The Rise and Effects of Homeowners Associations

Wyatt Clarke and Matthew Freedman*

April 2019

Abstract

In the U.S., nearly 60 percent of recently built single-family houses, and 80 percent of houses in new subdivisions, are part of a homeowners association (HOA). We construct the first near-national map of HOAs using publicly recorded mortgage records for single-family homes. We use these data to document the growth and characteristics of HOAs as well as to examine their relationship with housing prices. We find that houses in HOAs have prices that are on average at least 4 percent, or $13,500, greater than observably similar houses outside of HOAs. The HOA premium correlates with the stringency of local land use regulation, local government spending on public goods, and measures of social attitudes toward race. The data also paint a detailed picture of the people living in HOA neighborhoods, who are on average more affluent and racially segregated than those living in other nearby neighborhoods.

Keywords: homeowners associations, planned unit developments, capitalization, property values, private government, mortgages
JEL: H1, R3, R5

*Clarke: IBM (email: wyatt.clarke@gmail.com). Freedman: University of California, Irvine (e-mail: matthew.freedman@uci.edu).

1 Introduction

Most single-family houses constructed in the United States in recent years are part of a *common interest development* ("CID"), governed by a *homeowners association* ("HOA"). While rare until the 1960s, HOAs have come to house a fifth of Americans, a proliferation that Guberman (2004) called "one of the most significant privatizations of local government functions in history." HOAs are not as prevalent outside the U.S., but comparable privately governed communities now exist in a number of countries, including Canada, the United Kingdom, Australia, Japan, and China (McKenzie, 2011; Glasze et al., 2006; Wang, 2014). Popular and academic opinions are divided over whether homeowners actually like to live in HOAs. Some contend that HOAs represent a sensible market solution to local public goods problems (Foldvary, 1994; Barton and Silverman, 1994; Nelson, 2005), while others see HOAs as a sort of unregulated hostage crisis with unwitting homeowners harassed by busybody neighbors (McKenzie, 1994; Lucas, 2013; Benson and DeBat, 2014).

Economists typically assess residents' opinions about the value of local amenities or disamenities, such as school quality, air quality, property tax levels, or the presence of an HOA, by examining house prices. If the presence or quality of one of these features raises the price of a house, it indicates that the marginal buyer values that feature, other characteristics of the house being held constant. Households might be willing to pay more to live in a community with an HOA because HOAs provide tailored amenities and can enforce another layer of zoning, use restrictions, and minimum quality maintenance on top of that which local governments provide.[1] To the extent that HOAs screen on household attributes or prohibit activities more commonly undertaken by individuals in certain groups, they could also be used as a tool for exclusion and achieving relative homogeneity of residents within a local area. However, not only are HOA residents obliged to pay regular dues, but they also

[1] An annual survey of large HOA communities conducted by the Community Associations Institute indicates that common amenities and services include sports and recreational facilities, enforcement of building restrictions, disaster preparedness and mitigation, gated access to the community, ownership of streets, social programming, landscaping, and maintenance of natural areas (Community Associations Institute, 2016*b*).

must forsake some degree of personal control over their property. Whether the perceived benefits of living in a community with an HOA outweigh these costs will determine whether the presence of an HOA will tend to increase or decrease house prices.

This paper presents the first ever (nearly) national estimate of how HOAs affect single-family house prices, using public data curated by Zillow, Inc.[2] The data include prices, HOA membership indicators, and other building characteristics for 34 million housing transactions, observed for 44 U.S. states over 35 years.[3] Using a methodology that could be useful in a range of housing market applications, we derive HOA membership indicators from "planned unit development" and "condominium" riders described in Zillow's collection of public mortgage records. The size of the sample allows us to estimate HOA premiums for specific locations, which can be usefully compared to each other since they are calculated using consistent data elements and methodology. Market-specific estimates also allow us to explore which local factors influence the value of living in an HOA. Our estimation approach is based on hedonic theory à la Rosen (1974). Our baseline estimates of the HOA premium control for a rich set of housing unit characteristics and are identified off variation in HOA status among homes in close proximity. We also consider a flexible, semi-parametric generalized additive model (GAM) specification, which is common in the hedonic literature but has not yet been applied to HOAs. In another first for this literature, we also characterize HOA fees based on data we collected from Redfin.com, an online real estate listing site that serves as a complementary data source.

The paper's overarching conclusion is that single-family houses with an HOA generally sell at a premium, but with substantial local variation. On average, buyers pay at least 4 percent, or $13,500, extra for a house that is subject to an HOA. This is near the middle of the range of estimates from the existing literature on HOAs, which has generally focused on

[2]Although Zillow's data describe both single- and multi-family housing, we only study the relationship between house price and HOA membership for single-family housing. The reason is that sales of HOA and non-HOA multi-family homes cannot be compared as directly as sales of HOA and non-HOA single-family homes, as discussed in Section 3.

[3]We exclude states in New England, where the requisite data are not reported.

a single state or city at a time because data on these associations has been scarce. Assuming our rich set of controls limits the scope for unobservable differences between HOA and non-HOA homes to influence prices (an assumption we test in the data), this price premium implies that each dollar paid to an HOA in dues buys about $1.19 worth of benefits.[4]

Additionally, we find that the HOA price premium varies by location. Southern core-based statistical areas (CBSAs) tend to have higher HOA premiums, while New York, CBSAs in the Rust Belt, and CBSAs on the West Coast exhibit lower premiums. Relating HOA premiums to local regulatory environments, historical and demographic features, and social attitudes sheds light on what drives homeowners' willingness to pay for HOAs. In particular, HOA price premiums correlate with a CBSA's degree of land use regulation and expenditures on public goods in a way that suggests that HOAs serve as a valuable substitute for local government, particularly with respect to zoning regulations and the provision of law enforcement and other highly localized public goods. Further, we find evidence of heterogeneity in HOA premiums by a CBSA's levels of economic inequality, attitudes toward racial integration, historical demographic composition, and several other characteristics.

Along with price premiums, the data we use for this paper reveal an unprecedentedly detailed picture of where and when HOAs were established, what houses in HOAs look like, and who lives in HOA-regulated communities. HOA membership is most common where land was developed recently. Accordingly, HOA houses tend to be newer and slightly bigger. HOA neighborhoods are also more racially homogeneous than nearby non-HOA neighborhoods.

The paper proceeds as follows. Section 2 explains in more detail what an HOA is, provides reasons why HOAs are an important institution to study, and reviews the previous literature on HOAs. Section 3 presents the local government real estate records we use and explains our methodology for identifying HOAs. Section 4 depicts the people, places, and structures found in HOAs. Section 5 uses hedonic regression to evaluate whether and how HOAs affect the value of single-family homes contained in them. Section 6 discusses the key implications

[4]The average HOA member on Redfin pays $237 per month in dues. The net present value of $237 per month at a four percent annual return is $71,100. See Section 6 for details.

of our results, and Section 7 concludes.

2 Background and Literature Review

2.1 Background on Homeowners Associations

Buyers of property in a common interest development purchase a *divided interest,* which belongs only to the individual buyer, plus a share in a *common interest,* which is owned corporately with the owners of nearby homes. Common interests in residential housing can range from pools and parks to streets and sewers—or the building itself in the case of condominiums. Owners in CIDs are also subject to a set of contractually enforceable rules, known as *covenants, conditions, and restrictions* (CC&Rs). By law, each CID must have a homeowners association to manage shared property and enforce the CC&Rs. The owner of every home in a CID automatically becomes a voting member of its HOA. This is analogous to buying shares of stock in a corporation; in fact, most HOAs are organized as non-profit corporations with a board elected from their membership. HOAs are empowered to collect dues and fines, which are enforceable in civil court (McKenzie, 1994; Esquivel and Alvayay, 2014).[5]

Common interest developments evolved in England during the 19th century and first made it to America with the establishment of Gramercy Park in Manhattan in 1831: houses were built around a private park and charged with the park's perpetual maintenance (McKenzie, 1994). Reflecting this first use, Foldvary (1994) argues that HOAs are able to provide "territorial public goods" more effectively and efficiently using "contractual government" than can be accomplished by municipalities via "coercive government." Helsley and Strange (1998) present a theoretical model that predicts that the existence of private governments (e.g., HOAs) reduces provision of goods by the public sector but increases aggregate wel-

[5]For simplicity, the rest of the paper will mostly ignore the distinction between CIDs and HOAs and refer to both the physical housing and governing organization as an "HOA."

fare. The intuition for this result is that private governments allow tailoring of government service levels to individuals' heterogeneous preferences. Total welfare of non-members, who demand less services, increases since their taxes fall, while the welfare effects for members is indeterminate since they pay for extra services through membership dues.

McKenzie (1994) portrays HOAs more negatively. He argues that HOA buyers are often unaware of extensive CC&Rs before buying or lack a non-HOA alternative in their local market for the type of housing they want. He also characterizes HOAs as an instrument of exclusion and a successor to racially restrictive covenants that the Supreme Court ruled unenforceable in 1948 and that the Fair Housing Act outlawed entirely in 1968. McKenzie (1994) additionally explores how HOAs' basis in contract law allows them to legislate details of residents' lives and limit speech in ways that exceed the police power of municipal governments.[6] The authority of HOAs to regulate personal activities could be viewed as particularly problematic given that HOAs are exempt from the one-person, one-vote rule that applies to general-purpose governments, and that only owners, and not renters, may vote in HOA elections.

Helsley and Strange (2000) formally model welfare effects when private social organizations (e.g., HOAs or private schools) allow individuals to "secede" from the larger society. The model predicts that individuals with high socioeconomic status will secede, leaving those who do not secede unambiguously worse off. Nelson (2009) postulates a rich political economy of HOAs wherein existing residents extract rents from new residents by forcing new housing to be built in HOAs. New residents pay for services they receive via HOA fees and contribute to public goods they do not receive via local taxes. Nelson (2009) also asserts that the popularity of HOAs has impeded the creation of new municipalities, since HOA residents in unincorporated areas want to avoid double taxation.

[6] Congress passed the Freedom to Display the American Flag Act of 2005 in response to HOA restrictions against the display of the American flag.

2.2 Empirical Studies on HOAs' Effects on House Prices

A number of empirical studies have focused on how HOAs affect house prices, the topic of this paper. It is an important question because home price capitalization of local public goods, like neighborhood or school quality, is a measure of how the market values that good. Hopkins (2016) provides a summary of empirical literature on home price capitalization of HOAs.

The most relevant prior study of HOA price effects is Meltzer and Cheung (2014). They obtain a list of home addresses for Florida HOA board members, then match those addresses to publicly recorded transaction and assessment data and subdivision boundaries. They treat subdivisions with board member residents as HOAs, under the twin assumptions that board members live in the HOAs they serve and that HOA boundaries match subdivision borders. Applying hedonic regression, Meltzer and Cheung (2014) estimate that HOAs raise house prices in Florida suburbs by nearly 5 percent. They also conclude that the HOA price premium falls as a subdivision ages, that larger HOAs offer a smaller premium, and that homes located nearby an HOA also exhibit a price premium, indicating positive spillover effects.[7]

The current paper improves upon Meltzer and Cheung's (2014) method of observing HOAs, even within the state of Florida, by utilizing HOA indicators for each house. House-level data mitigates the risk of entire subdivisions being mis-classified by one wrong address,

[7] Meltzer and Cheung also use these data for a number of other papers. Cheung and Meltzer (2013) show that a higher density of HOA housing in a city is correlated with fewer forms of land use regulation. Cheung and Meltzer (2014) show that HOAs formed earliest in census tracts that were predominantly white, higher income, further from the city center, and with high vacancy rates. HOAs formed more slowly in cities with high expenditures on public services. Cheung et al. (2014) show that Florida home prices react less negatively to rising mortgage delinquency rates in cities where HOAs are more prevalent, suggesting that HOAs mitigate neglect of foreclosed homes. Meltzer (2013) explores whether increased presence of HOAs over time is associated with cities becoming more segregated. She concludes that a 10 percent increase in HOA share leads to a 1–2 percent increase in local racial segregation, but has no impact on economic segregation. In a slightly older paper with different data, Cheung (2008) tests whether growth of HOAs leads to relative decreases in spending by California cities, as predicted by Helsley and Strange (2000). Using a list of HOAs derived from the Secretary of State of California business registry, he estimates the number of HOA homes that are located in 110 California cities over time and finds that a 10 percent increase in prevalence of HOA housing leads to a 1.5 percent fall in local government spending.

and also allows observation of HOAs that might not align with subdivision boundaries. Additionally, the data we use were provided free of charge by Zillow, instead of requiring purchase (see Section 3).

An earlier group of papers estimates the house price effects of *private covenants*, an overlapping but wider category than HOAs.[8] These studies have small sample sizes and rely on data provided by local multiple listing services. Speyrer (1989) performs hedonic regression with 230 single-family home sales in Houston, Texas, a city known for its lax zoning laws. She finds that prices of houses bound by private covenants are nearly 9 percent higher than those of houses with no zoning. Hughes and Turnbull (1996) run hedonic regression using 1,314 transactions in Baton Rouge, Louisiana and find positive price premiums that decrease over time for homes with private covenants. A 10 year-old covenanted home in their sample is expected to enjoy a 6 percent premium over comparable non-covenanted homes, while a 20 year-old covenanted home would sell at only a 2 percent premium. Rogers (2006) uses 1,487 single-family home sales in Greeley, Colorado to consider the effect of different forms of covenants that can be implemented by HOAs. He finds price premiums of 2–3 percent for membership in an HOA and attributes the gains to use restrictions as opposed to building restrictions. Cannaday (1994) compares pet restrictions in 1,061 high-rise condominium units in Chicago to conclude that units in buildings that ban dogs but allow cats sell for a 16-17 percent premium over units in buildings that allow large dogs. Groves (2008) is the first study to focus explicitly on whether HOAs (or "Residential Community Associations") raise house prices. He is also the first to introduce a larger data set, based on state government sale records and an extensive search of local deed restrictions: 124,878 sales of single-family homes in St. Louis County from 1992 to 2001. Groves' regression specifications that are most comparable to other studies find a 1–3 percent price premium for living in an HOA.[9]

[8]Houses can be subject to private covenants but not be members of an HOA as long as there is no common interest, i.e., property owned jointly with other home owners (Esquivel and Alvayay, 2014).

[9]Groves's (2008) more complicated, preferred regression specification accounts for "spatial lag" of home price by including the average sale price of previously sold nearby homes as an independent variable. This

3 Data

Information on the prevalence, much less the activities, of HOAs is difficult to collect. As McCabe (2011) notes, "Data concerning the number and location of HOAs are inexact because most of the information about them resides in local property records, where their founding is recorded. They are not included as discrete entities in the U.S. Census, tracked by state governments, or comprehensively mapped by local governments." That is, public records of the details of HOAs exist, but are "buried" in local property records. It is easy to find founding documents for an individual HOA, but prohibitively time-consuming to compile them for the whole country. However, an industry has grown up around collecting and harmonizing local property records for business purposes. Zillow, a company that compiles property data for use in their online real estate viewing platform, has recently offered to share data with qualified researchers, calling it the Zillow Transaction and Assessment Dataset (ZTRAX).[10] Most founding documents to which McCabe refers, such as parcel maps and text of deed restrictions, are not captured in ZTRAX. However, one key indicator is included in ZTRAX: mortgage riders.

Mortgage lenders require that an addendum (or "rider") be added to publicly recorded loan documents for properties that are included in a CID, since that membership abridges the lender's ownership rights in case of foreclosure.[11] Thus, ZTRAX includes two relevant flags for mortgage transactions, indicating whether a mortgage has either a *condominium rider* or a *planned unit development rider* attached to it. The presence of either of these riders

addition alters the proper interpretation of his regression, since identification of an HOA price premium now relies only on houses at the edge of an HOA, which have non-HOA neighbors. With this specification, Groves's (2008) estimated HOA price premium falls to a precise zero, indicating that the value of living in an HOA may be linked to a house's seclusion from other neighborhoods.

[10] More information on accessing ZTRAX can be found at http://www.zillow.com/ztrax. The results and opinions are those of the authors and do not reflect the position of Zillow Group.

[11] Riders also provide HOAs the right to alert the lender when an owner has fallen behind on their HOA due payments, and to give the lender the opportunity to pay the assessment (which the lender can then add to the cost of the loan). This is important because HOAs can in principle place a lien on one's house. Fannie Mae provides a sample planned unit development rider here: https://www.fanniemae.com/content/legal_form/3150.pdf.

indicates that the home is bound by an HOA, as explained more fully later in this section.[12] (For simplicity, this paper references a notional "HOA rider" to indicate when either of these mortgage riders are present.) Along with mortgage records, ZTRAX also includes past sale prices, details of each property's physical features, and each parcel's geographic coordinates.[13]

3.1 Geographic Coverage

ZTRAX is an amalgamation of local databases, mostly collected at the county level. Presentation of the data is harmonized across jurisdictions, but there is substantive local variation in the data that are reported. At the most basic level, ZTRAX includes at least one hundred residential real estate transactions from 2,619 of the 3,142 U.S. counties. Of those, 2,021 counties have at least one mortgage with an HOA rider. It can be difficult to know whether the counties that lack any reported mortgage riders simply have no HOAs or if they are not reporting mortgage riders. (In the case of six New England states, no mortgage riders are reported for the whole state.) When estimating the effect of HOA membership on house values, we exclude larger counties in which we never observe a transaction with an HOA rider in order to ensure comparisons between similar houses. Specifically, when reporting descriptive statistics, we use a simple, admittedly arbitrary decision rule: any county with at least 5,000 new houses built since 1960 that does not report any HOAs is treated as a non-reporter and left out of the calculation. Counties with fewer than 5,000 houses built since 1960 are left in the calculation with the understanding that, even if they really do have HOAs, the counties' moderate size means omitting their HOAs will cause only modest down-

[12] The distinction between these two most common types of CID is a legal construct, rather than a difference of building type, so this paper combines them to form a single HOA indicator. Condominiums are typically built as multi-family developments, but can be composed of single-family homes (Esquivel and Alvayay, 2014). It seems more informative to distinguish between physical form of housing developments rather than legal forms, as described below.

[13] ZTRAX geocodes are "enhanced Tiger coordinates." This means each block face is located using Census Bureau data, and locations of individual rooftops are interpolated based on their house number. Viewed on a map, many ZTRAX geocodes are slightly shifted from their true location but correctly capture the distance between neighboring properties.

ward bias in the estimated share of houses with an HOA. By these criteria, ZTRAX contains information about the HOA status of homes in 2,526 counties, containing 90 percent of the U.S. population. Figure 1 provides a visualization of which years' real estate transactions are included in ZTRAX by state.

3.2 Inferring HOA Status from Mortgage Riders

As stated above, mortgage origination rules dictate that loan documents for houses bound by an HOA should include either a planned unit development rider or a condominium rider, depending on the legal type of HOA. Thus, an ideal version of ZTRAX would yield a simple binary indicator of HOA inclusion for each house that has a recorded mortgage. Unfortunately, however, mortgage riders are inconsistently reported within localities. Wrong HOA rider flags can arise at several points: mortgage originators might mistakenly omit (or include) an HOA rider in the actual document, local governments might record the loan documents with an error, or Zillow might err when adding public records to the ZTRAX database. In any case, inconsistency between multiple mortgages on the same property and visual inspection of ZTRAX data displayed on a map suggests non-trivial levels of both Type 1 and Type 2 errors. That is, some mortgage records of homes that truly do belong to an HOA lack the appropriate HOA flag, and mortgages for some homes that truly do not belong to an HOA have an HOA flag.

If a large number of mortgages were observed for each house, occasional wrong mortgage rider flags would not be a big problem. The modal HOA status of mortgages tied to each house would reliably indicate whether the house was in an HOA, thanks to the law of large numbers. Unfortunately, most houses have a record of being mortgaged a few times at most.

Thus, this paper uses two distinct approaches for determining whether a house is in an HOA. The first approach is simple: houses are designated as "HOA" if any mortgage on that house has ever included an HOA rider. Houses with no recorded mortgage are disregarded.

The second approach works by finding groups of houses that probably all share the same HOA status, and deciding their HOA status together, based on the larger number of mortgages. When should neighboring properties share the same HOA status? The legal process of establishing an HOA requires that all involved land owners agree to restrict their own property rights. Such complete agreement is usually only achieved when all the land is owned by one entity—a real estate developer. Developers subdivide larger plots of land, establish an HOA, and build houses that future HOA members will occupy. The houses are built in physically connected clusters, either all at once or in phases (Esquivel and Alvayay, 2014). Thus, if a house is included in an HOA, it is likely that a neighboring house built in the same subdivision or in the same year is also in the HOA. A large enough group of neighboring houses built at the same time will be tied to enough mortgage records to reliably conclude whether or not the whole cluster is bound by an HOA.

With the above real estate development process in mind, two options arise for associating nearby houses that likely share the same HOA status. First, houses can be grouped by subdivision whenever that information is included in ZTRAX. Second, even when a house is not in a recorded subdivision, clusters of houses that were built near each other at the same time can be observed using an algorithm. Once houses are grouped—by subdivision or by cluster—the whole group's HOA status can be judged based on how many of the houses have an HOA rider. The following decision rule is used: if less than 20 percent of ever-mortgaged houses in a subdivision or cluster have been flagged with an HOA rider, none of the houses in the group is considered part of an HOA. If over 60 percent, all houses in the group are treated as belonging to an HOA. Between 20 and 60 percent, no determination is made.

The details of the clustering algorithm used, analysis of how well subdivisions and clusters work in finding groups of houses with homogeneous HOA status, and justification for the cutoffs to decide HOA membership are presented in Appendix A. Both methods — grouping by subdivision or by cluster — work similarly well and each produce a classification for nearly half the houses in the full sample. By using both grouping methods, and accepting the result of either classification scheme as long as they do not contradict each other, we can assign an HOA status based on neighboring houses for 33 million houses, or 61 percent of the full sample. We refer to this amalgamation of the subdivision-based and cluster-based approaches as the neighborhood-based method.[14]

To summarize, a house's HOA status will be judged in this paper based on (1) whether any mortgage on the house has ever included an HOA rider (the "house-based" method) or (2) whether most nearby houses' mortgages include HOA riders, in cases where we think the nearby homes all share the same HOA status (the "neighborhood-based" method).[15]

3.3 Other Data Elements in ZTRAX

In addition to houses' HOA status and sale prices, ZTRAX provides a wealth of information on the characteristics of land parcels, the structures on the land, and transactions involving the parcels. Table 1 lists and briefly explains the data elements from ZTRAX used in this paper. Table 2 presents descriptive statistics involving those data elements.

3.4 Focus on Single-Family Houses

HOAs govern both single- and multi-family housing, but we look exclusively at single-family housing. Property in a common interest development consists of a *divided interest* bundled together with a *common interest*, which must be governed by an HOA. A neighborhood of single-family homes without any common interests does not need an HOA. Thus, one can imagine comparing sale prices of identical single-family homes, where some of the homes include a common interest *cum* HOA. Systematic differences between sale prices of HOA and non-HOA homes would be attributable to the value of the common interest and HOA bundle.

[14]For the 14 million houses where both methods assign an HOA status, the two methods agree over 99 percent of the time.

[15]The classification of homes using the house-based and neighborhood-based methods differ, but not substantially so. If we assume that the neighborhood-based method is correct, the house-based assignment method has a Type 1 error rate (false positives) of 5.5 percent and a Type 2 error rate (false negatives) of 4.5 percent.

With multi-family homes, comparing sales of otherwise identical units with and without an HOA is not possible. A multi-family building must have an HOA if the units are to be owned by multiple parties, since the walls are shared. The building may not have an HOA if all units have one owner, but, in that case, any sale is of the whole building. That cannot be compared to the sale price for a single unit. Additionally, sales of multi-family HOA homes cannot be compared to sales of single-family non-HOA homes, since there is no way to separate the value of HOA membership from the value of being in a multi-family building. Thus, multi-family HOA units have no suitable comparison group.

Apart from this conceptual problem, ZTRAX has data limitations that make the study of HOAs in multi-family housing difficult. For single-family homes and multi-family condominium units, each property tax record corresponds to one dwelling. However, for multi-family apartment buildings, one tax record can correspond to many units. The same is true for cooperatives, a less common form of multi-family HOA popular in New York City. Even though ZTRAX includes a field for the number of units in a building, that field is frequently not populated.

3.5 Excluding Home Sales with No Mortgage

ZTRAX draws from public data sources that record *all* real estate transactions. Many transactions occur between related parties and do not reflect a market price. Non-market transactions decrease the precision of regression results and can also bias results if non-HOA homes are more frequently transacted above or below the market price than HOA homes.

The logical step is to eliminate apparent non-market transactions. We do this partially by dropping transactions with an *intra-family transfer flag* and sales priced below $1,000. However, exploratory regression results still indicate a likely problem: on average, prices are 30 percent higher for homes that transact with a mortgage relative to similar homes that do not transact with a mortgage. Such a large difference raises a concern that many homes sold without a mortgage are not sold at market prices. For this reason, estimates of the value of

HOA membership presented later in this paper only use sale records that are accompanied by a mortgage, representing around 60 percent of sale records. While observable characteristics of mortgaged HOA and non-HOA homes are comparable to those of the broader sample of single-family HOA and non-HOA parcels (see Table 2), to the extent that mortgaged homes are not representative of all homes, the results will not necessarily generalize to the full universe of houses in the U.S.

3.6 Redfin Data

A secondary, also novel source of data for this paper is HOA fees described in online house listings. Standard real estate listing forms include a field for monthly HOA fee. Redfin.com is a real estate listing site that allows visitors to download a text file with details from up to 350 house listings in a given ZIP code at a time. Those details include the monthly HOA fee, if any. Using a computer program, we gathered a list of 900,000 house listings that appeared on Redfin on June 5, 2017.[16] Notably, these data represent a single cross-section of listings; we cannot match them to the historical transactions in the ZTRAX data. Also, the data include a limited number of housing characteristics and only cover a selected set of zip codes nationwide. Figure 2 illustrates the geographic coverage and cross-sectional variation in HOA fees based on the Redfin data. The mean (median) price of a single-family home listed in our Redfin sample is $483,277 ($279,000). One third of single-family homes in the Redfin sample are listed as being in an HOA with positive monthly dues. The average monthly HOA dues for a single-family home in an HOA in our Redfin sample is $237, which is consistent with Realtor.com's estimate that a typical single-family home in an HOA faces a monthly fee of $200-$300 (Mandell, 2016).[17]

[16] We adapted the code from the GitHub repository of Lopez (2017).

[17] In calculating average monthly dues in the Redfin sample, we only include reported dues less than $9,999/month due to apparent misreporting in a few cases. In the cross-section of data we have from Redfin, listing prices for single-family homes in HOAs are 10 percent higher than listing prices for single-family homes not in HOAs. Among single-family homes in HOAs, listing prices also covary positively with HOA dues.

4 Descriptive Statistics

The data we use in this paper provide unprecedented insight into the characteristics and prevalence of HOAs. Before estimating how membership in an HOA influences house prices, we explore where HOAs are, who lives in them, and what the buildings look like. This section focuses on our simplest measure of HOA membership—whether any mortgage on each house has ever included an HOA rider—in order to capture a representative sample of homes. The neighborhood-based method of eliciting HOA membership restricts the focus to subdivisions and clusters of homes built together.

4.1 HOA Growth and Coverage

Figure 3 indicates the share of single-family homes built in each year that are governed by an HOA. There has been a steady increase in the prevalence of HOAs since the 1960s. Almost 60 percent of new single-family home construction today is in an HOA. An even larger share of houses on newly developed land are in HOAs; the top line in Figure 3 shows single-family homes built in new subdivisions and utilizes the neighborhood-based measure of HOA membership. We see that around 80 percent of single-family homes built in new subdivisions today are bound by an HOA.[18]

The top panel of Figure 4 shows the share of houses in an HOA by census division. We see that HOAs are most common in the Mountain States, the states around Texas, and the southern Atlantic coast. The bottom panel of Figure 4 shows the share of houses that are in an HOA by county, demonstrating that HOAs are mostly found in and especially around cities. This fact reflects patterns of new residential land development since 1960. Figure 5 mirrors Figure 4, but restricts to houses built in new subdivisions since 2000 to show the ubiquity of HOAs in recent new land development.

[18]This group includes homes whose age is within 5 years of the mode in their subdivision, constituting more than half of single-family homes constructed in recent years.

4.2 Building Characteristics of HOA Housing

Table 2 shows average characteristics of single-family homes with and without an HOA. Homes in an HOA are much newer on average, and while they are situated on smaller parcels, have around 20 percent more floor space than non-HOA homes on average. Additionally, homes in HOAs have higher property taxes on average, are more likely to have access to a golf course, and are more likely to be waterfront. The raw observed differences along each of these dimensions highlights the importance of conditioning the comparison of HOA and non-HOA home prices on a rich set of housing and neighborhood characteristics.

4.3 HOA Residents

This subsection links ZTRAX to 2010 Decennial Census data to learn about the people living in HOAs. The smallest area for which the Census produces detailed resident economic statistics is a block group, but racial composition is available at the smaller block level. The U.S. is divided into over 200,000 block groups, generally housing from 600 to 3,000 people in 2010. These block groups are further subdivided into blocks, the smallest geographic unit used by the Census to tabulate data. There are 11.2 million blocks nationwide, each housing 28 people on average in 2010.

The top panel of Table 3 presents estimates of the median annual household income and race of HOA residents in single-family homes by block group. This table is calculated by weighting each census block group by its population multiplied by the share HOA. That is, a block group with 1,000 residents with 60 percent HOA homes will contribute a weight of 600 residents in the HOA column and 400 residents in the non-HOA column. This method of weighting reveals that HOA households earn over one third more than non-HOA residents. Comparing the ethnicity of HOA- versus non-HOA residents, HOAs contain slightly more residents who are white and Asian, and fewer that are black or of another (or mixed) race.[19]

[19] The races in the other category include American Indian/Alaska native, native Hawaiian or other Pacific islander, some other race, and two or more races.

For these statistics to reflect the characteristics of HOA residents accurately, one must assume that residents of HOAs are representative of the block group in which they live. This may not be true, especially if HOAs are formed within areas to isolate small groups of homeowners. In light of this possibility, and given that variation in HOA status is substantially less at the block level than at the block group level, we go on to exploit block-level information from the 2010 Decennial Census, which allows us to paint a more accurate picture of the population living in HOAs along racial lines (albeit not income lines). As shown in the bottom panel of Table 3, we find results similar to those at the block group level, with white and Asian individuals overrepresented in HOAs.

One concern about HOAs is that they might be used as a tool for exclusion and thereby foster residential segregation by race. To shed light on whether HOAs are associated with heightened segregation, we examine the degree to which racial diversity differs across neighborhoods with and without HOAs. Specifically, for each block in which we observe a home in the ZTRAX database, we calculate the Gini-Simpson Index, which measures the probability that two randomly drawn residents from that block are of different races (where the different races we consider are white, black, Asian, and other).[20] We define a block as an "HOA block" if any home on that block transacted with an HOA rider. On average, HOA blocks have Gini-Simpson Index values higher than non-HOA blocks; in particular, the probability that two randomly chosen residents on an HOA block are of different races is 26 percent, whereas the probability that two randomly chosen residents on a non-HOA block are of different races is 20 percent. As shown in Table 4, the 0.06 difference is statistically significant even if we allow errors to be correlated across blocks within the same county.[21] However,

[20]Specifically, we calculate for each block j in the sample

$$D_j = 1 - \frac{\sum_i n_{ij}(n_{ij} - 1)}{N_j(N_j - 1)},$$

where n_{ij} is the number of individuals belonging to the ith race (i.e., white, black, Asian, or other) in block j, and N_j is the total number of individuals residing in block j. Relative to the standard Gini-Simpson Index, which assumes sampling with replacement, this formulation assumes sampling without replacement, which is arguably more appropriate for this setting in which the population in many blocks is small.

[21]We obtain very similar results if instead of using an indicator for whether any home on the block

this difference is partly attributable to the fact that block population is positively correlated with HOA status as well as diversity; controlling for (log) block population narrows the differential to a marginally statistically significant 0.01. Restricting comparisons to blocks within states (in column (3)) reverses the sign on HOA status, suggesting that on average, HOA blocks tend to be less diverse than similarly populated non-HOA blocks within the same state. This difference is further magnified when we compare similarly populated blocks within the same county (column (4)); the probability that two randomly selected residents from an HOA block are of different races is two percentage points lower in HOA blocks as compared to similarly populated non-HOA blocks in the same county. This result is consistent with several previous papers that suggest that HOAs facilitate segregation within local areas (e.g., Gordon (2003), Meltzer (2013)), and also foreshadows subsequent results relating CBSA- and state-level HOA premiums with social attitudes toward race.

5 Hedonic Valuation of HOA Membership

This section will test if and how HOA membership is capitalized into home values by examining how the average selling price of homes in HOAs compares to the average selling price of observationally equivalent homes that are not in HOAs. A central econometric challenge in studying how HOA membership affects house price is that HOA membership is always sold in a bundle with a private home. There is no market price of HOA membership, which is the same dilemma faced when studying the value of other features of a home. We approach the problem as other authors have by using hedonic regression in the tradition of Rosen (1974) to estimate the contribution of HOA membership to house prices. We begin with an easily interpretable log-linear OLS model that incorporates a rich set of controls and identifies off variation in HOA status among observationally similar homes sold within the same month and within the same small geographic area. As robustness, we also consider a more flexible semi-parametric regression model, intended to better fit the potentially non-linear house

transacted with an HOA rider, we use the share of homes transacted with an HOA rider.

price function. Additionally, we present estimates using both measures of HOA membership described in Section 3, and broken out by state and CBSA.[22]

The largest regression sample used includes 10.1 million transactions. As discussed in Section 3, this group includes sales of single-family homes built since 1960, sold since 2005, and recorded along with a mortgage loan. When restricted to the neighborhood-based measure of HOA membership, the national regression includes 8 million transactions.

5.1 Hedonic Regression Models

The first regression model presented is log-linear OLS. Log sale price is regressed on key property and transaction characteristics described in Table 1 plus fixed effects for the geographic area (census block group) and month of sale. That is,

$$\ln P_{ijt} = \beta_0 + \beta_1 HOA_i + \mathbf{X}_{it}\mathbf{\Omega} + g_j + \eta_t + \epsilon_{ijt} \tag{1}$$

where $\ln P_{ijt}$ is the log sale price of home i, located in geographic area j, and sold in month t; HOA_i is an indicator for whether home i is part of a common interest development; X_{it} is a vector of house- and transaction-specific characteristics; g_j is a vector of geographic area (block group) fixed effects; η_t is a vector of month-by-year fixed effects; and ϵ_{ijt} is the error term.

In equation (1), the main coefficient of interest, β_1, is identified off variation in HOA status across observationally equivalent homes sold at a similar point in time within a small geographic area. Our use of block group fixed effects is motivated by a desire to restrict comparisons to nearby homes that not only share similar observable and unobservable property characteristics, but also are subject to the same local government taxes and regulations and have access to the same set of (non-HOA) public goods and services. One concern, however,

[22] CBSAs are geographical units that constitute one or more counties anchored by an urban core of at least 10,000 people and that are linked socioeconomically by commuting. They include both metropolitan statistical areas and micropolitan areas; the latter are areas with urban clusters with at least 10,000 people but fewer than 50,000 people.

is that in this case we are only identifying the effects of HOA status off a very small and potentially unrepresentative sample of "mixed" block groups; i.e., those that contain both HOA and non-HOA homes. However, 81 percent of CBSAs, 70 percent of counties, and 48 percent of tracts in the sample have a positive number of block groups containing both HOA and non-HOA houses. Also, estimates of β_1 are qualitatively similar (although larger in magnitude) if we instead use fixed effects for larger geographic areas (e.g., tracts).

Another potential concern is that locations with more HOA homes could exhibit different trends in house prices. Therefore, in some specifications, we replace the block group and month fixed effects with block group-by-month fixed effects. In that case, we restrict identification to deviations in HOA home prices from the neighborhood-level trend for similar non-HOA homes. One additionally might worry that some unobserved characteristics of houses might be correlated with HOA status and also independently affect prices.[23] While the rich set of controls and fixed effects we include in our baseline regressions help to mitigate this concern, along with our main results, we consider additional tests aimed at assessing the importance of unobserved housing characteristics in biasing our estimates in Section 5.2.1.

Notably, the log-linear model implicitly assumes that house prices are determined as a log-linear function of the included variables. Hedonic price theory gives no reason to think the function is log-linear. Thus, we also consider the following generalized additive model ("GAM") for the hedonic price function. GAMs are a relatively popular form of semi-parametric estimator in which at least some of the individual terms are estimated non-parametrically, but they are combined additively with other terms (Hastie and Tibshirani, 1990; Wood, 2017). The GAM model we use is:

$$\ln P_{ijt} = \beta_0 + \beta_1 HOA_i + \sum_p f_p(x_p) + e_{ijt} \tag{2}$$

[23]Nowak and Smith's (2017) finding that, when used in conjunction with standard housing attributes, information from the text descriptions of homes in MLS listings can decrease pricing error supports the notion that unmeasured housing characteristics could be important.

where f_p is a non-linear data driven function of x_p, known as a *smooth*.[24] A limitation of GAM models is that only continuous variables can be smoothed. Discrete variables, such as the number of bedrooms, can either enter linearly or be combined into a continuous index that can then smoothed. The model presented here does the latter by first predicting log sale price as a linear function of the discrete variables to create an index, then modeling log sale price as a function of HOA status and a vector of continuous variables.[25] The vector of continuous variables x_p includes all continuous variables used in the OLS model, the index of discrete variables, and an interaction of latitude and longitude which helps the model account for spatial autocorrelation (Wood, 2017).

5.2 Results

5.2.1 National Results

Table 5 presents national regression coefficients for six model specifications. The first column shows the results of running equation (1) using the house-based measure of HOA status and including census block group fixed effects and month fixed effects (in addition to the full complement of individual house characteristics). The second column replaces the block group and month fixed effects with block group-by-month fixed effects. The third and fourth columns show results from the same regressions as the first and second, only using the neighborhood-based measure of HOA status instead of the house-based measure. The last two columns of the table show results using the GAM model, first for the house-based measure of HOA status and then for the neighborhood-based measure of HOA status.

In the sparsest model we consider and using the house-based HOA measure (in column (1) of Table 5), we find that HOA status is associated with $e^{0.058} - 1 = 6.0\%$ higher house prices on average. The estimated premium falls to 5.3 percent with the inclusion of block

[24] *Smooths* work the same as non-parametric functions applied to histograms to provide a smooth, local estimate in one variable.

[25] All continuous variables are de-meaned within census tract, which is equivalent to including a vector of census tract dummies but is more computationally tractable.

group-by-month fixed effects (column (2)). The results are little changed when we instead use the neighborhood-based HOA measure (columns (3)-(4)).[26] These effect sizes are well within the range of those found in earlier studies using linear models to estimate the effects of HOAs on house prices. Also notably, the estimated relationships between house prices and other housing characteristics, such as lot size, age, exterior material, etc. generally align with economic intuition and earlier hedonic housing studies.[27]

The final two columns of Table 5 present results using the GAM model with discrete variables combined into a continuous index. We only show estimated coefficients for HOA status since a GAM model fits a curve to continuous variables rather than producing a coefficient. The results using both house-based and neighborhood-based HOA measures are of very similar magnitude, albeit slightly smaller, than those from the linear models. The estimate in column (5) indicates that houses that have been flagged with an HOA mortgage rider sell for 4.5 percent more on average. Similarly, the estimate in column (6) suggests that houses in neighborhoods where most homes have been flagged with an HOA mortgage rider sell for 3.7 percent more on average.

Despite our rich set of controls, an important concern is that unmeasured characteristics of houses that are correlated with HOA status could independently affect prices. In that case, our estimated HOA premium could over- or under-state the value households place on living in an HOA. To evaluate the extent to which unobserved housing characteristics might explain our HOA premium, we follow Oster (2017) in determining the share of the variation in house prices that unobserved characteristics need to explain, relative to included

[26]Using the house-based measure of HOA membership on the smaller sample of homes covered by the neighborhood-based measure yields very similar results to when we use the full sample.

[27]If we replace block group fixed effects with tract fixed effects, thereby expanding the set of HOA homes contributing to identification but also increasing the scope for unobserved property and neighborhod characteristics to bias our results, we find similar but larger estimates of the HOA premium. Specifically, depending on whether we use tract-by-month fixed effects and the house- vs. neighborhood-based measure of HOA status, we find estimates of the HOA premium that range between 7.5 percent and 8.2 percent. While the larger estimates may reflect omitted variable bias introduced by comparing less similar houses, they could also reflect the presence of positive spillovers of HOAs on nearby non-HOA houses (as in Meltzer and Cheung (2014)). Such spillovers would have the effect of biasing our estimates toward zero when using fixed effects for finer geographic levels.

variables, to reduce the estimated HOA premium to zero. In conducting her formal test, we must assume a maximum R-squared from a hypothetical regression that includes unobserved as well as observed controls; Oster (2017) suggests based on experimental results specifying an R-squared that is 30 percent larger than the R-squared from the regression with observed controls. Under that assumption, based on the specification in column (1) of Table 5 (including block group and month fixed effects), we find that unobservable characteristics of housing units would need to be 1.52 times as important as observable characteristics to reduce the HOA premium to zero.[28] If we take a stronger stand on which of the many controls are correlated with unobserved characteristics, and in particular assume that unobserved characteristics are only correlated with observed physical features of the house itself (building area, lot size, number of bedrooms, number of bathrooms, flooring type, exterior wall material, garage area, other building area, fireplace, presence of deck, presence of fence, presence of pool, and roofing material), we find that unobserved characteristics of houses would not only have to be negatively correlated with observed characteristics, but would also have to be over eight times more important than observed characteristics, to produce a zero effect. Taken together, results from tests in the spirit of Oster (2017) suggest that we can reject the hypothesis that the HOA premium is driven entirely, or even largely, by unobservable house characteristics.

We conduct a number of additional tests to determine whether HOA membership's effects on home values are mediated by specific property and neighborhood characteristics. The results of these tests, which all take as a starting point the specification that appears in column (1) of Table 5, appear in Table 6. In the first column, we show results from adding an interaction of the house-based HOA dummy with the age of the home at the time of sale; while not a perfect measure of the age of the HOA, given that most HOA homes today have been HOA homes since they were built, this represents a reasonable proxy. We find with

[28] For this test, we treat block group and month of sale as nuisance parameters and, as such, unrelated to the set of proportionally related variables. To the extent that features of a community, as opposed to attributes of the house itself, drive

the HOA premium, we want to capture those features in our estimate of the premium.

this specification that on average, a newly constructed HOA home sells for about 8.3 percent more than an observationally equivalent newly constructed non-HOA home. For each year of age, the HOA premium declines by 0.17 percent. Thus, a 10-year old HOA home sells for a premium of only about 6.6 percent on average, a 20-year old home sells at a premium of only 4.9 percent on average, and homes over about 40 years old sell at close to zero premium on average. This fading of the HOA premium over time, which is consistent with Hughes and Turnbull (1996) and Meltzer and Cheung (2014), could reflect differences in the structure and functions of older vs. newer HOAs. It could also reflect changes over time in the effectiveness of a given HOA in providing public goods at a reasonable cost, changes in the competitiveness of nearby local governments that might be offering substitutable amenities, or changes in the composition of HOA residents as neighborhoods age that might influence the scope of HOA activities.

We also consider how the HOA premium varies with building area, lot size, the implied property tax rate (i.e., the property tax bill divided by the assessed value in the most recent year), and subdivision size. We convert log building area, log lot area, and the tax rate into z-scores for ease of interpretation. Our measure of subdivision size is based on the number of transactions observed among units with the same legal subdivision name in ZTRAX that are also in the same county; since this is not an exact measure of the number of units in a subdivision, we simply use a dummy for whether we observe at least 50 units in the subdivision.[29] We also drop houses that have no legal subdivision name recorded in the data from the sample for this test.

As shown in columns (2) and (3) of Table 6, we find that larger homes, whether measured in terms of building area or lot size, tend to have a higher premium associated with HOA membership. This could reflect the greater value owners of more expensive homes place on ensuring nearby houses are congruent and that the neighborhood as a whole is well-maintained. Although the coefficient on the interaction is not statistically distinguishable

[29]The results of this exercise are qualitatively similar when we use different thresholds as well as when we use a continuous measure of subdivision size.

from zero, we also find that homes with higher tax rates tend to have lower HOA premiums.[30] This hints that HOAs may be viewed as substitutes for local governments, an issue we explore further in subsequent sections. Finally, consistent with Meltzer and Cheung (2014), we find that homes in larger HOAs tend to have lower premiums than homes in smaller HOAs. This suggests that the value households attach to living in HOAs, including whatever value they provide in terms of highly tailored amenities and perhaps a more homogeneous group of neighbors, is diluted when the subdivision is larger.

Having established a near-national estimate of how HOA membership affects house value, as well as how people's willingness to pay for HOAs varies with property and neighborhood characteristics, it is also interesting to explore variation among U.S. states and CBSAs within a unified framework. Doing so is not only intrinsically interesting, but it also opens the door for direct comparison to related work that is based on data specific to particular markets. Further, it allows us to consider the extent to which the HOA premium interacts with other local factors.

5.2.2 State- and CBSA-Level Results

We break out the analysis by states and CBSAs to examine how the HOA premium varies across geography. For these sub-sample analyses, we show results from the linear specification using the house-based measure of HOA status that includes block group and month fixed effects (as in column (1) of Table 5). The rank ordering of geographies based on the estimated HOA premium, as well as how that premium varies with other local characteristics (as discussed in the next subsection), is fairly robust to the particular specification and HOA measure.

We illustrate our state-level estimated HOA premium in Figure 6. The scatterplot shows the fraction of single-family houses built after 1960 transacted since 2005 that are in HOAs on the horizontal axis and the estimated HOA premium on the vertical axis. The markers

[30] Contributing to the small and insignificant effect of the tax rate is the fact that tax rates do not vary substantially within block groups among HOA and non-HOA homes.

are coded according to census region, with red circles for the Northeast, blue diamonds for the Midwest, green squares for the South, and yellow triangles for the West. In general, we see that states with a greater share of transacted homes in HOAs tend to be those with higher HOA premiums. The highest HOA premiums tend to be in states in the South. Price differences between homes that are in vs. not in HOAs, but that are otherwise similar, tend to be smallest in states in the Northeast and West (with the exception of Hawaii, where the HOA premium is in excess of 18 percent).[31]

We next estimate HOA premiums for each CBSA in our sample.[32] Some of the smallest HOA premiums are found in Detroit and New York City, with estimated premiums of -3 percent and -1 percent, respectively. CBSAs with relatively large HOA coefficients include Greenville, South Carolina (18 percent); Oklahoma City, Oklahoma (16 percent); Buffalo, New York (12 percent); Tulsa, Oklahoma (11 percent), and Charlotte, North Carolina (10 percent).[33]

We can compare our state- and city-specific results with results from previous studies that have focused on smaller geographies due to data limitations. For example, we estimate a 5.9 percent HOA price premium for the state of Florida, which is statistically different and slightly higher than Meltzer and Cheung's (2014) estimate of 4.9 percent; a 3.9 percent price premium in the Greeley, Colorado CBSA, which is higher than Rogers's (2010) estimated premium; and a 7.0 percent price premium in the St. Louis CBSA, which is higher than Groves's (2008) estimated premium for St. Louis County.

[31] In results not shown, we also break the results out by census division. There are nine census divisions, each comprised of a few states. We find the smallest price difference between HOA and non-HOA single-family homes in the mid-Atlantic (including New York) and along the West Coast. The highest HOA price premiums are in the southern U.S., peaking in the East South Central division.

[32] In Appendix B, we provide coefficient and standard error estimates for HOA membership from OLS and GAM regressions for the largest fifty CBSAs in the sample by population. Coefficient estimates from each model are very similar; for the 50 CBSAs shown in Appendix B, the correlation between estimates of the HOA premiums from the OLS model with block group and month fixed effects and those from the GAM model is 0.86.

[33] Houston, Texas has one of the highest estimated premiums based on the OLS model with block group and month fixed effects (over 20 percent), but it is one case in which the estimate is not robust to alternative model specifications. The differences in estimated premiums across models suggest that in Houston, HOAs tend to be located in neighborhoods with different trajectories on average, and in particular places where house price growth is relatively strong.

5.2.3 Mechanisms

In order to shed additional light on the potential sources of the HOA premium, we relate our estimated premiums to a variety of local factors hypothesized to influence the perceived value of living in a community governed by an HOA. Table 7 shows the results of a series of simple linear regressions relating CBSA-level or state-level HOA price premiums to CBSA-level or state-level conditions, one at a time. The coefficients in Table 7 are equivalent to the slope of the line of best fit for a scatter plot with HOA premiums on the vertical axis and the CBSA/state-level characteristic on the horizontal axis.

One metric in Table 7 with a particularly strong relationship to the CBSA-level HOA premium is the Wharton Residential Land Use Regulation Index, which measures the stringency of local housing market regulations. The results indicate that HOAs are less highly valued where land use regulations are more stringent. This finding is consistent with the idea that HOAs provide "private zoning," which is most valuable in the absence of public zoning. This is in line with Speyrer (1989), who makes this argument in explaining the relatively high HOA premium she observes in Houston, Texas, a city that lacks strict zoning regulations.

HOAs may similarly serve as a substitute for local governments in the provision of other public goods, such as roads, law enforcement, and parks. To test this proposition, we compiled data from the 2012 Census of Governments on all local government revenues and expenditures, and aggregated to the CBSA level. We scale each of the revenue and expenditure measures by the number of housing units in the CBSA in 2010. Consistent again with the perception of HOAs as a substitute for local government, we find that HOA premiums are greater in CBSAs where local government revenues, and in particular property tax revenues (which are used to finance many of the same types of public goods for which HOA dues pay), are lower. Meanwhile, HOA premiums tend to be higher in CBSAs with lower public spending, including public spending on roads, police, parks and recreation, and administration (although only the effects for police and administration are statistically significant).[34]

Helsley and Strange (1998) model HOAs as a means of providing differentiated levels of public services within the same municipality. Thus, demand for HOAs could be higher when residents' demands for public services are more varied. Contexts with greater income inequality might feature a larger customer base willing to pay for higher service levels. In partial support of this idea, the coefficient for Gini index in Table 7 is positive and statistically significant.

HOAs have also been cast by some as a tool for exclusion of some groups. We might therefore expect that the HOA premium would be higher in places where people hold more discriminatory attitudes. To examine this, we first consider the relationship between the CBSA-level HOA premium and results of the "Black-White Implicit Association Test." This is an online test administered by a Harvard University-affiliated group called Project Implicit. Test subjects' reaction speed is measured when asked to identify positive and negative adjectives associated to pictures of European- and African-origin faces. The positive coefficient in Table 7 indicates that HOA premiums are higher in cities where the average white resident who takes the IAT has a harder time associating good adjectives with "black" faces, relative to their speed at the same task with "white" faces. We similarly find that state-level HOA premiums are negatively correlated with Brace et al.'s (2002) index of racial tolerance, which is derived from General Social Survey questions from 1974 to 1988. Finally, we examine whether the historical racial composition of CBSAs correlates with current HOA premiums within those CBSAs. We focus on racial composition in 1960, when HOAs were virtually non-existent (see Figure 3). We find that that the share of a CBSA's population that was black in 1960 correlates positively with its modern-day HOA premium, whereas

[34] We also examined the relationship between CBSA-level HOA premiums and mean and median CBSA-level HOA fees collected from Redfin. We found no statistically or economically meaningful relationship between the HOA premium and mean or median HOA fees. It is not theoretically clear that HOA premiums should correlate positively or negatively with dues, as the magnitudes of the premiums should in part reflect the dues and dues might also endogeneously respond to the premiums. We also hesitate to put too much emphasis on these results because Redfin's geographic coverage is limited to selected zip codes (see Figure 2), and for some CBSAs there were very few Redfin listings off which to calculate average and median dues.

the share white in 1960 correlates negatively.[35] This suggests that HOAs are more valued in contexts in which there were large minority populations historically. Taken together, these results are consistent with McKenzie's (1994) assertion that demand for HOAs is driven at least in part by a desire for exclusion.

6 Discussion of Results

HOA fees are analogous to property taxes, in that they are a mandatory ongoing stream of payments required of property owners, generally set by an elected board of HOA officials. It is also reasonable to think that HOA fees are determined via a market process wherein property owners can "vote with their feet," similar to local property taxes (Tiebout, 1956). The consensus view among academic studies is that property taxes are usually partially capitalized into real estate prices (Sirmans et al., 2008). This means that the net effect of raising both property taxes and public expenditures is to lower real estate prices.

This paper and most prior related papers argue that HOAs have the opposite effect. Prices of HOA housing are higher, despite the increased costs of ownership from HOA fees. Redfin.com listings show that HOAs charge an average of \$237 per month in fees. If prices went down by the full value of those fees (i.e., full capitalization), we might expect the average HOA house to cost \$71,100 less than an observationally equivalent non-HOA house.[36] However, our most conservative estimate from Section 5 (from the GAM model) show that HOA houses are priced 4 percent higher than their non-HOA counterparts, or \$13,500 more for the average sale in an HOA subdivision from our sample (from Table 2). The above contrast implies that homeowners obtain \$1.19 in benefits for each \$1 paid in HOA fees.

[35] In results not shown, we find that current racial composition correlates in the same way with HOA premiums as historical racial composition, but the relationships are weaker and not statistically significant. We also find that HOAs tend to be more highly valued in places with lower house prices, higher rates of HOA affiliation, newer buildings, higher rates of owner-occupancy, more individual householders, and smaller CBSA size.

[36] This represents the net present value of a \$2844 perpetual annuity, paid annually and discounted at a 4 percent interest rate. Do and Sirmans (1994) estimate that homeowners discount the value of property taxes at 4 percent and note that most studies use discount rates ranging from 3–6 percent.

The net positive value homeowners derive is consistent with some polls suggesting that most people living in HOA communities have neutral if not positive experiences with their HOA.[37] The value they derive may come from the strong tax (i.e., dues)-benefit link that exists in HOAs as well as the highly tailored amenities HOAs offer. Indeed, our results suggest that HOAs are particularly highly valued in places where local government capacity is more limited and where public spending on police and other public goods is lower. However, our findings also indicate that some may value HOAs as an instrument to achieve a level of homogeneity among neighbors not otherwise possible. From a policy perspective, the extent to which HOAs are reinforcing segregation and possibly limiting access to educational and other opportunities for some groups should be an important consideration going forward.

7 Conclusion

The homeowners association has become a staple of the U.S. housing market. Indeed, HOAs govern 80 percent of houses built in new subdivisions today, and a fifth of all existing single-family homes. However, HOAs have received relatively little academic attention, much less rigorous empirical evaluation, in part as a result of data limitations.

Using data derived from public real estate records for counties housing 90 percent of Americans, this paper documents the proliferation of HOAs across the country and estimates the value homeowners place on them. We find that on average nationwide, homes with an HOA sell for at least 4 percent more than observationally equivalent homes with no HOA. We find that the premium diminishes as homes age and that it tends to be higher for larger houses and for homes in smaller subdivisions. We also find that the HOA premium correlates with local land use regulations and government expenditures in a way that suggests that HOAs are perceived by homebuyers as a valuable substitute for local government.

In addition to studying how HOAs relate to house values, this paper examines the features

[37] For example, in a poll conducted by Zogby Analytics in 2016 for the Community Associations Institute, 87 percent of respondents reported a positive or neutral experience with their HOA (Community Associations Institute, 2016*a*).

of HOA houses and who lives in them. Homes in an HOA are around 400 square feet larger on average and occupy smaller plots of land. The average HOA fee listed on Redfin.com is $2,800 per year, comparable to the $2,200 per year paid by the average homeowner in property taxes (Kiernan, 2018). HOA residents tend to be higher income and are disproportionately white and Asian relative to non-HOA residents. HOA residents also tend to live on less racially diverse blocks than do their non-HOA counterparts in nearby neighborhoods. Our methodology exploiting information on mortgage riders could be useful to researchers interested in not only HOAs, but also other aspects of homes that are recorded in mortgage documents.

We hope our research spurs additional work investigating the causes and consequences of HOAs in the U.S. as well as those of similar entities in other countries. As HOAs play a significant role in constraining the activities of property owners, but at the same time provide a variety of amenities and address certain externalities through zoning-like policies, further study of their prevalence, functions, and impacts is critical to understanding the dynamics of local housing markets, and in particular the market for newly constructed single-family homes. If HOAs continue to proliferate, it also raises a host of questions about the implications of HOAs for residential segregation and inequality, local government finances and activities, and the extent to which the threat of voting with one's feet serves to discipline the market and ensure choice.

References

Barton, S. E. and Silverman, C. J. (1994), *Common Interest Communities: Private Governments and the Public Interest*, Institute of Governmental Studies Press, Berkeley, CA.

Benson, S. E. and DeBat, D. (2014), *Escaping Condo Jail*, CreateSpace Independent Publishing Platform.

Brace, P., Sims-Butler, K., Arceneaux, K. and Johnson, M. (2002), 'Public Opinion in the American States: New Perspectives Using National Survey Data', *American Journal of Political Science* **46**(1), 173–189.

Cannaday, R. E. (1994), 'Condominium Covenants: Cats, Yes; Dogs, No', *Journal of Urban Economics* **35**(1), 71–82.

Cheung, R. (2008), 'The Interaction Between Public and Private Governments: An Empirical Analysis', *Journal of Urban Economics* **63**(3), 885–901.

Cheung, R., Cunningham, C. and Meltzer, R. (2014), 'Do Homeowners Associations Mitigate or Aggravate Negative Spillovers from Neighboring Homeowner Distress?', *Journal of Housing Economics* **24**, 75–88.

Cheung, R. and Meltzer, R. (2013), 'Homeowners Associations and the Demand for Local Land Use Regulation', *Journal of Regional Science* **53**(3), 511–534.

Cheung, R. and Meltzer, R. (2014), 'Why and Where Do Homeowners Associations Form?', *Cityscape* **16**(3), 69–92.

Community Associations Institute (2016*a*), Homeowner Satisfaction Survey, Technical report, Foundation for Community Association Research.

Community Associations Institute (2016*b*), Large-Scale Association Survey Results, Technical report, Foundation for Community Association Research.

Do, A. Q. and Sirmans, C. F. (1994), 'Residential Property Tax Capitalization: Discount Rate Evidence from California', *National Tax Journal* **47**(2), 341–48.

Esquivel, A. and Alvayay, J. (2014), A Guide to Understanding Residential Subdivisions in California, Technical report, California Bureau of Real Estate.

Foldvary, F. (1994), *Public Goods and Private Communities*, Edward Elgar, London.

Glasze, G., Webster, C. and Frantz, K. (2006), *Private Cities: Global and Local Perspectives*, Routledge.

Gordon, T. (2003), 'Moving Up by Moving Out? Planned Developments and Residential Segregation in California', *Urban Studies* **41**(2), 441–461.

Groves, J. R. (2008), 'Finding the Missing Premium : An Explanation of Home Values within Residential Community Associations', *Land Economics* **84**(2), 188–208.

Guberman, R. (2004), 'Home is Where the Heart Is', *Legal Affairs* .

Hastie, T. and Tibshirani, R. (1990), *Generalized Additive Models*, Chapman and Hall/CRC, Boca Rotan, FL.

Helsley, B. R. W. and Strange, W. C. (2000), 'Social Interactions and the Institutions of Local Government', *American Economic Review* **90**(5), 1477–1490.

Helsley, R. W. and Strange, W. C. (1998), 'Private Government', *Journal of Public Economics* **69**(2), 281–304.

Hopkins, E. A. (2016), 'The Impact of Community Associations on Residential Property Values', *Housing and Society* **43**(3), 157–167.

Hughes, W. and Turnbull, G. (1996), 'Uncertain Neighborhood Effects and Restrictive Covenants', *Journal of Urban Economics* **39**, 160–172.

Kiernan, J. S. (2018), '2018's Property Taxes by State', Wallethub.com.

Lopez, S. (2017), 'GitHub Repository'. GitHub.com.
URL: *https://github.com/slopez15*

Lucas, W. (2013), *Neighbors at War: The Creepy Case Against Your Homeowner's Association*, Hogback Publishing, Denver, CO.

Mandell, L. (2016), 'What Are HOA Fees? How These Dues Make Homeownership Easier Than Ever'. Realtor.com.
URL: *https://www.realtor.com/advice/buy/what-are-hoa-fees/*

McCabe, B. C. (2011), 'Homeowners Associations as Private Governments: What We Know, What We Don't Know, and Why It Matters', *Public Administration Review* **71**(4), 535–542.

McKenzie, E. (1994), *Privatopia: Homeowner Associations and the Rise of Residential Private Government*, Yale University Press, New Haven, CT.

McKenzie, E. (2011), *Beyond Privatopia*, Urban Institute Press, Washington, DC.

Meltzer, R. (2013), 'Do Homeowners Associations Affect Citywide Segregation? Evidence From Florida Municipalities', *Housing Policy Debate* **23**(4), 688–713.

Meltzer, R. and Cheung, R. (2014), 'How Are Homeowners Associations Capitalized into Property Values?', *Regional Science and Urban Economics* **46**(C), 93–102.

Nelson, R. H. (2005), *Private Neighborhoods and the Tranformation of Local Government*, Urban Institute Press, Washington, DC.

Nelson, R. H. (2009), 'The Puzzle of Local Double Taxation: Why Do Private Community Associations Exist?', *Independent Review* **13**(3), 345–365.

Nowak, A. and Smith, P. (2017), 'Textual Analysis in Real Estate', *Journal of Applied Econometrics* **32**(1), 896–918.

Oster, E. (2017), ‘Unobservable Selection and Coefficient Stability: Theory and Evidence’, *Journal of Business and Economic Statistics* **Forthcoming**.

Rogers, W. H. (2006), ‘A Market for Institutions: Assessing the Impact of Restrictive Covenants on Housing’, *Land Economics* **82**(4), 500–512.

Rogers, W. H. (2010), ‘Measuring the Price Impact of Municipal Incorporation on Homeowner Associations’, *Land Economics* **86**(1), 91–116.

Rosen, S. (1974), ‘Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition’, *Journal of Political Economy* **82**(1), 34–55.

Sirmans, G. S., Gatzlaff, D. H. and Macpherson, D. a. (2008), ‘The History of Property Tax Capitalization in Real Estate’, *Journal of Real Estate Literature* **16**(3), 327–344.

Speyrer, J. F. (1989), ‘The Effect of Land-Use Restrictions on Market Values of Single-Family Homes in Houston’, *The Journal of Real Estate Finance and Economics* **2**(2), 117–130.

Tiebout, C. M. (1956), ‘A Pure Theory of Local Expenditures’, *Journal of Political Economy* **64**(5), 416–424.

Wang, F. (2014), ‘Determinants of the Effectiveness of Chinese Homeowner Associations in Solving Neighborhood Issues’, *Urban Affairs Review* **50**(3), 311–339.

Wood, S. (2017), *Generalized Additive Models: An Introduction with R*, second edn, Chapman and Hall/CRC, Boca Rotan, FL.

Table 1. Data Elements Used from ZTRAX

Property Characteristic	Description
Building area	Usable area of primary home, in square feet
Lot area	Land area, in square feet
Garage area	Usable area of the garage, in square feet
Other structure area	Usable area of any secondary buildings, in square feet
Implied property tax rate	Property tax bill divided by assessed value in most recent year
Owner occupied	Whether the property is occupied by its owner
Year built	Year the main building was constructed
Bedrooms	Number of bedrooms
Baths	Number of bathrooms
Total rooms	Total number of rooms in the house
Topography	Note about the property's terrain (e.g., hilly/level ground)
Tile roof	Indicator for a tiled roof
Near golf course	Indicator for proximity to a golf course
Waterfront	Indicator for having a view of water
Flooring	Type of interior flooring: carpet, wood, or other
Exterior wall	Type of exterior walls: brick, siding, stucco, or wood
Fence	Indicator for having a fence
Fireplace	Indicator for having a fireplace
Pool	Indicator for having a pool
Deck	Indicator for having a deck
Census block	Most basic census geography; includes state, county, and tract
Legal subdivision name	Links groups of parcels previously subdivided from one larger parcel
Property address latitude	TIGER block face coordinates, interpolated based on house number
Transaction Characteristic	**Description**
Planned unit development rider	Indicator for inclusion of a PUD rider, on mortgage records
Condominium Rider	Indicator for inclusion of a condominium rider, on mortgage records
Sale price amount	Price paid for property deed transfer
Sale price amount type	Indicator of how sale price was observed
Transaction date	Date the transaction was finalized
Intra-family transfer flag	Indicator for buyer and seller having a family relationship
Deed type	Type of property deed: warranty, foreclosure, or other
Public record type	Deed transfer, mortgage, deed transfer with concurrent mortgage

Notes: This table lists and briefly explains the data elements relating to land parcels and real estate transactions used for analysis from the Zillow Transaction and Assessment Dataset (ZTRAX).

Table 2. Descriptive Housing Statistics, by HOA Status

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All single-family parcels				Built since 1960, transacted since 2005, with accompanying mortgage			
	HOA Measure				HOA Measure			
	House-Based		Neigh.-Based		House-Based		Neigh.-Based	
	Non-HOA	HOA	Non-HOA	HOA	Non-HOA	HOA	Non-HOA	HOA
Unduplicated parcels								
Building area (sqft)	1,907	2,299	1,793	2,274	1,907	2,364	1,828	2,355
Lot area (sqft)	15,954	11,241	12,816	10,121	16,584	10,954	13,151	9,750
Garage area (sqft)	279	321	279	324	294	339	299	343
Other structure area (sqft)	56	36	54	35	53	35	50	32
Implied property tax rate	2.50%	2.90%	2.50%	3.00%	2.50%	2.80%	2.50%	2.90%
Pool	11%	14%	12%	14%	8%	10%	9%	10%
Deck	11%	10%	9%	9%	12%	8%	10%	7%
Bedrooms	3.4	3.5	3.3	3.5	3.4	3.5	3.4	3.5
Baths	2.2	2.6	2.2	2.6	2.2	2.6	2.2	2.6
Total rooms	6.8	6.9	6.7	6.9	6.8	7.0	6.8	7.0
Tile roof	5%	15%	5%	16%	4%	16%	5%	17%
Golf	0.10%	0.40%	0.00%	0.40%	0.10%	0.30%	0.00%	0.30%
Waterfront	1.10%	1.20%	1.10%	1.20%	0.70%	1.30%	0.80%	1.30%
Flooring---carpet	16%	19%	17%	20%	15%	21%	16%	21%
Flooring---wood	2%	1%	2%	1%	2%	1%	2%	1%
Fence	4%	6%	5%	6%	3%	6%	4%	7%
Sample size (millions)	36.6	9.6	24.1	9.1	2.2	1.5	1.5	1.4
Built since 1960	21.6	9.1	14.7	10.3				
+ transacted since 2005	4.0	2.3	2.8	2.2				
+ with mortgage	2.2	1.5	1.5	1.4				
Percent of Sample	79%	21%	73%	27%	59%	41%	52%	48%
Transactions								
Sale Price ($1,000, nominal)					$287	$352	$282	$352
Transaction Date (year)					2009	2009	2009	2009
Age at sale (years)					24	6	25	5
Deed type---warranty					61%	73%	60%	74%
Deed type---foreclosure					3%	7%	3%	8%
Deed type---other					36%	20%	37%	19%
Sample size (millions)					5.7	4.4	4.2	3.8

Notes: Selected property and sale characteristics of single-family houses recorded in the Zillow Transaction and Assessment Dataset (ZTRAX). Columns (1) and (2) include all houses for which there is record of a mortgage, allowing a house-specific determination of HOA status. Columns (3) and (4) include houses for which a neighborhood-specific HOA status can be determined, based on mortgage riders for that house and neighboring houses as described in Section 3. Columns (5)-(8) mirror columns (1)-(4), but the sample is limited to houses used for regressions in Section 5; specifically, those built post-1960 and sold since 2005, concurrent with a new mortgage.

Table 3. Income and Race, by HOA Status

	HOA	Non-HOA
Block Group Level		
Median Annual Household Income	\$83,401	\$59,656
Percent White	75.0%	72.9%
Percent Black	8.9%	12.3%
Percent Asian	7.5%	4.4%
Percent Other Races	8.6%	10.4%
Block Level		
Percent White	74.8%	73.0%
Percent Black	9.1%	12.2%
Percent Asian	7.9%	4.3%
Percent Other Races	8.2%	10.5%

Notes: Calculated as the weighted average of census block group (top panel) or census block (bottom panel) characteristics, where weights are population multiplied by the percentage of single-family houses in (or not in) an HOA. House-based HOA status is used.

Table 4. Racial Diversity of HOA Blocks

	Diversity			
HOA Block	0.062***	0.013*	-0.011***	-0.018***
	(0.007)	(0.007)	(0.004)	(0.003)
Log Population		0.063***	0.054***	0.041***
		(0.004)	(0.002)	(0.001)
State FEs			✓	
County FEs				✓
Observations	3.6M	3.6M	3.6M	3.6M

Notes: Unit of observation is a census block. Diversity represents the probability that any two randomly drawn individuals from a census block are of different races (where the races include white, black, Asian, and other). Standard errors (in parentheses) are adjusted for heteroskedasticity and clusters at the county level. Significant at the *10% level, **5% level and ***1% level.

Table 5: National Regression Results

	Log Price					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS Model				GAM Model	
	HOA Measure				HOA Measure	
	House-Based		Neigh.-Based		House-Based	Neigh.-Based
HOA	0.058***	0.052***	0.063***	0.056***	0.044***	0.037***
	(0.003)	(0.005)	(0.005)	(0.007)	(0.0003)	(0.0003)
Log Building Area	0.499***	0.512***	0.503***	0.512***		
	(0.011)	(0.016)	(0.012)	(0.018)		
Log Lot Area	0.105***	0.113***	0.117***	0.125***		
	(0.005)	(0.007)	(0.006)	(0.008)		
Age at Sale (Years)	-0.005***	-0.006***	-0.006***	-0.007***		
	(0.0002)	(0.0004)	(0.000)	(0.001)		
Implied Prop. Tax (%)	-0.101	-0.020	-0.150	-0.037		
	(0.122)	(0.171)	(0.114)	(0.171)		
Deed Type - Foreclosure	0.022	-0.009	0.030	0.001		
	(0.019)	(0.025)	(0.028)	(0.036)		
Deed Type - Warranty	0.089***	0.083***	0.101***	0.097***		
	(0.013)	(0.020)	(0.015)	(0.024)		
Owner-Occupied	0.011***	0.019***	0.009***	0.017***		
	(0.003)	(0.003)	(0.003)	(0.003)		
Pool	0.087***	0.091***	0.085***	0.089***		
	(0.005)	(0.007)	(0.006)	(0.009)		
Waterfront	0.214***	0.187***	0.164***	0.138***		
	(0.029)	(0.040)	(0.023)	(0.027)		
Exterior Wall - Brick	0.035***	0.038***	0.026***	0.030***		
	(0.005)	(0.008)	(0.005)	(0.010)		
Exterior Wall - Siding	-0.030***	-0.028***	-0.032***	-0.032***		
	(0.005)	(0.007)	(0.009)	(0.009)		
Month FEs	✓		✓			
Block Group FEs	✓		✓			
Block Group × Month FEs		✓		✓		
Observations	10.1M	10.1M	8.0M	8.0M	10.1M	8.0M

Notes: This table presents coefficient estimates from linear regressions and GAMs. In addition to the variables listed in the table, other controls include number of bedrooms, number of bathrooms, total number of rooms, floor material, other exterior materials (stucco or wood), log garage space, log other building area, presence of a fireplace, precens of a deck, roof material, presence of a fence, access to golf course, and topographical features. Information about the property variables we include appear in Table 1, and descriptive statistics appear in Table 2. Standard errors (in parentheses) are adjusted for heteroskedasticity and clusters at the county level. Significant at the *10% level, **5% level and ***1% level.

Table 6: Tests for Heterogeneous Effects of HOA Membership by Property and Neighborhood Characteristics

	Log Price				
	(1)	(2)	(3)	(4)	(5)
HOA	0.081***	0.051***	0.054***	0.058***	0.067***
	(0.006)	(0.003)	(0.003)	(0.003)	(0.004)
HOA × Age at Sale (x10)	-0.017***				
	(0.004)				
HOA × Log Building Area		0.043***			
		(0.003)			
HOA × Log Lot Area			0.035***		
			(0.002)		
HOA × Implied Prop. Tax (%)				-0.001	
				(0.003)	
HOA × Large Subdivision					-0.018***
					(0.004)
Month FEs	✓	✓	✓	✓	✓
Block Group FEs	✓	✓	✓	✓	✓
Observations	10.1M	10.1M	10.1M	10.1M	6.8M

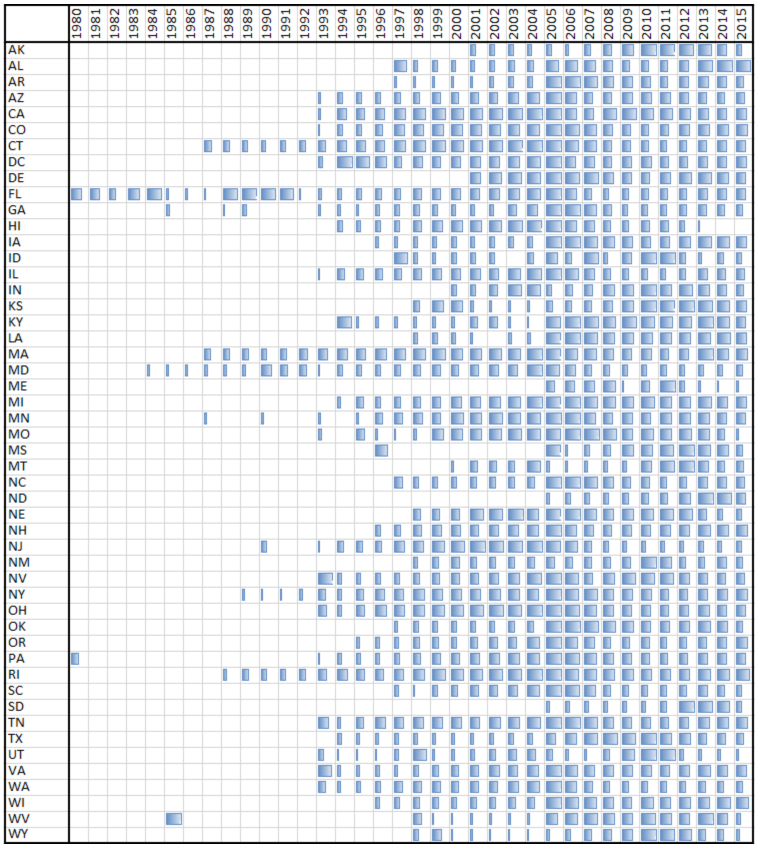
Notes: This table presents coefficient estimates from linear regressions. Controls included in the regressions but not shown in the table include all those used in the regression shown in column (1) of Table 5. Standard errors (in parentheses) are adjusted for heteroskedasticity and clusters at the county level. Significant at the *10% level, **5% level and ***1% level.

Table 7: Relationships Between CBSA and State-Level Characteristics and HOA Premiums

	Coefficient	Standard Error	CBSAs
Local Government Characteristics			
Zoning Strength, WRLURI	-0.0205***	(0.0074)	179
Total Local Government Revenues†	-0.0015	(0.0016)	194
Property Tax Revenues†	-0.0089**	(0.0045)	194
Total Local Government Expenditures†	-0.0031**	(0.0012)	194
Expenditures on Roads†	-0.025	(0.0265)	194
Expenditures on Police†	-0.0502*	(0.0275)	194
Expenditures on Parks & Recreation†	-0.0049	(0.0529)	194
Expenditures on Administration†	-0.0548***	(0.0120)	194
Current Socioeconomic Characteristics			
Gini Coefficient	0.4702**	(0.2201)	205
IAT Test for White Preference	0.2589**	(0.1145)	205
Racial Integration Z-Score (State-Level)	-0.0680*	(0.0371)	34
Historical Racial Composition			
Share Black in 1960	0.1060***	(0.0355)	193
Share White in 1960	-0.1117***	(0.0374)	193

Notes: Each row presents the coefficient estimate from a regression of the CBSA- or state-level HOA premium on the specified CBSA- or state-specific characteristic. †Scaled by number of housing units in the CBSA. Standard errors (in parentheses) are heteroskedasticity robust. Significant at the *10% level, **5% level and ***1% level.

Figure 1: Geographic and Temporal Extent of ZTRAX Transaction Records



Notes: Box size in each cell represents the number of single-family home sales recorded in that state an dyear, relative to the year with the most sales.

Figure 2: HOA Fees

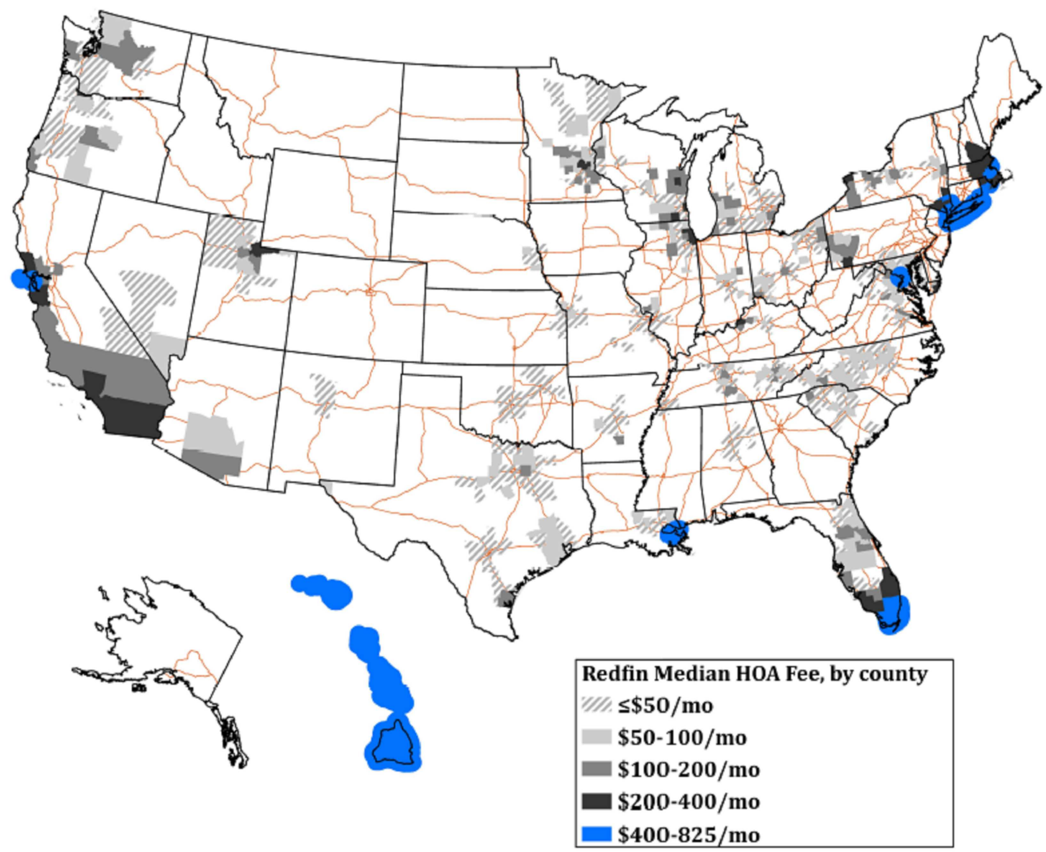


Figure 3: Share of U.S. Single Family Homes Built with an HOA, by Year Built

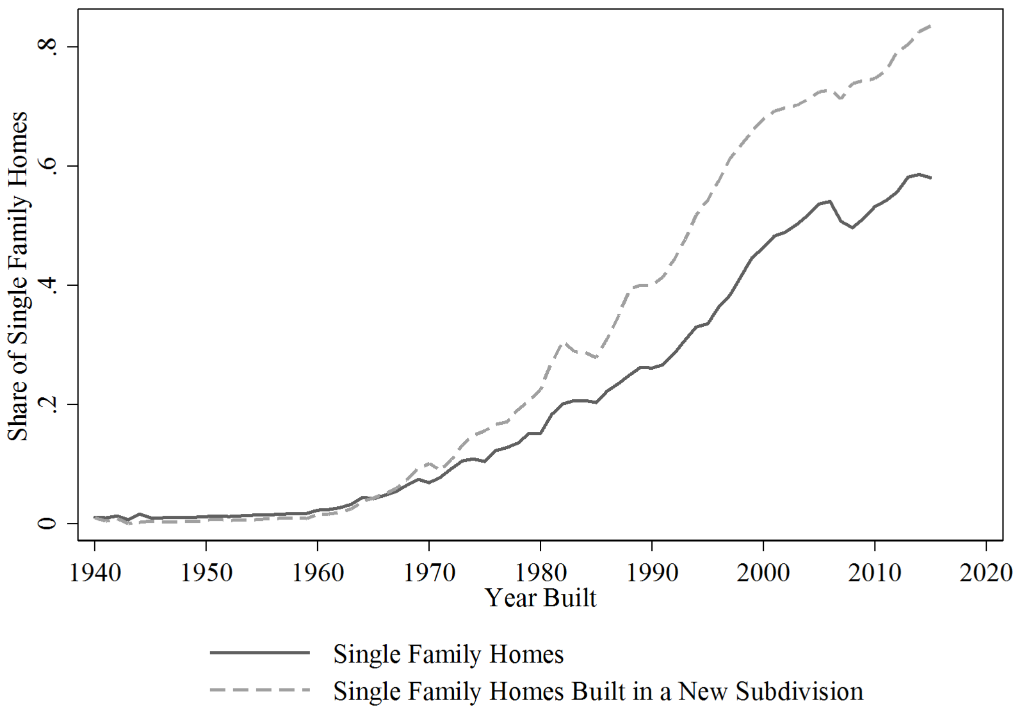


Figure 4: Single-Family Housing with an HOA, by Census Division and County

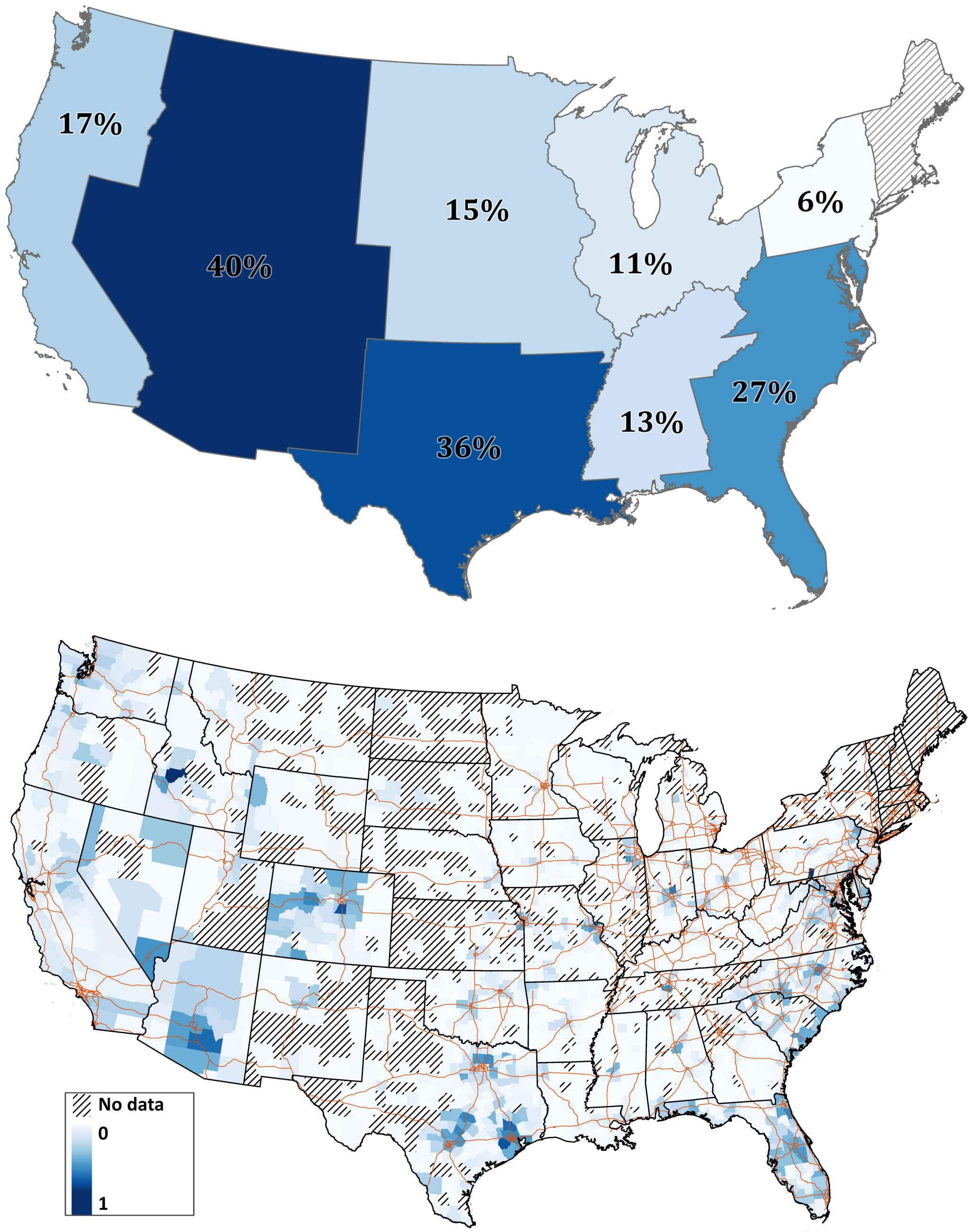


Figure 5: Single-Family Housing Built in New Subdivisions with an HOA, by Census Division and County

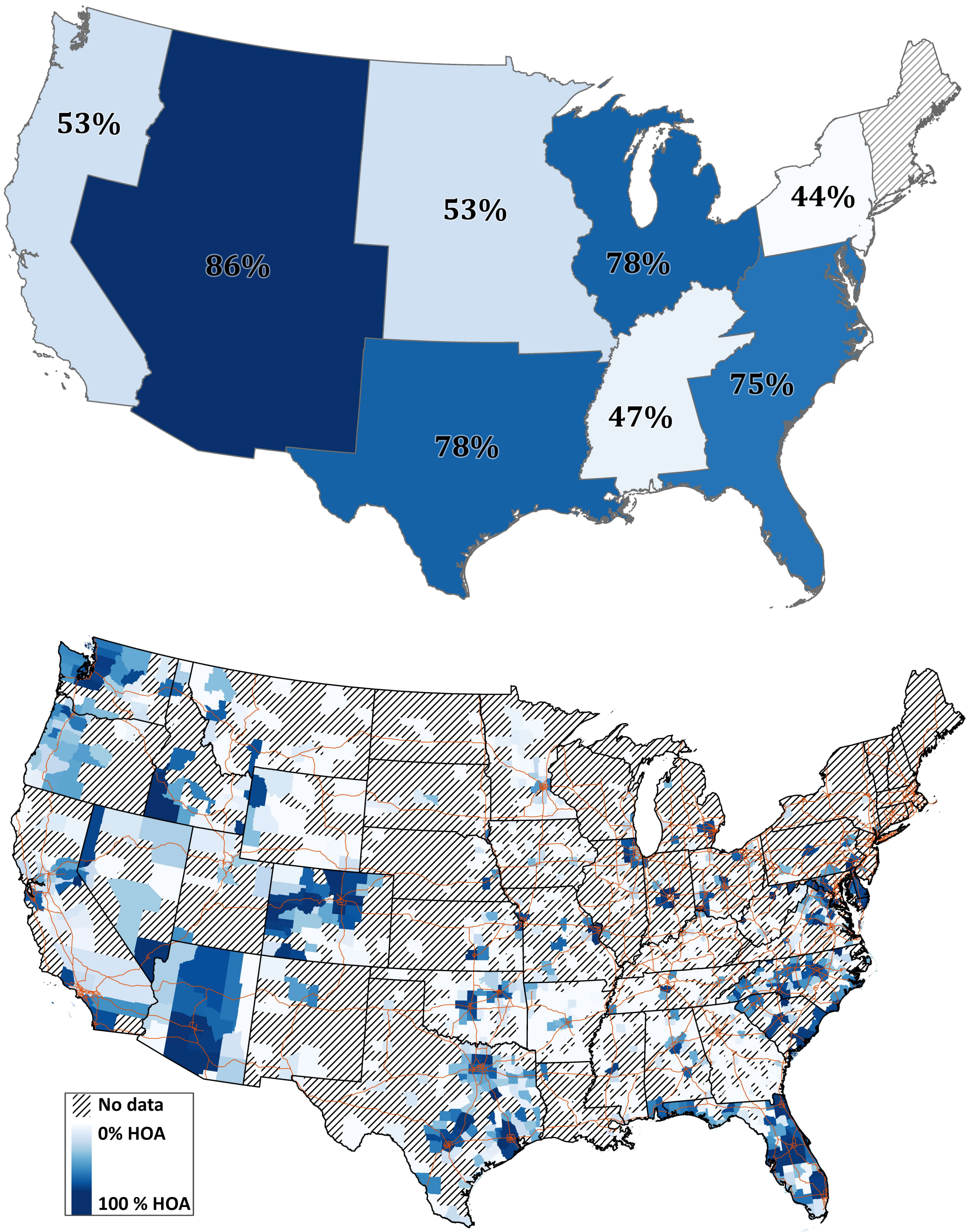
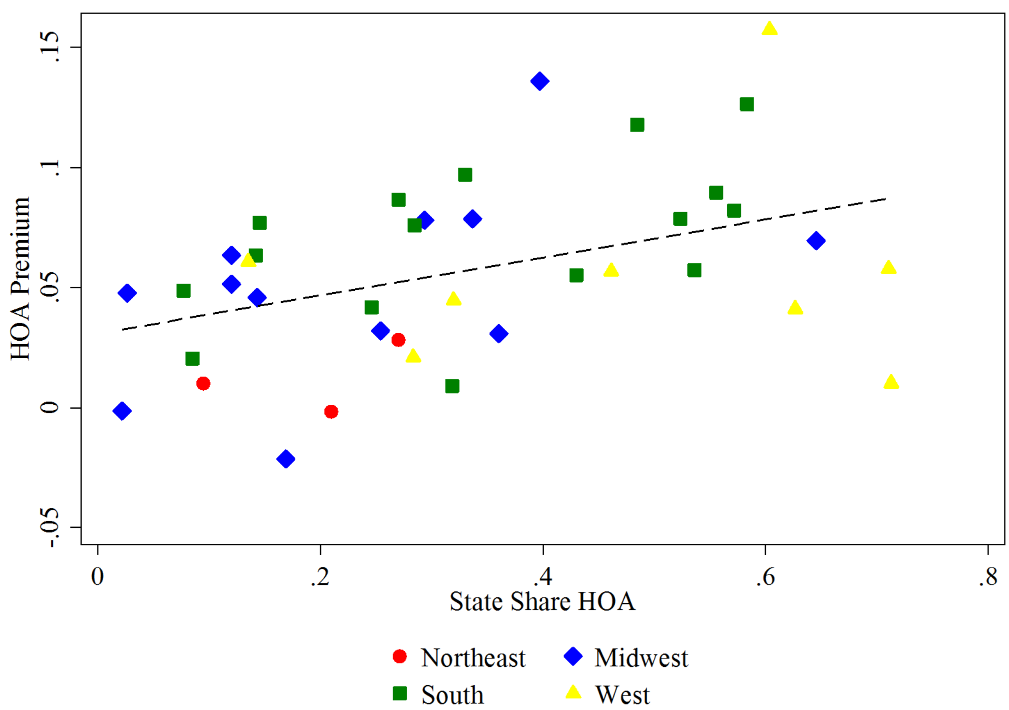


Figure 6: State-Specific HOA Premiums by Share of Transacted Homes in HOAs



Appendix A

Clustering Algorithm Used to Attribute HOA Status to Neighborhoods

DBSCAN ("density-based spatial clustering of applications with noise") is an unsupervised clustering algorithm used to find groups of homes that likely share a common HOA status. The benefit of DBSCAN, relative to grouping by subdivision, is that it can be applied to houses where subdivisions are not recorded. The disadvantages of DBSCAN relative to grouping by subdivision is that the process is somewhat arbitrarily chosen, it is difficult to explain, and DBSCAN does not include homes in an HOA if, for some reason, they were built at a different time than the other homes. Altogether, it is worth grouping by both subdivision and DBSCAN cluster to (1) include more homes in the neighborhood-level measure of HOA status and (2) use the methods to validate one another.

Methodology

DBSCAN works by grouping houses that are within x distance of at least n other houses into core clusters. Any additional houses that are within x distance of at least one house in the core cluster are added to the periphery of the cluster. Remaining houses are not placed in a cluster. The distance used is simple Euclidean distance plus a penalty for dierence in year built. For this paper, clusters are formed with at least $n = 5$ houses in their core, located within $x = 75$ meters of each other if built in the same year or within $\frac{x}{2} = 37.5$ meters of each other if built one year apart. These parameters were chosen to balance (1) confidence that houses in the same cluster really do share the same HOA status and (2) inclusion of as many HOA-member houses as possible, which helps avoid focusing on a small and potentially unrepresentative group. DBSCAN is implemented using the scikit-learn package in Python, which provides full documentation.

Results

Figure A1 shows the degree to which houses in groups formed using DBSCAN and by subdivision all have, or do not have, HOA mortgage flags. If the groups matched HOA boundaries perfectly and all house-level indicators of HOA status were correct, the histograms would show spikes at 0 percent and 100 percent with nothing in the middle. Table A1 demonstrates that grouping by subdivision or by DBSCAN results in the same HOA designation for over 99 percent of the 14 million houses covered by both methods.

Figure A1: Histograms of the Percentage of Homes within a Subdivision or Cluster Having a House-Level HOA Indicator

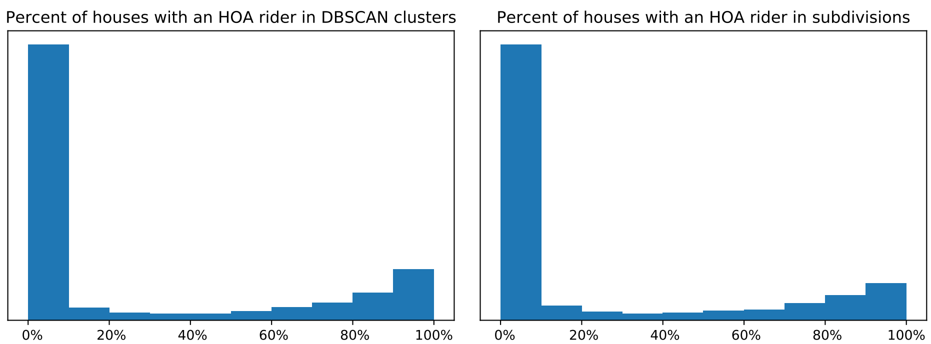


Table A1: Agreement Between Subdivision and DBSCAN Methods for Assigning Neighborhood HOA Status

		Cluster Classification	
		No HOA	HOA
Subdivision	No HOA	8,487,937	40,900
Classification	HOA	17,783	5,513,593

Appendix B

Each line in the table below presents the coefficient and standard error estimates for HOA membership (plus sample size) from OLS and GAM regressions of house price on housing characteristics identical to the ones found in columns (1), (2), and (5) of Table 5 (OLS with block group and month fixed effects, OLS with block group-by-month fixed effects, and the GAM model), but restricted to sales from the corresponding CBSA. The table includes the largest 50 CBSAs in the sample by population.

CBSA	(1) OLS	(2) OLS	(3) GAM	(4) Obs.
New York-Newark-Jersey City, NY-NJ-PA	-0.012	-0.020	-0.013	191616
	(0.003)	(0.004)	(0.002)	
Chicago-Naperville-Elgin, IL-IN-WI	0.033	0.027	0.028	254648
	(0.002)	(0.002)	(0.002)	
Dallas-Fort Worth-Arlington, TX	0.068	0.001	0.040	4359
	(0.032)	(0.082)	(0.02)	
Houston-The Woodlands-Sugar Land, TX	0.204	0.055	0.078	3999
	(0.042)	(0.093)	(0.022)	
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.049	0.038	0.045	174481
	(0.002)	(0.003)	(0.002)	
Washington-Arlington-Alexandria, DC-VA-MD-WV	0.014	0.006	0.024	167249
	(0.002)	(0.002)	(0.002)	
Miami-Fort Lauderdale-West Palm Beach, FL	0.022	0.019	0.018	266536
	(0.002)	(0.003)	(0.002)	
San Francisco-Oakland-Hayward, CA	0.009	0.009	0.014	157246
	(0.003)	(0.003)	(0.002)	
Phoenix-Mesa-Scottsdale, AZ	0.048	0.041	0.060	637585
	(0.001)	(0.001)	(0.001)	
Riverside-San Bernardino-Ontario, CA	0.024	0.020	0.048	486965
	(0.001)	(0.001)	(0.001)	
Detroit-Warren-Dearborn, MI	-0.032	-0.033	-0.027	31868
	(0.006)	(0.013)	(0.005)	
Seattle-Tacoma-Bellevue, WA	0.058	0.053	0.067	269475
	(0.002)	(0.002)	(0.002)	
Minneapolis-St. Paul-Bloomington, MN-WI	0.058	0.068	0.065	55767
	(0.005)	(0.007)	(0.005)	
San Diego-Carlsbad, CA	0.019	0.022	0.020	190429
	(0.002)	(0.002)	(0.001)	
Tampa-St. Petersburg-Clearwater, FL	0.065	0.062	0.088	287488
	(0.002)	(0.003)	(0.002)	
St. Louis, MO-IL	0.068	0.062	0.070	86708
	(0.003)	(0.004)	(0.003)	
Baltimore-Columbia-Towson, MD	0.075	0.073	0.064	28703
	(0.005)	(0.007)	(0.004)	
Denver-Aurora-Lakewood, CO	0.025	0.019	0.047	289605
	(0.002)	(0.002)	(0.001)	
Charlotte-Concord-Gastonia, NC-SC	0.093	0.095	0.070	223357
	(0.003)	(0.003)	(0.003)	

Pittsburgh, PA	0.034	0.030	0.021	65436
	(0.006)	(0.008)	(0.005)	
Portland-Vancouver-Hillsboro, OR-WA	0.045	0.044	0.062	138205
	(0.002)	(0.003)	(0.002)	
Orlando-Kissimmee-Sanford, FL	0.053	0.046	0.069	224913
	(0.002)	(0.003)	(0.002)	
Sacramento--Roseville--Arden-Arcade, CA	0.011	0.008	0.029	176576
	(0.002)	(0.002)	(0.002)	
Cincinnati, OH-KY-IN	0.080	0.071	0.080	108652
	(0.004)	(0.005)	(0.003)	
Las Vegas-Henderson-Paradise, NV	-0.004	-0.004	0.010	339585
	(0.002)	(0.002)	(0.001)	
Kansas City, MO-KS	0.054	-0.275	0.070	1467
	(0.042)	(0.181)	(0.020)	
Cleveland-Elyria, OH	0.074	0.066	0.068	68868
	(0.005)	(0.007)	(0.004)	
Columbus, OH	0.056	0.040	0.069	88884
	(0.004)	(0.005)	(0.003)	
San Jose-Sunnyvale-Santa Clara, CA	0.026	0.028	0.049	83769
	(0.005)	(0.006)	(0.004)	
Nashville-Davidson--Murfreesboro--Franklin, TN	0.062	0.067	0.060	110001
	(0.004)	(0.005)	(0.004)	
Virginia Beach-Norfolk-Newport News, VA-NC	0.016	0.011	0.033	69989
	(0.004)	(0.004)	(0.004)	
Jacksonville, FL	0.052	0.049	0.064	141650
	(0.003)	(0.004)	(0.003)	
Memphis, TN-MS-AR	0.040	0.036	0.029	82043
	(0.004)	(0.004)	(0.004)	
Oklahoma City, OK	0.146	0.165	0.122	97174
	(0.006)	(0.008)	(0.005)	
Louisville/Jefferson County, KY-IN	0.041	0.034	0.059	54468
	(0.004)	(0.005)	(0.004)	
Richmond, VA	0.012	0.006	0.032	29915
	(0.006)	(0.007)	(0.006)	
New Orleans-Metairie, LA	0.067	0.062	0.056	7988
	(0.01)	(0.014)	(0.009)	
Raleigh, NC	0.084	0.097	0.073	168938
	(0.003)	(0.004)	(0.003)	
Birmingham-Hoover, AL	0.071	0.065	0.077	27167
	(0.009)	(0.012)	(0.008)	
Buffalo-Cheektowaga-Niagara Falls, NY	0.105	0.116	0.087	23746
	(0.012)	(0.018)	(0.010)	
Tucson, AZ	0.067	0.058	0.069	102041
	(0.003)	(0.004)	(0.002)	
Tulsa, OK	0.098	0.106	0.087	47954
	(0.007)	(0.009)	(0.006)	
Fresno, CA	0.069	0.067	0.069	73993
	(0.004)	(0.004)	(0.003)	
Omaha-Council Bluffs, NE-IA	-0.007	-0.034	0.005	32486
	(0.007)	(0.010)	(0.007)	
Bakersfield, CA	0.046	0.060	0.057	86593
	(0.005)	(0.005)	(0.004)	
Greenville-Anderson-Mauldin, SC	0.168	0.158	0.178	24099
	(0.008)	(0.012)	(0.008)	

Knoxville, TN	0.069	0.063	0.074	47351
	(0.006)	(0.006)	(0.006)	
Oxnard-Thousand Oaks-Ventura, CA	0.037	0.033	0.049	54220
	(0.004)	(0.006)	(0.004)	
Allentown-Bethlehem-Easton, PA-NJ	0.023	0.021	0.032	27392
	(0.007)	(0.010)	(0.006)	
Dayton, OH	0.017	-0.022	0.028	11669
	(0.016)	(0.023)	(0.013)	