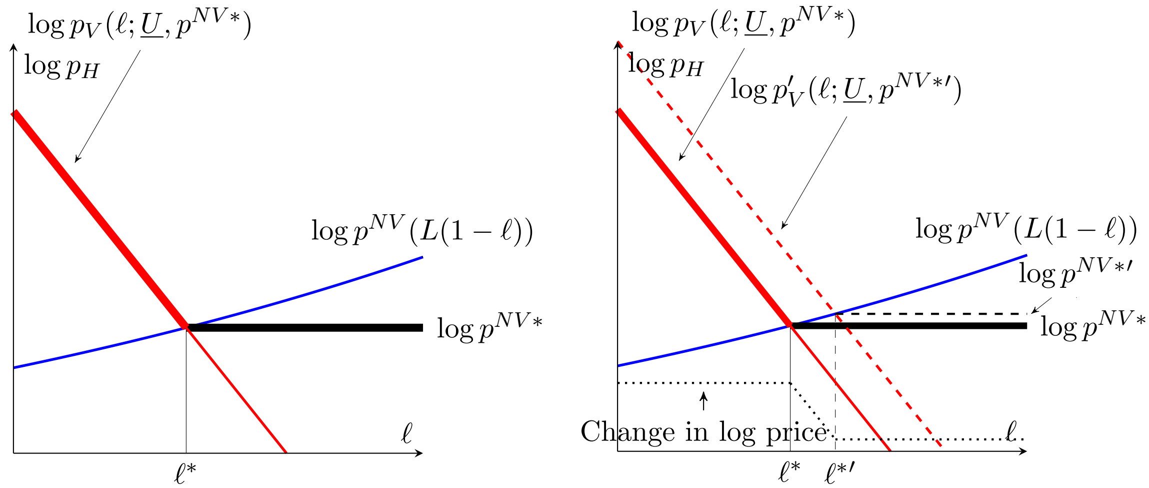

Figure 3: Model illustration. Determination of equilibrium prices, and a shift in the value of vacation properties in a market with vacation and non-vacation homes

papers that illustrate why this can happen and what patterns might be expected.[8] Two primary reasons for why sellers can be better off (and welfare higher) when platforms make markets thicker should be relevant in the home-rental environment: with differentiated products a thicker market will allow owners to rent to consumers whose idiosyncratic preferences are better matched to the property; and thickness increases efficiency by reducing the small-numbers noise that leads to some periods having excess and some insufficient demand at the price that clears the market in the average-demand limit. Theory suggests that the benefits can be very large when idiosyncratic preferences have a thick upper tail, and is particularly important in the initial growth of a market that had been very thin.[9]

The benefits of platform growth, however, will typically have substantial decreasing returns. Wolinsky (1984) and Gehrig (1998) note that benefits asymptote to an upper bound if idiosyncratic benefits are drawn from a distribution with bounded support or search costs lead consumers to stop searching once they have found a fairly good match. And Ellison et al. (2004) note that the benefits from eliminating price variation also asymptote to an upper bound. To bring out this observation in a simple way and think about how decreasing returns might vary across properties, consider a simple model along the lines of that in Ellison et al. (2004). Suppose the market for vacation rentals in some area

[8]See Rochet and Tirole (2006), Rysman (2009), and Jullien et al. (2021) for excellent surveys.

[9]See Perloff and Salop (1983) for an analysis of oligopoly with general distributions of idiosyncratic preferences. Ellison and Ellison (2024) estimate large benefits to the sellers of used books when sales initially moved online.

consists of S sellers and a Poisson(kS) number of potential renters with $k > 1$. Suppose that each renter only finds properties to be an acceptable match with probability p, and that if the match is acceptable, renter i's willingness to pay would be $\theta_i \overline{v}$, with $\theta \sim U[0,1]$ and $\overline{v}$ reflecting the upper bound generated by the ability to get something preferable at a hotel/resort. Write B for the random variable giving the number of renters for whom the rental is acceptable. When the market clears at the $S + 1^{\text{st}}$ highest valuation and $B \geq S$, the seller's expected revenue conditional on B is $\overline{v}\frac{B-S}{B+1} = \overline{v} - \overline{v}(S+1)\frac{1}{B+1}$. The expectation of this expression when B is Poisson with parameter kpS gives an approximation to each seller's expected revenue:[10]

$$E(p^*) \approx \overline{v}(1 - \frac{1}{kp}) - \frac{\overline{v}}{kp}\frac{1 - e^{-kpS}}{S}.$$

If we think of an increase in the popularity of the vacation rental platform as increasing the number of sellers and buyers on the platform, keeping the ratio k constant, then price is asymptoting (at rate $1/S$) to the price in a continuum-market limit, $\overline{v}(1 - \frac{1}{kp})$, as S grows. The final term, reflecting small-market losses from inefficient price variation is decreasing as S grows. Idiosyncratic properties with p small have fewer interested renters. Such properties will see benefits accrue more slowly with initial gains in S and won't gain as much in total in the $S \to \infty$ limit because the number of interested renters per property remains lower. But, the decreasing returns set in more slowly for such properties. Hence, there will be stages of home-rental platform growth at which they are deriving greater benefits from incremental growth (because the more popular homes are already receiving very close to their upper bound benefits).

Figure 4 provides an illustration. It graphs the expected price in a market with S sellers and $B \sim$ Poisson(pS) buyers for $p = 2$ and $p = 4$. In the market with (on average) four times as many interested renters as homes, there is a big benefit from the introduction of a market with $S = 1$ and $E(B) = 4$, and a substantial incremental gain from growth to $S = 2$ and $E(B) = 8$. The benefits from further growth are more moderate. In the thinner market with just twice as many interested potential renters as homes, the benefits from the establishment and initial growth of the market are barely more than half as large, but the decreasing returns set in more slowly. Even in absolute terms, the incremental benefits are larger as S increases from 3 to 4 to 5 and so on.

[10] If X is Poisson with parameter λ, then $E\left(\frac{1}{X+1}\right) = \frac{1-e^{-\lambda}}{\lambda}$. The approximation given for $E(p^*)$ is only an approximation because it ignores that the the formula for the expected price conditional on B only holds when $B \geq S$.

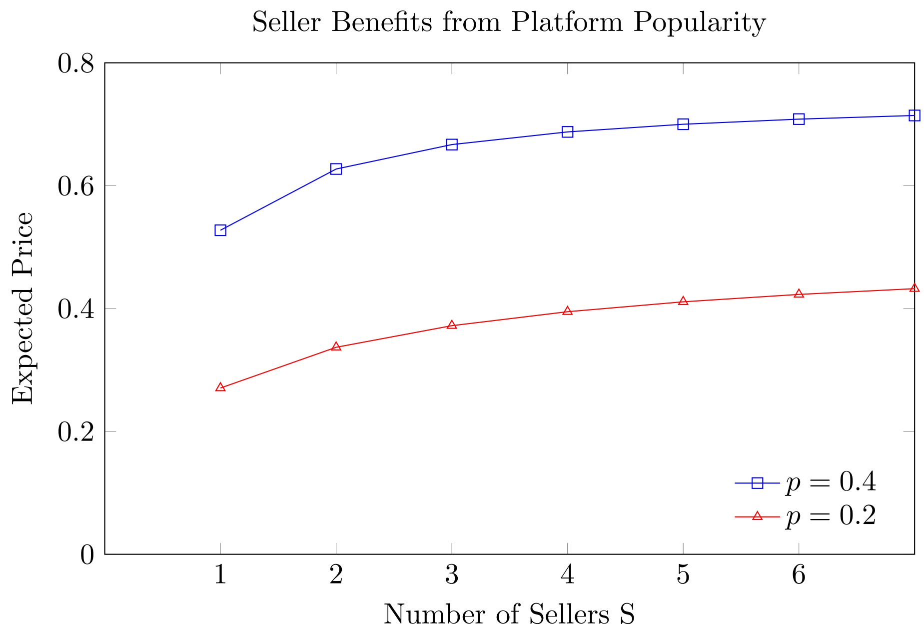

Figure 4: Expected seller benefits as a function of platform popularity

4.2 Empirical framework

Our empirical analysis aims to provide evidence that the growth of home-rental platforms have had such effects, and to estimate the magnitude of the effect on the prices of vacation and standard homes. Our primary identification strategy is to identify/bound the effects using two different sources of variation in the data. One can loosely be thought of as a difference-in-difference strategy exploiting the time-series variation within each submarket, examining how the difference between the prices of vacation and standard homes changes as platform use grows. The second can loosely be thought of as a triple-difference strategy examining how the relationship between the vacation/standard gap and the growth of home-rental platforms varies when we compare markets with smaller and larger numbers of potential vacation homes.

The time-series difference-in-difference identification strategy looks for the patterns one would see in comparing data generated as in the left and right panels of Figure 3. Consider a fully log-linear approximation to this model in which the curves for market j at time t are:

$$\begin{aligned} \log P_{jt}^{V}(\ell) &= (A_j + \nu_{jt}) + \beta X_t - M\ell \\ \log P_{jt}^{NV}(\ell_{jt}^*) &= (a_j + \eta_{jt}) + m\, \ell_{jt}^* \end{aligned}$$

with X_t representing the level of diffusion of the home-rental platform at t, and ν_{jt} and η_{jt} reflecting market-time level shocks to the levels of vacation and standard demand. We can find the equilibrium fraction of houses in the two submarkets by setting the two curves equal when $\ell = \ell^*_{jt}$.[11] Suppose the transaction price p_{ijt} at which house i in market j transacts at t is given by

$$p_{ijt} = \begin{cases} P^V_{jt}(\ell(i)) + Z_i\Theta + \zeta_i + v_{ijt} & \text{if } \ell_i \leq \ell^*_{jt} \\ P^{NV}_{jt}(\ell^*_{jt}) + Z_i\Theta + \zeta_i + v_{ijt} & \text{if } \ell_i > \ell^*_{jt} \end{cases}$$

where Z_i is a vector of the observable differences between house i's characteristics and mean house characteristics, ζ_i reflects unobserved quality differences, and the v_{ijt} are idiosyncratic transaction-specific errors. Plugging the equilibrium value ℓ^*_{jt} into these equations we find

$$p_{ijt} = \delta_{jt} + Z_i\Theta + \frac{M}{M+m}(A_j - a_j - \frac{1}{2}M\bar{\ell}) \cdot V_{ijt} + \frac{M}{M+m}\beta \cdot V_{ijt}X_t + u_{ijt},$$

where δ_{jt} is a term that varies only at the market-time level, V_{ijt} is an indicator for whether property i is in the vacation market, and u_{ijt} is a composite error term given by $u_{ijt} = \frac{M}{M+m}(\nu_{jt} - \eta_{jt})V_{ijt} + \xi_i + v_{ijt}$.[12]

A regression of transaction prices on Z_{ijt}, V_{ijt}, $V_{ijt}X_t$ and market-by-year fixed effects will provide consistent estimates of the coefficients in front of each of these terms if the error term in the final line is orthogonal to each of the regressors. The primary estimate of interest would be the coefficient on $V_{ijt}X_t$, which provides a lower bound on the effect β of a one-unit increase in platform penetration X_t on the value to owners of (and price increases for) vacation homes. It is a lower bound, because it is estimating the difference between the price increases for vacation homes and standard homes. In other words, it is estimating the difference between the level of the dotted black line in the right panel of Figure 3 for low and high values of ℓ, rather than the level of the line when ℓ is low. Other coefficients estimate the average premium paid for vacation homes relative to non-vacation homes and the hedonic values of house attributes.

The primary threat to this identification strategy is the possibility that the $(\nu_{jt} - \eta_{jt})V_{ijt}$ component of the error term might be correlated with $V_{ijt}X_t$. There are two factors that make it plausible that this is not a substantial problem. First, the component of the error term is the **difference** between the shocks to valuations in the vacation and standard components of each market, which does not have an obvious correlation with trending

[11] The equilibrium is $\ell^*_{jt} = \frac{1}{M+m}\left((A_j - a_j) + \beta X_t + (\nu_{jt} - \eta_{jt})\right)$.

[12] Here, we have written ξ_i for the sum of both components of a property's value that are unobserved by the econometrician: ξ_i is equal to ζ_i for non-vacation properties and to $\zeta_i + M(\bar{\ell} - \ell(i))$ for vacation properties. The market-time fixed effect is $\delta_{jt} = \frac{m}{M+m}(A_j + \nu_{jt}) + \frac{M}{M+m}(a_j + \eta_{jt}) + \frac{m}{M+m}X_t$.

variables like productivity and income growth. Second, the New Hampshire market had a large cyclical rise and fall in the time period in question, so even the individual shocks ν_{jt} and η_{jt} would not be expected to have the same monotone pattern as X_t. The presence of the $(\nu_{jt} - \eta_{jt})V_{ijt}$ and ξ_i terms in the errors also suggest that standard error calculations should allow for correlation across properties within market-time-vacation cells and across time for the same property.

In theory, the relevant parameters could be estimated separately using the data from any market within New Hampshire. In practice, this would be quite difficult to do, particularly in markets that do not have many vacation homes. We will estimate the model on the entirety of the state, essentially assuming that the coefficient on the key interaction $V_{ijt}X_t$ only varies across markets with variation in two covariates.[13] The first proxy we will use for the vacation dummy is an indicator for whether a home is located on a vacation-relevant waterfront.[14] We can measure the vacation premium while using quite finely defined place-by-time fixed effects: we will primarily use ZIP code-year fixed effects for the market-by-time dummies that the approach requires even though adjacent ZIP codes would typically be thought of as part of the same local housing market.[15]

We will also use a second vacation proxy that reflects proximity to general vacation amenities. By design, this variable changes more slowly across space because most of these amenities are sites to which vacationers would drive.[16] For example, there are no homes from which one would walk to the Storyland Amusement Park, so with visitors arriving exclusively by car, one would expect a moderate drop off in vacation utility as one moves from 1 to 5 to 10 miles away. As a result, we there will not be as much variation in the vacation amenities index within a ZIP code. This limits the power and precision of estimates obtained in models with ZIP-by-year fixed effects, but we obtain some significant results from this approach as well.

In the discussions above we have assumed that the benefits that owners of vacation homes derive from being able to rent their homes are proportional to the value of the

[13]We will allow the coefficient on the vacation dummy to vary with the distance of a market from Boston and with proxies for the fraction of homes in a county which are vacation homes. Non-vacation homes are more expensive as one gets closer to Boston, which makes the vacation-premium smaller.

[14]Though usually taking on values close to 0 or 1, the measure is a prediction and will take on continuous values reflecting likelihoods.

[15]There are 250 ZIP codes in New Hampshire, so the average ZIP is about 6 miles across and has about 6000 residents.

[16]In future work we may try to construct a very sharp proxy along the lines of our waterfront indicator by identifying slopeside ski condominiums, but have not yet done so because we are skeptical that we will be able to identify a large enough number of such properties to obtain significant results.

home. In practice, there are probably departures from this. If there are fixed costs to using a house as a short-term rental, then the net-of-costs value to homeowners may not be strictly proportional to the home's value. This issue may be particularly relevant when one is thinking about extracting low-season value: a house that is far from all vacation amenities would not benefit from incremental platform growth because it would not make sense to even incur the fixed cost of listing it on the platform. We will also implement an identification strategy based on this nonlinearity in Section 6.2.

We also present results using a different identification strategy, making additional use of cross-market differences. To provide intuition for this "triple difference" approach, Figure 5 draws a version of our price-change graph appropriate for two different situations: the graph on the left depicts a market in which very few homes have amenities that make them candidates for vacation use, whereas the graph on the right depicts a market with many more potential vacation homes. The dashed lines in each graph illustrate the effects of an increase of the same magnitude in the value of vacation homes. In the graph on the left, very few homes shift from standard to vacation use, and the effect on prices in the standard market is barely visible. In the limit as the number of potential vacation homes goes to zero, a within-market estimate of how the vacation-premium changes when platform use X_t increases would be estimating the full effect β of platform use on vacation home prices.

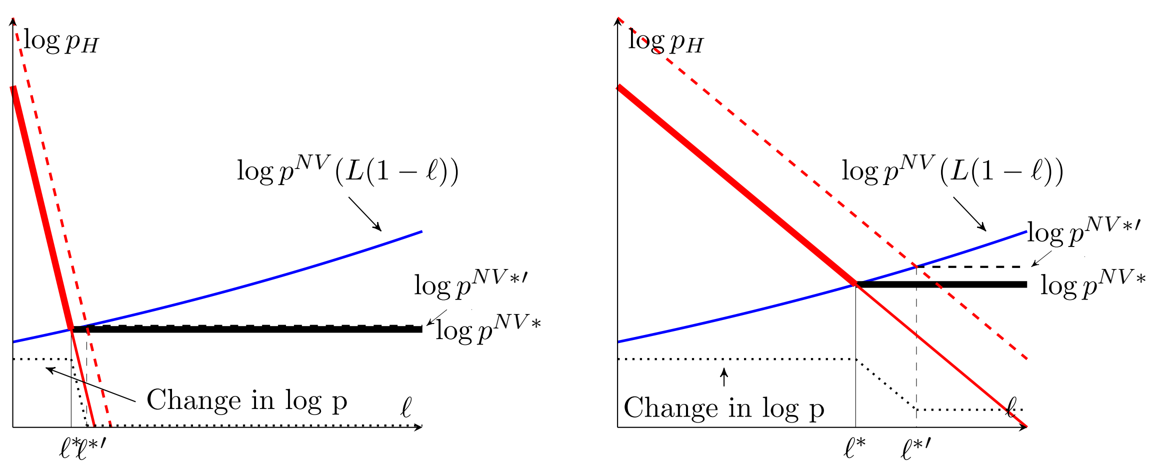

Figure 5: Model illustration. Determination of equilibrium prices, and a shift in the value of vacation properties in a market with vacation and non-vacation homes

The figure on the right depicts a market with many more homes that are vacation-home candidates, for instance, an area with more lakes. Here, a shift of the same magnitude in the value of owning a vacation property causes many more homes to shift from standard to vacation use. As a result, the increase in the price of vacation homes is much larger, and the change in the measured "vacation premium" (the difference between the prices of vacation and standard homes) is smaller. Crucially, note that the amount by which the increase in the vacation-premium is smaller is the amount by which the prices of non-vacation homes increased when X_t increased. We take two important ideas from this comparison. First, by examining how much less sensitive the vacation-premium is to changes in X_t when the number of potential vacation homes is larger, we can estimate the effects of home-rental platforms on the submarket for standard, or non-vacation, homes. Second, by examining the limit as the fraction of potential vacation homes goes to zero, we can identify the full magnitude of the effect of home-rental platforms on the prices of vacation homes.

5. Data

Just as digitization has had profound effects on markets and the questions that economists study, so has it affected the data that we have at our disposal to study these questions and markets. We take advantage of two so-called "big data" sets made available to researchers by private firms, Zillow data on property characteristics and real estate transactions and SafeGraph data on personal mobility. We have also drawn from other sources, which we discuss below, but let us start by describing the Zillow and SafeGraph data sets.

5.1 Zillow real estate data

The first of the two primary data sets we use is a near-census of real estate transactions over approximately the last two decades in the state of New Hampshire. These data are a matter of public record, of course, but we are enormously grateful to Zillow for compiling them from different sources, standardizing format and variables, and making these processed data sets available to researchers. For our purposes, we made most use of two particular data sets, 1) a list of real estate transactions sourced from public records identified by a property ID, and 2) a detailed set of property characteristics from tax assessments identified by the property ID. We needed to further process, filter, and merge these two sets to obtain our final analytical data set, as we explain below.

We have filtered the Zillow data to tailor it for our purposes, choosing single-family homes that were sold in arms-length transactions, between the start of Zillow's data in 1996 and the end of 2019.[17] Zillow includes detailed location data on each property, in-

[17] We have data on transactions beyond 2019, but we omitted those data to avoid COVID effects. COVID-related impacts on the real estate market are of great interest, but we abstract away from them here to focus on the impact of digital rental platforms.

cluding street address and precise latitude and longitude. Zillow also includes standard measures of property characteristics, such as the lot size, interior square footage, number of bedrooms, number of bathrooms, and condition.[18] These measures could vary over time for a property as renovations or additions are made, and, in particular, could be different across transactions for properties transacted multiple times during our data period. The raw data contain dozens of such variables, and we have retained the ones that will be most useful to us. Finally, Zillow has details concerning each transaction, most importantly sales price.[19] The price data have some clear errors, e.g. a substantial number of prices appear to have been entered with too many or too few trailing zeros. To eliminate these, we ran a preliminary hedonic regression and dropped all observations with prices that are not within a factor of 8 of the predicted value.[20] These filters leave us with a total of 488,143 transactions. The top section of Table 1 includes variables from this data set that we use in our analysis. We we have omitted from the table variables reported in categories, such as the heating system type, and locational variables such as longitude, latitude, and ZIP code. We use the former in our analysis, but summary statistics on the category labels would not be very informative. We use the latter to construct our "amenities" measures as described in the next paragraph, and include zip-code dummies and variables related to the properties distance from Boston in some regressions.

To give a sense of some patterns in the transaction-price data, Table 2 presents estimates from a hedonic regression with *log(SalesPrice)* as the dependent variable and various characteristics and time dummies as the right-hand-side variables.[21] Most of the property characteristic effects are as one would expect. The house's interior size and condition both have very large effects on the sale price. The bathroom coefficient implies that having one more bathroom is associated with the price being about \$20,000 higher. Sales prices are also higher for properties with larger lot sizes, but the magnitudes of the effect is modest.[22] One notable result is that houses with more bedrooms do not sell for more, holding square

[18] We drop observations with missing data on interior size, but retain observations with other attributes missing, setting the variable equal to the non-missing mean and adding (unreported) dummy variables for each variable being missing to our regressions.

[19] Note that we have information on the characteristics of *all* properties, even those that have not been transacted in the last two decades, but we retained only those that had at least one transaction during the period 1996-2019 since our dependent variable is based on sale price.

[20] This dropped 8147 transactions, which is about 100 times the number of such outliers that would be expected if errors in log prices were normally distributed.

[21] In addition to the variables shown in the table, the regression included zip-code dummies, year dummies, 23 dummies for different types of heating and air conditioning systems, and dummies for missing values of the lot size, bedroom, bathroom, and property condition variables.

[22] Given that the mean lot size is over two acres, it need not be surprising that the marginal value of additional acreage is modest.

Table 1: Summary Statistics for Sales Price Regressions

Statistic	Mean	St. Dev.	Min	Max
Transaction-level variables				
SalesPrice (000's)	216.5	157.5	3.6	4,950
log(SalesPrice)	12.06	0.72	8.19	15.41
TransactionYear	2008	6.87	1996	2019
Bedrooms	2.96	0.97	1	8
Bathrooms	1.92	0.75	1	5
LotSize (acres)	1.94	3.29	0.04	48.04
log(LotSize)	-0.04	1.22	-3.19	3.87
BuildingSize (sf)	1,779	808.3	430.0	5658.0
Property-level variables				
WaterFront	0.035	0.150	0.000	0.998
VacationAmenities	2.091	0.719	1.159	5.680
VacationAmenWin	1.662	0.417	0.936	3.266
VacationAmenSpr	1.918	0.485	1.232	3.562
VacationAmenSum	2.592	1.700	1.069	13.243
VacationAmenFal	2.193	0.689	1.325	5.464
VRAmen_Frac	0.129	0.164	0.000	0.668
Monthly variables				
log(PlatformUsage)	17.00	2.90	10.41	20.15
S.PlatformUse	0.00	1.00	-2.27	1.08

footage fixed. The data are from the pre-COVID era, and it could be that preferences for having larger numbers of smaller rooms have since shifted. The month-of-year dummies indicate that prices are lower for houses transacted in January through March and higher for houses transacted in the summer and fall. These could reflect that houses are less-well matched to the buyers in the months when real estate activity is thinner, or that the houses that transact at these times are worse in unobserved dimensions. Figure 10 in the Appendix shows the coefficient for the year dummies in the hedonic regression with 95% confidence intervals.[23] It shows an up-down-up pattern nearly identical to that in Figure 2.

5.2 Waterfront classification

One of the primary ways in which we will look for evidence of the impact of home-rental websites on real estate markets is by examining relative price changes for waterfront and non-waterfront properties. A Zillow database contained a variable indicating whether a

[23]The omitted year dummy in the regression is 1996.

Table 2: New Hampshire house prices: a hedonic regression

	Dependent var.: *log(SalesPrice)*	
	Coef.Est	(Std.Err.)
LogBuildingSize	$0.579^{\dagger}$	(0.005)
Condition_Poor	$-0.659^{\dagger}$	(0.026)
Condition_Fair	$-0.333^{\dagger}$	(0.012)
Condition_Good	$0.124^{\dagger}$	(0.004)
Condition_Excellent	$0.288^{\dagger}$	(0.013)
LogLotSize	$0.038^{\dagger}$	(0.002)
Bathrooms	$0.108^{\dagger}$	(0.002)
Bedrooms	$-0.061^{\dagger}$	(0.002)
February	-0.008	(0.004)
March	0.012^{**}	(0.004)
April	$0.031^{\dagger}$	(0.004)
May	$0.061^{\dagger}$	(0.004)
June	$0.086^{\dagger}$	(0.004)
July	$0.093^{\dagger}$	(0.004)
August	$0.094^{\dagger}$	(0.004)
September	$0.088^{\dagger}$	(0.004)
October	$0.090^{\dagger}$	(0.004)
November	$0.084^{\dagger}$	(0.004)
December	$0.075^{\dagger}$	(0.004)
Observations	488,143	
Adjusted R^2	0.591	
Residual Std. Error	0.463	

Notes: **p<0.01. † $p < 0.0001$. All regressions include year dummies reported in Figure 10, unreported ZIP code fixed effects, dummies for types of heating and air conditioning systems, and dummies for missing values of some attributes. Standard errors clustered at county-year level.

property was waterfront. We manually checked a sample of those properties reported to be waterfront and found it was quite accurate, but it was missing for a large majority of properties, including most of another sample that we manually identified as being waterfront. Hand-gathering this variable for the additional several hundred thousand properties was not a viable solution, but we were able to generate two additional useful proxies. The first is a flag for whether the distance from the Zillow-reported latitude and longitude of a property is within 75 meters of a lake or the ocean shoreline according to the NOAA

National Geodetic Survey 2023 vector shoreline data for New Hampshire.[24] The second is a flag for whether the property had an address on a relatively short street that went in close proximity to a lake or the ocean.[25] We then chose a small sample of properties, balanced across each of the seven categories defined by the values of the three waterfront flags, omitting the category where all flags were 0. In that small sample, we used Google Maps and other sources to determine manually whether we would consider each property to be waterfront. Using a logit model, we then regressed our hand-coded variable on the three waterfront flags and county fixed effects to determine their relative importance in predicting whether a property was, in fact, waterfront. We constructed our own omnibus variable *PWaterfront* as the predicted value from this regression for all properties that have at least one positive flag.[26] See Appendix Table 7 for coefficient estimates from this regression. We interpret *PWaterFront* as the predicted conditional probability of a property being on a waterfront given its county and its values for the three waterfront flags. As such, it will vary between 0 and 1. The predictions of the logit model were sufficiently sharp so that the resulting variable can be thought of as close to a dummy variable: over 86% of the values are exactly zero, another 9% are less than 0.2, and the majority of the values above 0.2 are at least 0.8.

5.3 Safegraph activity data

The second way in which we will estimate effects of vacation-rental sites on real estate markets is by comparing prices of homes that are more and less suitable for vacation rental use due to their proximity to other vacation amenities. The nearby vacation-amenity variable we construct for this purpose is based on our second substantial data source, comprehensive cellphone ping data from SafeGraph. We used these data to give us a richer map, varying by season, of proximity to vacation amenities across the state of New Hampshire. We are enormously grateful to SafeGraph for making these anonymized data available to researchers. Although their measures are aggregated over many cellphone users, they provide remarkable detail of patterns of movement over time, allowing us to create maps of amenities favored by tourists.

Let us first take a step back and describe the information contained in the SafeGraph data. SafeGraph collects cell-phone ping data—triangulating cell-phone users' locations

[24] We define a "lake" as any water body over 500,000 square meters according to the US Census Road and Area Water Shapefiles, and the shoreline data excludes the Isle of Shoals.

[25] More precisely, the property received a 1 if it had an address on a road that was shorter than 3/4 of a mile that, at some point, was within 65 meters of a "lake" or the coastline.

[26] *PWaterfront* is set to zero for properties that have all three flags equal to zero.

using pings off of cell towers—and processes these data to provide estimates of the number of people visiting each location in their database and how far visitors are on average from their home.[27] We will be using exclusively data pre-COVID, January 2018 to December 2019. Our data set includes monthly visit counts to many thousands of locations both in the state of New Hampshire and in nearby parts of Maine, Massachusetts, and Vermont. SafeGraph also provides a categorization for those locations. So, for instance, we have a measure of traffic at Gunstock Mountain Resort every month during that two-year period, and SafeGraph classifies it in the category of ski resorts and tells us its precise location. By comparing the number of visitors to different locations within a category, we can obtain a proxy for which areas are more appealing to vacationers. For example, in January of 2018, Gunstock Mountain Resort's visitor index is 2404 whereas that of the nearby Abenaki Ski Area was 318.[28] One limitation of the SafeGraph categories is that some of the categories are fairly broad, e.g. "All Other Amusement and Recreation Industries" (which we shorten to "Amusements") includes go-cart tracks, batting cages, laser tag facilities, disc golf courses, some local playgrounds, some farms, any many other types of businesses.[29] A second aspect that affects how we use the data is that the distance-from-home variable reflects how far people are from their home, not how far they traveled on their final drive to the location. For example, customers visiting the escape room near the the Waterville Valley ski resort are recorded as being on average more than 100 miles from home.

Our measure of each location i's proximity to vacation amenities that are valuable in season s will be constructed as a weighted average of the number of visits to each nearby vacation-relevant location j

$$\text{VacationAmenities}_{is} = \sum_j w_{ij} \text{Visits}_{\text{js}}$$

The weights w_{ij} are intended to reflect both (1) the relevance of amenity j to location i given the distance between the two, and (2) the degree to which visit counts at location j are indicative of location j being attractive to tourists. For example, a gas station may have as many daily visitors as a theme park, but we would not expect that proximity to a popular gas station would be as important to those searching for rentals on VRBO as proximity to a

[27] These data became famous and quite important during COVID because they provided a measure of how much economic activity had fallen off for different types of businesses, how good was compliance over time and across locations during lockdowns, and many other aspects of individual mobility.

[28] Abenaki is a small community-owned downhill facility. It has a vertical drop of 200 feet, is closed on Mondays and Tuesdays, and does not open until 4pm on Wednesday, Thursdays and Fridays.

[29] There are also a number of businesses that appear to be misclassified and we have only attempted to correct a few of the most visited ones.

theme park that draws the same number of visitors. To reflect these two forces, the weights will be constructed as a product of the importance of a particular category's locations to tourists and a geographical discount factor reflecting how far tourists are willing to drive to those locations: $w_{ij} = \beta_{c(j)}\delta_j^{d(i,j)}$. $\beta_{c(j)}$ is a measure of the importance to tourists of visits to various locations in location j's category, $c(j)$. δ_j is a per-mile discount factor reflecting how the value that tourists derive from being close to amenity j declines with distance, with $d(i,j)$ being the distance from home i to amenity j. The distance discounts would also be expected to differ across categories given different propensities to drive to activities of different types.

Before describing how we construct the components $\beta_{c(j)}$ and δ_j of the weights, it is useful to look at some summary statistics. Table 3 contains summary statistics on the categories of SafeGraph data that we use in our analysis. The first twenty-one rows of the table contain information on categories that we thought of as potential proxies for the presence of "vacation-relevant amenities": amenities that would help us predict the value that tourists would place on renting a home nearby. This could be because of a direct causal effect—people like renting homes adjacent to ski resorts—or because the presence of the amenity otherwise indicates that this is an area is one in which tourists like to stay—fudge shops in proximity could suggest that there are amenities not otherwise captured in our data that lead tourists to want to be in the area. The summary statistics are at the level of individual site in each category, and reflect averages across the 24 months in the data. The ski resorts and casinos have the largest average visitor counts (despite the limited ski season) and the bakeries and confectioneries have the smallest. Average distance-from-home is also largest for the ski resorts and smallest for the nature parks and bakeries.

Turning back to our construction of the measure of proximity to vacation-relevant amenities, we construct the required weights $w_{ij} \equiv \beta_{c(j)}\delta_j^{d(i,j)}$ in two steps. First, we chose the per-mile discount factor δ_j in an ad hoc manner, basing them on the Safegraph distance-from-home data as described in Appendix A.3. At the high end, the δ_j are 0.96 for ski resorts and above 0.9 for some of the major theme parks. At the low end, they are about 0.55 for bakeries, bars, and restaurants.[30] We then choose the category importance weights, $\beta_{c(j)}$, via a LASSO regression analysis. These regressions were run at the ZIP-code level,

[30] We think that allowing per-mile discount factors to vary across categories is important for capturing which areas would be attractive to vacation renters. Given what we do with the discount factors, we think their common scale matters less: we will eventually examine how prices changed for the houses that are better or worse located according to the index in regressions with zip-by-year fixed effects, and we will mostly report results using amenity indexes that have been standardized by dividing by their standard deviations.

Table 3: Summary Statistics for Activity Data

Visit Category	Number of Locations	Avg Monthly Visits Mean	Avg Monthly Visits St. Dev.	Avg Dist. from Home
Amenities				
Amusements	559	138	236	34
Arcades	16	170	162	32
Bakeries	294	47	73	18
Bars	465	112	127	29
Bookstores	227	155	272	24
Breweries	222	90	190	50
Casinos	10	827	804	27
Confectioneries	147	51	148	47
GasStations	1852	181	194	20
GolfCourses	299	270	758	28
HistoricSites	135	58	84	73
Marinas	48	56	49	39
Museums	179	122	262	56
MusicVenues	50	346	814	21
NatureParks	2541	197	561	17
Restaurants	6556	129	242	25
Sightseeing	36	65	104	51
SkiResorts	53	923	1295	125
SportsVenues	17	608	573	40
ThemeParks	183	152	680	67
Zoos+Botanic	34	94	142	33
Auxiliary Measures of Tourist Activity				
Hotels	358	180	220	127
RV&Campground	60	58	54	108

using the log of the number of VRBO-listed properties near each New Hampshire ZIP code as the dependent variable.[31] The right hand side variables were the aggregate activity relevant to the the ZIP in each of 21 categories c, i.e. $\sum_{j\in c}\delta_j^{d(z,j)}\text{Visits}_\text{j}$, where $d(z,j)$ is the distance from ZIP code z's centroid to the location j and Visits_j is the average across seasons of the seasonal visit count at loccation j. We also included the ZIP code's population and the log of the number of properties in the ZIP that were transacted at least once, so the the measure could reflect aggregate activity not only through the category-specific

[31] These counts were hand-collected by conducting searches on VRBO.com in late February of 2024 for a one week rental for two people for June 1-8, 2025. VRBO appears to return all listings that are within 10 miles of some location associated with the ZIP code. The count is top-coded at 300 for 7 of the 243 ZIPs due to the way VRBO returns results.

activity variables.

Table 4 presents estimates from two regressions. The left column reports OLS estimates from a regression run on all twenty-three RHS variables. We then ran a cross-validated LASSO regression to potentially select a subset of the variables. That procedure indeed selected just six of the amenity variables along with *log(Population)*. The second column reports estimates of an OLS regression run just on those variables. All RHS variables were standardized to have standard deviation 1, so the coefficient estimates can be taken as reflecting the importance of each variable.

Note that in the full regression in the first column, three of the RHS variables, those reflecting proximity to ski resorts, marinas, and other amusements, have particularly large coefficients and are significant at the 0.1% level. A few other variables are significant at the 5% level, with the nature parks, bakeries, and sightseeing activities having positive coefficients. As noted above, there are multicollinearities here and multiple ways that one could use data like these to identify locations frequented by vacationers. For example, the prevalence of VRBO listings near Hampton Beach is associated both with many people being at the beach itself (which is one of our "nature parks") or with many people being at the nearby arcades, marinas, restaurants, and sightseeing (e.g. fishing and whale watching) amenities. The LASSO regression ends up selecting all three variables that were significant at 0.1% level in the full regression, omitting all three variables that were significant just at the 5% level in the full regression, and selecting three other variables: activity at sightseeing, arcades, and theme parks. It also includes the log population at the ZIP level, which has a negative coefficient. One can think of that as partially offsetting the tendency of all activity-count variables to be larger in more populous areas.

Table 4: Estimates for Vacation Amenities Index

	Dependent var.: *log(VRBOListings)*			
	Full Model		Post LASSO	
	Coef.Est	(SE)	Coef.Est.	(SE)
S.Amusements	0.55***	(0.14)	0.17	(0.05)
S.Arcades	0.14	(0.08)	0.10	(0.06)
S.Bakeries	0.31*	(0.13)		
S.Bars	-0.40	(0.24)		
S.Bookstores	-0.15	(0.11)		
S.Breweries	0.16	(0.11)		
S.Casinos	0.00	(0.10)		
S.Confectioneries	-0.22*	(0.10)		
S.GasStations	-0.16	(0.16)		
S.GolfCourses	-0.19	(0.14)		
S.HistoricSites	-0.10	(0.09)		
S.Marinas	0.41***	(0.06)	0.33	(0.05)
S.Museums	-0.11	(0.11)		
S.MusicVenues	-0.15	(0.09)		
S.NatureParks	0.37*	(0.16)		
S.Restaurants	0.08	(0.31)		
S.Sightseeing	0.21*	(0.10)	0.41	(0.06)
S.SkiResorts	0.61***	(0.08)	0.79	(0.07)
S.SportsVenues	-0.0	(0.10)		
S.ThemeParks	0.11	(0.12)	0.06	(0.05)
S.Zoos+Botanic	-0.11	(0.07)		
log(Population)	-0.04	(0.08)	-0.10	(0.04)
log(Properties)	0.02	(0.06)		
Observations	242		242	
R^2	0.625		0.572	

Notes: Standard errors are in parentheses. $^{*}p<0.05$. $^{***}p<0.001$.

We want our vacation amenities index to reflect the ability of the owners of each property in our sample to extract additional value from the property by renting it through a home-rental website like AirBNB or VRBO. The predicted value from a regression like this seems a good candidate. Following the suggestion in Belloni and Chernozhukov (2013) we use coefficients from an OLS regression run on the variables selected by the LASSO procedure rather than the coefficients that come out of the LASSO procedure when doing this. Those coefficients are given in the second column of Table 6.[32]

Recalling the functional form for the index given above, coupled with results from the vacation amenities regression and our per-mile discount factors, we can construct a vacation amenities index for any location in the state of New Hampshire. Using its longitude and latitude, we compute its distance from every amenity, discounting visits by distance according to a category-specific discount rate and weighting them by category importance, and then add them up to obtain a single vacation amenities index number. Note, however, that we do have activity measures at a finer time-series division than the annual ones we used in the vacation amenities regression. In particular, we can use season-level numbers and construct the index for a particular property i in a particular season s.[33] The *Vacation-Amenities* variable varies tremendously across parts of the state. It also, however, varies, sometimes substantially, within a ZIP code. For example, a house located a block from Hampton Beach will be much closer to the amusements, arcades, marinas, and sightseeing locations located across the street from the beach and on the waterfront, than a house located a mile or two inland in from the beach.

To illustrate the variation in the vacation amenities index across seasons and across the broad geography of New Hampshire, Figure 6 presents heat maps showing the value that the *VacationAmenities* variable would take on for a house located at the centroid of each of New Hampshire's ZIP codes.[34] From left-to-right the maps in the figure show activity in the winter, spring, summer, and fall. Overall winter activity is lowest of any of the

[32]The reported standard errors are just those computed by OLS. We present them purely for interested readers. We are interested in this regression purely as a tool for prediction and have no interest in hypotheses about the relevance of the RHS variables to suitability for rental use, and hence have not calculated standard errors relevant to that exercise.

[33]Because the amenities regression was run on an annual dataset, predicted values will not necessarily align with the seasonal demand for vacation rentals. To align the index with vacation-rental demand we scale the predicted values up or down by season-specific scale factors chosen as the ratio between average log-visits to hotels and campgrounds in each ZIP in each season (which we also compute from Safegraph data) to the predicted values from an unscaled regression.

[34]Some parts of New Hampshire, e.g. uninhabited parts of the White Mountain National Forest, do not belong to any ZIP code. We also constructed the index at points within each such area to avoid having holes in the heat maps.

seasons[35] but, within winter, is highest in the White Mountain region, where many of the the largest ski resorts are located. For anyone who has visited the North Atlantic coast in the winter, it will not be surprising that activity along the coast is quite low. The third heat map is summer, which is the peak tourist season. Three other areas light up: the Seacoast (far southeast), the Lakes Region (near the middle of the state), and around Lake Sunapee (near the western border with Vermont). The index also takes on slightly higher values in the White Mountains in the summer than in the winter, reflecting that, although activity at the ski areas is lower, there is still substantial traffic at some of them with hiking, zip-lines, and the like, and that the region also has a number of theme parks and other amusements. Fall is a somewhat better tourist season than spring—visitors come to hike amidst the fall foliage, while the spring, known locally as "Mud Season" typically has less to offer. Downhill ski areas do remain open through March and early April, driving some of the spring traffic. These heat maps are interesting in their own right, but we mostly show them to provide validation that the method we used to construct the amenities index produces sensible results.

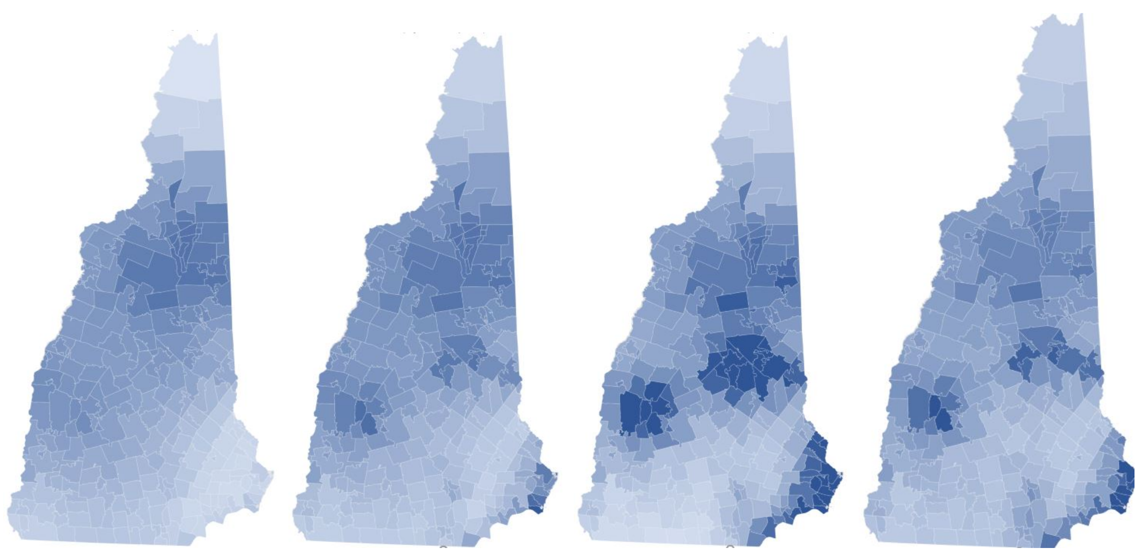

Figure 6: Activity at Vacation Amenities by Season

From left to right: Winter, Spring, Summer, Fall

As mentioned in the literature review, there have been a number of academic papers written using SafeGraph and other similar cellphone ping data, which discuss advantages

[35] The average value of the index across ZIPs is lowest in the winter, reflecting that we observe less activity at hotels then.

and limitations of the data. We will not reproduce those discussions here, but we will note that we are using data from the period of time when it was arguably most useful: a moment when 1) a large fraction of US adults were carrying smartphones and 2) smartphones had not yet adopted a default security feature that prevented the type of tracking that SafeGraph did.[36] Note that our SafeGraph data do not extend as far back as the Zillow data.[37] We do not consider this an important shortcoming. We would expect the amenities that tourists and vacationers care about—ski resorts, marinas, beaches, hiking trails, theme parks—to be quite stable over time, even decades. While we do not think that 20 years of SafeGraph data would be necessary to construct our map of tourist activity, using only one year of data might be less than ideal, though. Unusual weather in one year can have effects on skiing, beach-going, and other outdoor activities. Fortunately, we have two years of pre-COVID SafeGraph data, and the ski seasons and summer rainfall were fairly typical in those years—perhaps a bit more rainfall than normal at the beaches—so we feel that our data should provide an accurate measure of average tourist interest for the entire two decades of our analysis.

5.4 Home rental site traffic

Finally, we constructed a variable which represents overall traffic on the relevant online rental platforms, *log(PlatformUse)*. Figure 7 shows its evolution over time and also indicates the source data from which it was constructed. We did not have a single, reliable source of data that spanned the entire period of interest, so we collected four different data series, three from archived copies of the VRBO.com website and one from Google Trends, and used a regression model involving year-specific growth rates and month-of-year and data-source dummies applied to the partial series, some overlapping, to paste them together and construct our variable.[38] To create the figure, we graphed all of the constituent series, shifting them up or down for a consistent scaling.[39] You can see from the figure that even

[36] In other words, during the period of the data we use, it was possible for cell phone users to opt out of tracking, but no cell phones had that option as the default, so few users did.

[37] If they did, it would be a highly selected sample of consumers who went everywhere with their cellphones in 1996.

[38] VRBO changed its website design at various points in time. Early on it reported monthly user sessions. At other times it reports wordlwide listings and/or we can count New Hampshire listings. Our Google Trends data reflects searches for VRBO or AirBNB.

[39] Google Trends data has an artificial scale. Other series, such as number of listings in New Hampshire on VRBO or the number of monthly user sessions on VRBO, have entirely interpretable but different scales from each other. We are primarily interested in the trends, not the scale of the variable, but the units could be interpreted as those of the one series that is not rescaled: the log of total monthly user visits to vacation rental websites. Given that this series was only available from 1998 to 2001, we would not recommend taking the units too seriously as you move further away from that time frame to the extrapolated periods.

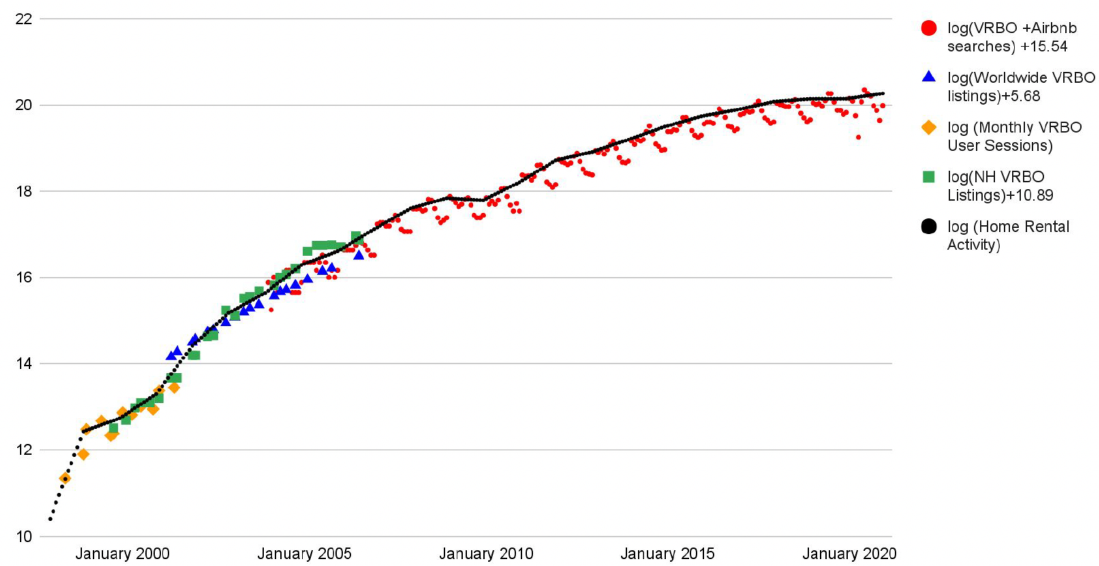

Figure 7: Index of Home Rental Platform Traffic, *log(PlatformUse)*

though the data are drawn from diverse sources, they tell a fairly consistent story.

In our main regression models examining prices of vacation and standard properties, we standardize some explanatory variables to facilitate interpretation of the coefficients on main effects and on interactions. For example, we define $S.VacationAmenities$ and $S.PlatformUse$ by subtracting off the means of the $VacationAmenities$ and *log(PlatformUse)* variables and dividing by their standard deviations. When interpreting some results it will be useful to keep in mind that the scale of the standardized $S.PlatformUse$ variable ranges from -2.27 at the start of the sample to 1.08 at the end.

6. Effects of vacation-rental platforms

In this section we report our main results of interest: estimates of the effects of vacation rental platforms on prices in vacation and standard markets. The first section presents the evidence that exploits our cleanest source of variation at a small geographic scale: the comparison of waterfront and non-waterfront properties. The second subsection discusses results examining our vacation amenities indexes.

6.1 Evidence from waterfront properties

Table 5 contains estimates of the waterfront-related coefficients from two transaction-level hedonic regressions. Each regression includes all of the property characteristic variables in the base hedonic regression we reported in Table 2 and ZIP-code-by-year fixed effects. The regression in the first column implements the difference-in-differences strategy we described in the Section 4.2. It uses ZIP-code-by-year fixed effects to control for all factors affecting average prices in a narrowly defined area at a particular point in time. It uses three variables *WaterFront*, *WaterFront* × *S.DistBoston*, *WaterFront* × *CountyMeanWF* to estimate the average premium paid for waterfront properties, allowing some variation across different parts of the state. And the main variable of interest, *WaterFront* × *S.PlatformUse*, measures how the waterfront premium has changed as home-rental websites have grown in popularity.

Table 5: Effects of home rental platform growth: evidence from waterfront properties

	Dependent var.: *log(SalesPrice)*	
	(1)	(2)
WaterFront	0.483***	0.479***
	(0.018)	(0.018)
WF × *S.DistBoston*	0.192***	0.190***
	(0.020)	(0.020)
WF × *CountyMeanWF*	0.693**	0.757*
	(0.232)	(0.234)
WF × *S.PlatformUse*	0.076***	0.099***
	(0.007)	(0.013)
WF × *S.PlatformUse* × *CountyMeanWF*		-0.283*
		(0.122)
Observations	488,144	488,144
Adjusted R^2	0.617	0.617
Residual Std. Error	0.448	0.448

Notes: *p<0.05. **p<0.01. ***p<0.001. All regressions include unreported ZIP-by-year and month-of-year fixed effects, dummies for types of heating and air conditioning systems, and dummies for missing values of various attributes. Standard errors clustered at county-year-waterfront level.

The coefficient on the *Waterfront* variable indicates that waterfront properties at the sample-mean distance from Boston are estimated to sell for 48 log points (≈ 62%) more than non-waterfront properties with similar characteristics in the same ZIP code. The coefficient on the interaction between the *Waterfront* variable and the standardized distance-to-Boston variable indicates that the waterfront premium is larger in areas farther from Boston, presumably reflecting that prices in the standard (non-vacation) submarket are higher in the parts of the state closer to Boston, which narrows the waterfront premium. The coefficient on the interaction between the *Waterfront* variable and a variable giving the fraction of transacted properties in the county that are on the waterfront is also positive.[40] It suggests that waterfront premia are 8-10% higher in the Lakes Region than elsewhere, presumably reflecting that the lakes there are better on some dimensions (e.g. being larger and with cleaner water) and/or have more complementary amenities (e.g. bait shops and lakeside restaurants).

The primary coefficient of interest in this regression is that on the interaction between the *Waterfront* dummy and the standardized *log(PlatformUse)* variable. It indicates that the waterfront premium went up by 7.6 log points with each standard deviation increase in the *log(PlatformUse)* variable. Recall that the standardized version of this variable ranges from -2.27 to 1.08, so it implies that the waterfront premium has grown by about 25 log points (an $\approx$ 29% price increase) as home rental sites have grown. This is a very large increase. Recall that we argued that this should be a lower bound on the effect of home-rental platforms on vacation-home prices given that the increase in the waterfront premium will not reflect the portion of the increase in the price of waterfront homes that has spilled over to non-waterfront properties.

The regression in the second column uses the triple-difference strategy we discussed to estimate the degree to which the expansion of home-rental websites have also increased the prices of non-waterfront properties. Recall that this strategy involved examining whether the vacation-home premium grew less slowly in markets with more potential vacation properties, with the motivation being that spillovers to standard property prices will be larger there, offsetting part of the increase in vacation-home prices. We implement this strategy by adding the interactions between the *Waterfront* $\times$ *S.PlatformUse* variable and the *CountyMeanWF* variable, which we are thinking of here as a proxy for the fraction of potential vacation homes in the market.[41] The estimated coefficient of -0.28 on the three-way inter-

[40] The *CountyMeanWF* variable is 14-16% in the "Lakes Region" counties (Belknap and Carroll) and between 0.5% to 4.5% in the others.

[41] Second homes have been reported to comprise about 29% and 44% of 2021 sales in the Lakes Region counties (Belknap and Carroll) and to be just 1.7% of 2021 sales in the Hillsborough country, in which we estimate that fewer than 1% of transacted properties are on waterfront.

action suggests that the waterfront premium has indeed been less sensitive to the growth of home-rental platforms in counties with more waterfront properties: the estimates imply that a one-standard deviation increase in the *PlaformUse* variable is associated with a 10 log point increase in the waterfront premium in counties with very few waterfront properties and an 0.06 log point increase in the counties with the most waterfront properties. Thus suggests that about 40% of the increase in vacation home prices may have spilled over to standard home prices in those counties. Over the full 24 year period, this corresponds to a total increase in standard home prices of about 15% in the Lakes Region counties. Home-rental-platform-driven increases in non-vacation prices in other counties would be less than 5%.

Given that our regression estimates will reflect the patterns in how the waterfront price premium has covaried with increases in home-rental platform usage, and both changes occur over a long time horizon, it seems useful to look in further depth at how the waterfront premium has evolved. To this end, we estimate a third version of our regression, similar to the specification in column (2), but replacing the *Waterfront* $\times$ *S.PlatformUse* variable with a full set of twenty four *Waterfront* $\times$ Year Dummy interactions.[42] Figure 8 plots the point estimates of the year-specific waterfront premia. Note that the time-series pattern looks very much like the growth in the usage of home-rental traffic in Figure 7. There is perhaps a small dip in the great recession, but otherwise the pattern shows steady growth from 1996 through 2011. A notable difference in the patterns is what has happened since 2011. While the growth of vacation-rental platforms has slowed, they do continue to grow, whereas the waterfront premium was roughly flat from 2011 to 2014 and then declined from 2014 to 2019. One potential reason for the difference is a difference between what we are actually estimating and the ideal estimating equation we described in Section 4: the model envisaged the use of a vacation-home indicator whose coefficient would capture the difference in prices between vacation and standard homes. In our application we think that almost all of the waterfront homes are potential vacation homes for which one must outbid those in the vacation-home submarket. But it is not true that all non-waterfront homes are non-vacation homes. Many vacationers prefer the quiet of the woods over proximity to water. Indeed, people in the Lakes Region discuss informally that vacation-rental platforms have grown to the point that they list near-water properties in addition to on-water properties. If vacation-rental platforms had already reached a level of popularity by 2011 sufficient to allow owners of waterfront homes to extract almost the full asymptotic value of their

[42] We also included the *Waterfront* $\times$ *CountyMeanWF* interaction.

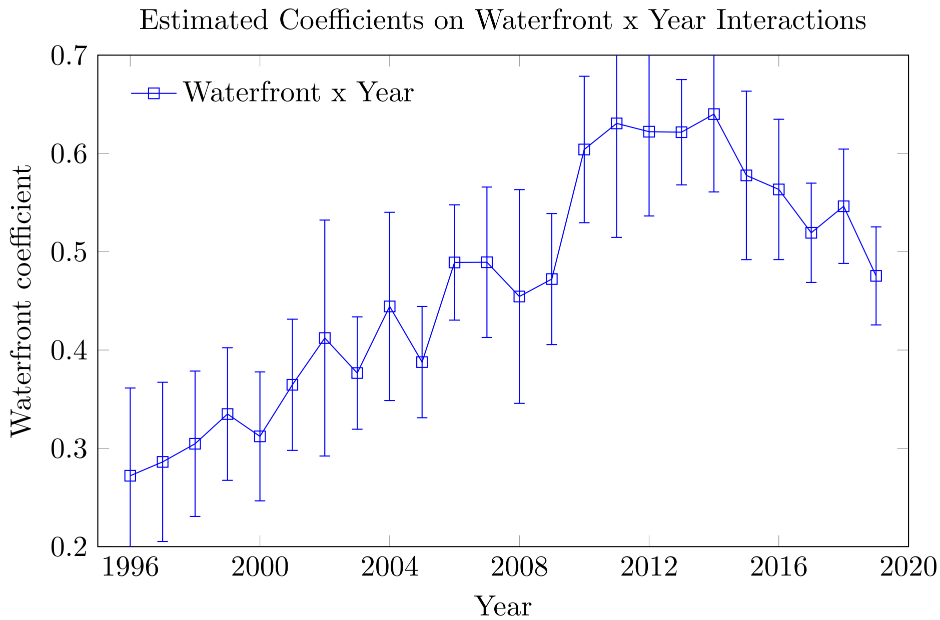

Figure 8: Estimates on interactions between *Waterfront* variable and year dummies with 95% confidence intervals.

properties, but owners of non-waterfront properties continued to receive increased benefits due to the thinner markets for their homes, this could reduce the vacation premium as measured by the waterfront indicator.

6.2 Evidence from proximity to vacation amenities

We now turn to the second way in which we will examine the effect of home-rental platforms on vacation and standard homes: examining the differential trends for properties with high and low values of our *VacationAmenities* index. The vacation amenities index will be a noisier measure of the likelihood that a property is in vacation home market, and it does not vary nearly as much within a ZIP code as our waterfront measure, so we would expect estimates in this section to be attenuated by the measurement error and less precise than those in the previous section. Nonetheless, the estimates in this section do provide additional evidence of the effects of home-rental websites on New Hampshire home prices.

The first column of Table 6 presents coefficient estimates from a regression that again implements our differences-in-differences identification strategy. The regression includes ZIP-code-by-year fixed effects (and all of the variables in the regression in the final column of Table 5.) It again uses three variables $S.VacationAmenities$, $S.VacationAmenities \times$

S.DistBoston, *S.VacationAmenities* × *CountyMeanVA* to estimate the premium paid for properties located closer to vacation amenities, allowing some variation across different parts of the state. It indicates that a property that is one standard deviation better in proximity to vacation amenities (and at the mean distance from Boston) sells for 5.5% more. Again the vacation premium varies as expected with the distance to Boston: the vacation premium would be just 0.6% in an area one-standard deviation closer to Boston than the mean, and 10.3% in an area that is one standard deviation farther from Boston than the mean.

Table 6: Effects of home rental platform growth: evidence from vacation amenities index

	Dependent variable: *log(SalesPrice)*		
	(1)	(2)	(3)
S.VacationAmenities	0.054***	0.054***	-0.023***
	(0.015)	(0.015)	(0.012)
S.VA × *S.VRFrac*			0.055***
			(0.005)
S.VA × *S.DistBoston*	0.049***	0.049***	0.079***
	(0.009)	(0.009)	(0.008)
S.VA × *CtyMeanVAmen*	0.023*	0.024*	-0.039***
	(0.009)	(0.009)	(0.009)
S.VA × *S.PlatformUse*	0.028***	0.033*	0.012
	(0.008)	(0.014)	(0.014)
S.VA × *S.PlatformUse* × *S.VRFrac*			0.012**
			(0.004)
S.VA × *S.PlatUse* × *CtyMnVAmen*		-0.004	-0.013*
		(0.007)	(0.005)
Observations	488,143	488,143	488,143
Adjusted R^2	0.619	0.619	0.620
Residual Std. Error	0.447	0.447	0.446

Notes: *p<0.05. **p<0.01. ***p<0.001. All regressions include all of the variables from the hedonic regression in Table 2 and the regression in column (2) of Table 5. Standard errors clustered at the county-year level.

The primary coefficient of interest is that on the interaction between the vacation amenities index and the platform use variable. The positive and significant coefficient indicates that the proximity-to-vacation-amenities premium grew as home-rental platforms became more popular. Together with the other coefficients, it indicates that in areas at the sample-mean distance from Boston, the price premium associated with being one standard deviation better in proximity to amenities was -1% at the start of the sample and grew to 11% at the end of the sample.

The regression in the second column uses the triple-difference strategy to examine whether the expansion of home-rental websites also increased the prices in the standard market. We implement this strategy by adding the interactions between the $S.VacationAmenities \times SPlatformUse$ variable and the $CountyMeanVAmen$ variable, which we are again thinking of as a proxy for the prevalence of vacation homes in the market. The point estimate on this variable is negative, but it is far from significant and the standard error is such that we also cannot conclude that price spillovers to standard homes are "small."

As in the waterfront analysis, we can again look directly at how the proximity-to-vacation-amenities premium has evolved over time by estimating a third version of our regression, similar to the specification in column (2), but replacing the $S.VacationAmenities \times S.PlatformUse$ variable with a full set of twenty four $S.VacationAmenities$ $\times$ Year Dummy interactions. Figure 9 plots the point estimates of the year-specific proximity-to-amenities premia. The yearly estimates are much less precise relative to their magnitude than were the year-by-year waterfront premium estimates. But the time series pattern of the point estimates is quite similar. It is mostly increasing from 1996 through 2014. And again we see a decline from 2014 through 2019. The overall pattern seems consistent with the growth of home rental sites being a driving force and does not suggest that the premia are instead aligned with housing prices in general. But the estimates are noisy enough that we cannot draw strong conclusions.

The seasonal variation we identified in constructing the vacation amenities index provides an additional potentially valuable opportunity to distinguish effects of the growth of home-rental websites from changes in tastes. A number of the homes in our sample have sufficiently little rental appeal that it would be challenging to find anyone who wants to rent them on VRBO at any time of the year. Many more would probably be hard to rent out except perhaps in the summer high season. Given the fixed costs that must be incurred to use a house as a rental (insurance, packing up possessions, arranging check-in and cleaning services, etc.), one would expect there to be nonlinearities in the potential benefits that vacation rental platforms provide to houses with different levels of vacation amenities: it is only vacation-attractiveness beyond some fixed-cost-determined threshold that should matter for vacation-rental use. As a result, the value of proximity to vacation amenities when used as a vacation rental will not be as aligned with the value of proximity to vacation amenities to a full-time occupant, and we can hope to exploit this additional

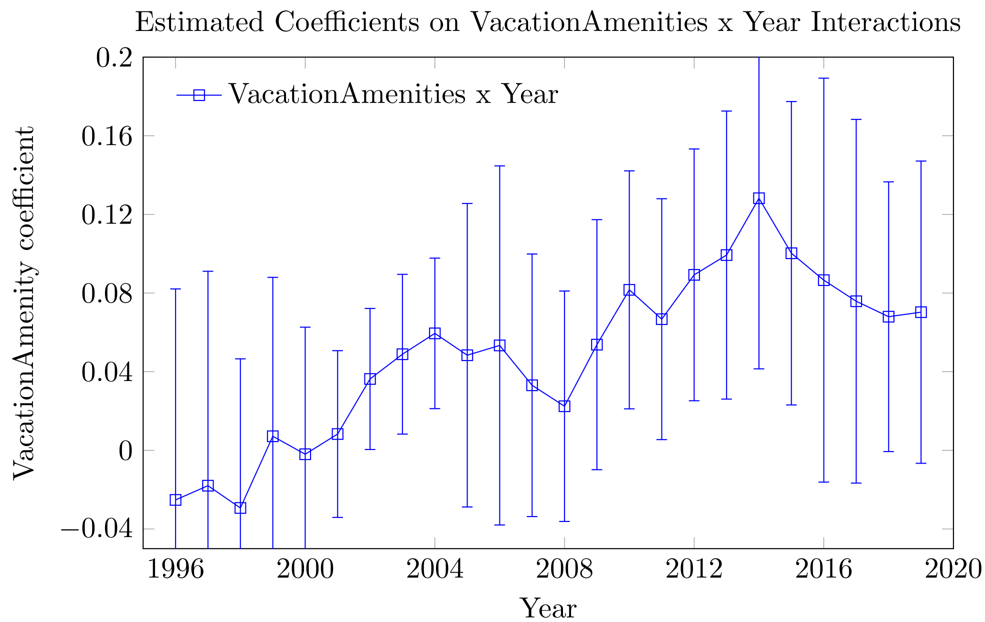

Figure 9: Estimates on interactions between *VacationAmenities* variable and year dummies with 95% confidence intervals.

source of information.

To formalize this idea, we constructed a variable intended to proxy for how well suited a home's proximity to vacation amenities was to exploitation as a vacation rental. Specifically, we defined

$$\text{VRAmenitiesFraction}_i = \frac{\sum_s \min(\text{VacationAmenities}_{is} - c, 0)}{\sum_s \text{VacationAmenities}_{is}},$$

where we set c, interpreted as the minimum level of vacation amenities that a property must have to be rentable given costs, was chosen so that about 25% of properties have a nonzero rental value in the winter and 55-60% have a nonzero rental value in the summer and fall. The resulting variable has a median of just 0.05, but is equal to 0.67 for the properties it regards as best-suited exploitation as vacation rentals. Large values occur for properties with high value of the vacation amenities index in multiple seasons—properties in the Lakes Region that are also close to the Gunstock Ski Resort, for example—and for some properties that have almost all of their value in one season—properties on coastal New Hampshire. The third column of Table 6 reports coefficient estimates from regressions that add this variable's interactions with the *S.VacationAmenities* variable and its interaction with the growth of home rental platforms. The coefficient on the *S.VacationAmenities*

$\times$ *VRAmenitiesFraction* interaction is positive and highly significant, indicating that the premium for proximity to vacation amenities is larger when the seasonal distribution of the amenities is well suited to being exploited via vacation rental. The positive significant coefficient on the textitS.VacationAmenities $\times$ *S.PlatformUse* $\times$ *VRAmenitiesFraction* indicates that the proximity-to-vacation amenities premium grew more rapidly as home rental sites became more popular for properties that were more exploitable as vacation rentals. Finally, another interesting coefficient in this regression is the triple-interaction in the final row, which again is the term that the triple-difference strategy suggests can capture price increases having spilled over to standard properties. It becomes larger than it was in the previous column, and suggests that a bit less than half of the home-rental-site driven price increases have spilled over to standard homes.

7. Conclusion

The New Hampshire real estate market has a number of features that make it a nice setting in which to study the effect that the growth of VRBO, AirBNB, and other vacation home rental websites has had on the markets for primary and vacation homes. These features include the large number of vacation homes in the state, the substantial variation in the share of vacation homes in different parts of the state, the fact that we can identify some properties as being much more likely than others to be suitable vacation homes, and the up-and-down pattern that the real estate market experienced during the Great Recession, reducing its correlation with the growth of vacation rental websites.

In this paper we have exploited detailed data from a number of different sources to measure items of interest. We extracted and combined information from Zillow's ZTRAX databases to build a nearly universal dataset of New Hampshire transactions going back to when vacation rental websites were first being introduced. We combined data from several sources to build a measure of how the usage of vacation rental websites grew. We develop a novel method for measuring the proximity of properties to vacation (and other) amenities that are important in each season of the year, exploiting cell phone ping data. And we combine data on property locations with data from the Census Bureau and NOAA to identify properties with vacation-relevant waterfront locations.

A simple economic model suggests that, given the type of information we have, one can estimate how the growth of vacation rental websites has affected prices for New Hampshire homes in a fairly straightforward manner. Regressions with ZIP-code-by-year fixed effects let us measure the price gap between waterfront and nearby non-waterfront properties. We find that this gap has gotten substantially larger as vacation rental sites grew in popularity, increasing from roughly 35% to roughly 75%. Two findings indicate that some of the increase has spilled over and raised the price of standard properties: our triple-difference estimates indicate that the measured waterfront premium grew more slowly in areas where spillovers should be larger, and the premium has been declining in the most recent pre-pandemic years. But the estimates suggest that spillover price increases in standard properties are much smaller, perhaps on the order of forty percent as large as the effect on vacation-home prices, even in the areas with the largest number of vacation homes.

A second analysis uses similar methods to explore price changes for properties with proximity to vacation amenities such as ski resorts, marinas, sightseeing businesses and amusement parks. The magnitudes of the effects we identify here are smaller, but, qualitatively, the results are similar. We estimate that the premium paid for homes proximate to vacation amenities grew as home rental sites grew in popularity. And the increases are particularly clear for properties whose seasonal distribution of amenities makes them well suited to being rented on home rental websites.

References

Alonso, W. (1964). Location and Land Use: Toward a General Theory of Land Rent. Harvard University Press, Cambridge.

Baldauf, M., Garlappi, L., and Yannelis, C. (2020). Does climate change affect real estate prices? only if you believe in it. The Review of Financial Studies, 33(3):1256–1295.

Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. Bernoulli, 19(2):521–547.

Bishop, K. C. and Murphy, A. D. (2011). Estimating the willingness to pay to avoid violent crime: A dynamic approach. American Economic Review, 101(3):625–629.

Black, S. E. (1999). Do better schools matter? parental valuation of elementary education. The quarterly journal of economics, 114(2):577–599.

Boeing, G. (2017). A multi-scale analysis of 27,000 urban street networks. Manuscript submitted for publication. doi, 10.

Brelsford, C., Moehl, J., Weber, E., Sparks, K., Tuccillo, J. V., and Rose, A. (2022). Spatial and temporal characterization of activity in public space, 2019–2020. Scientific Data, 9(1):379.

Brueckner, J. K. (1987). The Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model. Elsevier Science, Amsterdam.

Chay, K. Y. and Greenstone, M. (2005). Does air quality matter? evidence from the housing market. Journal of political Economy, 113(2):376–424.

Chen, M. K., Chevalier, J. A., and Long, E. F. (2021). Nursing home staff networks and covid-19. Proceedings of the National Academy of Sciences, 118(1):e2015455118.

Chen, M. K. and Rohla, R. (2018). The effect of partisanship and political advertising on close family ties. Science, 360(6392):1020–1024.

Chiou, L. and Tucker, C. (2020). Social distancing, internet access and inequality. Technical report, National Bureau of Economic Research.

Cunha, A. M. and Lobão, J. (2022). The effects of tourism on housing prices: applying a difference-in-differences methodology to the portuguese market. International Journal of Housing Markets and Analysis, 15(4):762–779.

Davis, L. W. (2004). The effect of health risk on housing values: Evidence from a cancer cluster. American Economic Review, 94(5):1693–1704.

Ellison, G. and Ellison, S. F. (2024). Match quality, search, and the internet market for used books. Technical report, Massachusetts Institute of Technology.

Ellison, G., Fudenberg, D., and Mobius, M. (2004). Competing auctions. Journal of the European Economic Association, 2(1):30–66.

Epple, D. (1987). Hedonic prices and implicit markets: estimating demand and supply functions for differentiated products. Journal of political economy, 95(1):59–80.

Farronato, C. and Fradkin, A. (2018). The welfare effects of peer entry in the accommodation market: The case of airbnb. Technical report, National Bureau of Economic Research.

Gao, S., Rao, J., Kang, Y., Liang, Y., and Kruse, J. (2020). Mapping county-level mobility pattern changes in the united states in response to covid-19. SIGSpatial Special, 12(1):16–26.

Gehrig, T. (1998). Competing markets. European Economic Review, 42(2):277–310.

Glaeser, E. L. (2008). Cities, Agglomeration, and Spatial Equilibrium. Oxford University Press, Oxford.

Glusac, E. (2020). The future of airbnb.

Juhász, L. and Hochmair, H. H. (2020). Studying spatial and temporal visitation patterns of points of interest using safegraph data in florida.

Jullien, B., Pavan, A., and Rysman, M. (2021). Two-sided markets, pricing, and network effects. Handbook of Industrial Organization, 4(1):485–592.

Kahn, M. E. (2021). Racial and ethnic differences in the financial returns to home purchases from 2007 to 2020. Technical report, National Bureau of Economic Research.

Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. American Economic Review, 57(2):197–210.

Morrison, J. (2023). A primer on new england's second-home markets — plus when and where people are buying.

Muth, R. F. (1969). Cities and Housing: The Spatial Pattern of Urban Residential Land Use. University of Chicago Press, Chicago.

Nolte, C., Boyle, K. J., Chaudhry, A. M., Clapp, C. M., Guignet, D., Hennighausen, H., Kushner, I., Liao, Y., Mamun, S., Pollack, A., et al. (2021). Studying the impacts of environmental amenities and hazards with nationwide property data: best data practices for interpretable and reproducible analyses. WVU College of Law Research Paper, (2021-013).

Palmquist, R. B. (1984). Estimating the demand for the characteristics of housing. The Review of Economics and Statistics, pages 394–404.

Perloff, J. and Salop, S. (1983). Equilibrium with product differentiation. Review of Economic Studies, 52(1):107–120.

Poursaeed, O., Matera, T., and Belongie, S. (2018). Vision-based real estate price estimation. Machine Vision and Applications, 29(4):667–676.

Roback, J. (1982). Wages, rents, and the quality of life. Journal of political economy, 90(6):1257–78.

Rochet, J.-C. and Tirole, J. (2006). Two-sided markets: a progress report. RAND Journal of Economics, 37(3):645–667.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. Journal of political economy, 82(1):34–55.

Rosen, S. (1979). Wages-based Indexes of Urban Quality of Life. Johns Hopkins Univ. Press.

Rysman, M. (2009). The economics of two-sided markets. Journal of Economic Perspectives, 23(3):125–143.

Sheppard, S. (1999). Hedonic analysis of housing markets. Handbook of regional and urban economics, 3:1595–1635.

Wheaton, W. C. (1974). A comparative static analysis of urban spatial structure. Journal of Economic Theory, 9:223–237.

White, M. (2023). Once an evangelist for airbnbs, she now crusades for affordable housing.

Witte, A. D., Sumka, H. J., and Erekson, H. (1979). An estimate of a structural hedonic price model of the housing market: an application of rosen's theory of implicit markets. Econometrica: Journal of the Econometric Society, pages 1151–1173.

Wolinsky, A. (1984). Product differentiation with imperfect information. Review of Economic Studies, 51(1):53–61.

Zervas, G., Proserpio, D., and Byers, J. W. (2017). The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry. Journal of marketing research, 54(5):687–705.

Appendix A

A.1 Year dummies in hedonic regression

Figure 10 graphs the estimated year dummies from the hedonic regression described in table 2 with 95% confidence bands.

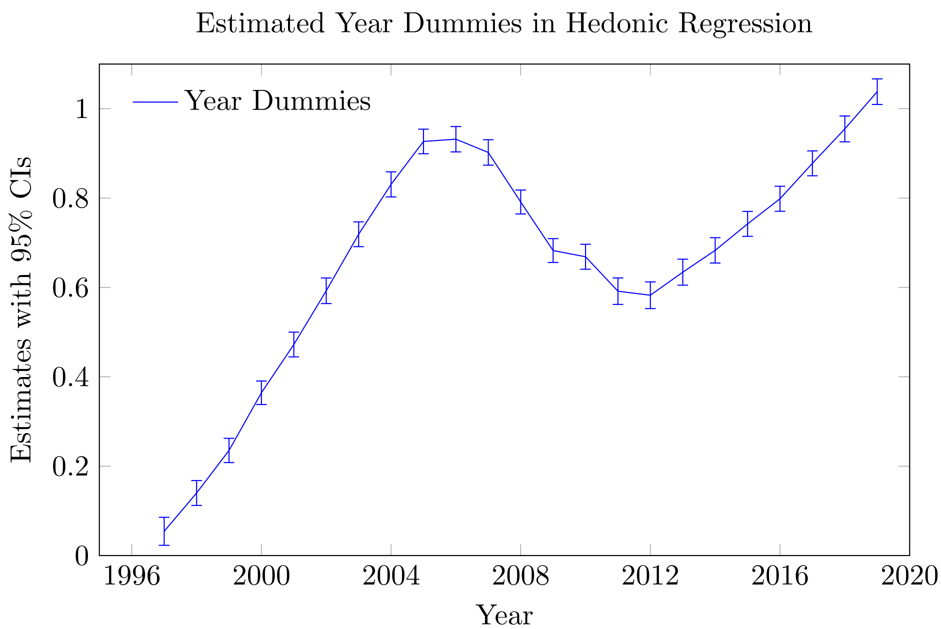

Figure 10: Hedonic estimates of log-price changes for New Hampshire homes 1996-2019 with 95% confidence intervals.

A.2 Waterfront variable regression

The table low presents coefficient estimates from the regression used to generate the *WaterFront* variable.

A.3 Distance discounting

In most categories, we use the same discount factor at all locations in the category setting $\delta_j = e^{-\kappa/\hbar_{c(j)}}$, where $\hbar_{c(j)}$ is the median value of the distance-from-home variable for locations in the category. In three of the categories (amusements, theme parks, and nature parks) we allow distance-discounting to vary across locations within a category. The motivation for this was that in the three categories we felt both that there was substantial heterogeneity in what SafeGraph included in the category and that the (noisy) distance-from-home variable contained useful information that would help us to better re-

Table 7: Logit regression used to construct *WaterFront* variable

	Dependent variable: *VerifiedWaterfront*
ZillowWF	3.118***
	(0.736)
RoadsWF	0.261
	(0.496)
WaterProximity	4.270***
	(0.754)
County dummies	Yes
Observations	210
Log Likelihood	−64.900

Notes: *p<0.1; **p<0.05; ***p<0.01

flect the how tourists would value being at some distance from a location. For example, the "amusements" and "nature parks" categories both include a number of generic neighborhood playgrounds, which can get a lot of visitors, but which have almost all of their visitors coming from very nearby. In these categories we used the geometric mean of the location j's distance-from-home and the category median in place of $\hbar_{c(j)}$. [43]

Table 8 lists the three most popular locations within each of the six vacation amenity categories that were selected for inclusion in the construction of the *VacationAmenities* by our LASSO procedure. The table also lists indexes of the activity levels at the locations and the distance-from-home of visitors taken directly from the SafeGraph data, and the per-mile discount factor that we apply when determining their relevance to nearby homes. Two of the categories, "Amusements" and "Theme Parks," are categories where we allowed the per-mile discount factor to vary within category. Note, for example, that the second most-visited "Amusement" (which is our shorthand for Safegraph's "All Other Amusement and Recreation Industries") is a nondescript local playground.[44] The location-specific discount factor results in this playground having little impact on the amenity index except for houses that are very close to it.

[43] To avoid having extreme values we limited the distance used to be within a factor of three of the category median.

[44] Given how nondescript the playground is in Google Streetview we suspect that this particular playground is a data error, but there are many other playgrounds in this category and the "Nature Park" categories that we think are truly often visited.

Table 8: Categories in *VacationAmenities* index: Most active New Hampshire locations

Category	Top NH locations	Activity index	Distance index	Per-mile discount
Amusements	Mel's Funway Park, Litchfield	698	17	0.70
Amusements	Belvedere Playground, Nashua	598	2	0.36
Amusements	FieldHouse Sports, Bow	577	17	0.70
Arcades	Playland Arcade, Hampton Beach	322	50	0.68
Arcades	Pinball Wizard Arcade, Pelham	229	8	0.68
Arcades	Break Free 603, Waterville Valley	225	174	0.68
Marinas	Wolfeboro Corinthian Yacht Club	154	42	0.73
Marinas	Akwa Marina Yacht Club, Laconia	120	77	0.73
Marinas	Great East Docks, Sanbornville	108	9	0.73
Sightseeing	AG Fishing & Whale Watching, Hampton Bch	180	37	0.85
Sightseeing	MV Kearsarge Restaurant Ship, Sunapee	157	32	0.85
Sightseeing	Mount Washington Cruises, Laconia	89	86	0.85
Ski Resorts	Gunstock Mountain Resort, Gilford	1623	63	0.96
Ski Resorts	Loon Mountain, Lincoln	1574	192	0.96
Ski Resorts	Attitash Mountin Resort, Bartlett	1543	174	0.96
Theme Parks	Canobie Lake Park, Lincoln	8754	25	0.82
Theme Parks	Story Land, Glen	1311	105	0.91
Theme Parks	Kahuna Laguna, North Conway	869	168	0.93