

# A Contemporary Reevaluation of Relative Affordability in the American Housing Market

DATA698-004 Capstone Project Mid-Term Draft

Zachary Palmore

## Contents

### 0.1 Introduction

It is important that housing be affordable because people need a place to live. Failure to provide adequate affordable housing has been shown to weaken local economies, reduce quality of life, and increase financial and mental health stress in the population. We note the two major national indices that focus on estimating a qualifying income as a proportion of the median family income for a given area to measure affordability. While these incorporate interest rates and make reasonable assumptions about down payments, amortization, and other general requirements for consumers to be granted a mortgage loan, they leave out a plethora of obstacles faced by consumers, especially for first-time homebuyers. Instead, focus should be placed on direct consumer access to housing without loan considerations and incorporate a reasonable safety net in these index calculations. We propose a new method of calculating housing affordability for consumers in the 21st century, one that incorporates a set of minimum thresholds that better exemplify the economic challenges of the modern home buying process.

There are two main indices used to calculate housing affordability in the United States, the National Association of Realtors (NAR) Housing Affordability Index and the California Association of Realtors (CAR) Index. When these national methods are applied at the city or county level it is considered a Local Housing Affordability Index (LHAI). We intend to focus on U.S. metropolitan statistical areas (MSAs) to make well known the local affordability challenges. These major indices fail to consider the full scope of challenges that potential homebuyers, especially first-time homebuyers, face when making the decision to purchase a home. The formula for the home affordability index (HAI) is as follows:

$$HAI = \left( \frac{\text{Median Family Income}}{\text{Qualifying Income}} \right) \times 100$$

Where the qualifying income value contains assumptions about the down payment, average federal mortgage interest rate, principal, interest on the loan, taxes (including property taxes), insurance, with the last four building the acronym PITI. The CAR Index incorporates a minimum income needed to qualify for the loan and income distribution of the area. In both cases, applying this equation to incomes at local levels improves our understanding of how affordable the area is but is narrowly focused on the average consumer's income having little if any debts or financial obligations. However, debt and financial obligations are realities of modern life.

For example, consumers who are of home buying age (25+) are mostly millennials which according to the 2020 Experian State of Credit report means they have an average of about \$27,251 in non-mortgage debts. It is important to understand that when the parents of these millennials (Colloquially Gen X's and Baby Boomers) were purchasing houses, they had far fewer non-mortgage debts on average, were often given generous loans while needing fewer qualifications, and could have used the same amount of non-mortgage

debts as millennials towards down-payments. Such an amount in their times would have more than covered the necessary upfront costs of becoming homeowners. Today, this is not the case. Lending practices have tightened, and costs have increased while the income of millennials has stabilized. These factors must be considered in the affordability index since they act as barriers to ensuring access to comparable housing.

In the United States the average cost to purchase a house has continuously increased since at least 1921 when the Bureau of Labor Statistics (BLS) began collecting data on the Consumer Price Index, or CPI, which included housing costs. As with all consumer goods, inflation has been a significant factor contributing to this increase in house prices. Market adjustments in personal and household income have also been made in response to inflation, generally increasing an individual's gross income. Overall, this has raised the chances of homeownership in the U.S. through the collection of individuals who chose to save money, reduce their debts, and leverage credit to finance their homes with mortgages. This trend has lowered the upfront costs needed to purchase a house while simultaneously limiting the burden placed on some homeowners who lack the means to afford their house on their own. Unfortunately, today, there is far more to purchasing a house than simply having enough money. For most Americans, becoming homeowners is a process.

Consider that it is a standard recommendation by financial institutions that any mortgage bearer not consume more than 28% of that person's monthly gross income on a monthly mortgage payment. In real dollars, this means if you make \$5,000 in a month, you should not pay more than \$1,400 on your mortgage payment. Additionally, the individual interested in purchasing a house using a mortgage, should expect to make a down payment on that house. Amounts vary by lender and the federal government has several programs available to assist prospective homebuyers should they fulfill the requirements of the program. For the average American these range from as low as 3% to 10% minimum down payments on government-backed specialty and conventional home loans. These are common among first-time home buyers. Private loans, however, require 20% as a minimum unless the individual has their mortgage insured. This makes understanding the home-buying process quite difficult and approval in any case is also dependent upon credit, existing real assets, financial assets, and other forms of wealth.

Evidence suggests that housing today is much less affordable than it was over the past century. In the 1920's it was common for an individual to purchase a home without a mortgage. Instead, the individual would pay for the house outright, fully owning the home and its property upon transfer of the deed. According to James L. Butkiewicz, an economics professor at the University of Delaware, the first home loans were typically fixed short-term loans of 3 - 5 years in length and given for no more than 50% of the value of the property. This trend of loaning less than half the value of the home for short-terms lasted through much of the 1900's until reforms were created in an attempt to stabilize the market. These methods are impractical or impossible for most Americans in the first two decades of the 21st century because we have failed to comprehend the magnitude of the problem; that is, owning homes has become less attainable for each subsequent generation since the 1980's.

During the years 2020 and 2021, there was a surge in home buying fueled by the reactions of businesses and employees to work remotely and relocate, often to locations with a lower cost of living. Anecdotal accounts of reasons these people decided to move ranged from their location not being as valuable to them due to mandatory quarantines, to moving out a little sooner than planned. The economic outcomes of this shift resulted in a rapid increase of home prices. There simply was not enough supply of new homes to meet the demand of consumers in the market. To further worsen relative affordability for potential homeowners, there was also a lumber shortage and multitude of supply chain issues that intensified the process. It was completely unconventional of extant generational experiences and led to bidding wars and houses selling minutes after they hit the market prior to the buyer ever stepping foot on the property. This had the greatest effect on those just entering the market, namely millennials. It is unclear how this may affect future generations, but an improved housing affordability index may help to mitigate risks to consumers and local economies.

Our hypothesis is that improving upon the housing affordability index will increase understanding among financial institutions and municipalities to empower prospective homebuyers in their areas. With an improved relative affordability measure, efforts to mitigate homelessness, reduce financial stress, increase quality of life, and boost economies can occur at national, state, and local levels. By incorporating a safety net within

the calculations of housing affordability, especially for metropolitan statistical areas, we can focus directly on the consumer’s access to housing and evaluate how comparable it is to that of previous generations.

## 0.2 Literature Review

Housing affordability is a significant contributing factor when determining the level of economic development and market stability in a country.<sup>9</sup> For over a century, policy experts and analysts have targeted economic factors as the most important determinants of how to make housing more affordable.<sup>13</sup> While this is supported by a larger share of historical publications than recent sustainability-focused socioeconomic examples, there has been a rapid increase since 1995 in the volume of literature concerning affordable housing.<sup>17</sup> These papers have a wide range of definitions for what affordable housing is but nevertheless have exposed conventional economic hypotheses as having only scratched the surface of how to accurately calculate housing affordability.<sup>4</sup>

A cross-disciplinary review of existing literature on affordable housing in 2012 noted that it is widely accepted “among policy analysts that an inadequate supply of affordable rental and owner-occupied housing can lead to a number of negative social outcomes.”<sup>15</sup> Current reviews of the literature from peer-reviewed articles to public releases of industry reports and data, found this remains the case.<sup>17</sup> Failure to provide even adequate housing has been shown to create circumstances that force households into lower economic classes and in some cases, poverty. Whereas any marked improvements in how we measure housing affordability, could lead to positive outcomes for society and individuals alike through policy and market-based resolutions. Such documented outcomes include “enhanced housing and transportation infrastructure, income adequacy, household wellbeing, reduced inequalities and improved rental housing” as well as access to new opportunities.<sup>9</sup>

There is also a consensus that we must include additional costs consumers face to fully understand housing affordability.<sup>11</sup> One example of this comes from studies that include household transportation expenses to create another index known as the location affordability index (LAI). The U.S. Department of Housing and Urban Development (HUD) filled a database with information on affordability by location and developed a method to estimate the LAI. While this has caused major improvements in public and private assessments of affordability, its reliability remains in question. Studies have shown that at census-tract level aggregation, LAI models tend to overestimate costs, especially for renters.<sup>11</sup> Due to the unreliability in its aggregation, variable construction, and modeling methods, some studies have claimed “reliability metrics for the Journey to Work (JTW) data on transit use warrant jettisoning these data altogether” citing ethical reasons as well.<sup>10</sup> Here again, there is also disagreement on how to calculate these satellite variables of the overall housing affordability index.<sup>6</sup> Other studies have made it clear that an optimal measurement that precisely estimates the true value of affordability, especially at local levels, remains a major concern.<sup>3,7,9</sup>

As the direct costs of housing have increased on average across the United States, research efforts to address the growing concern of housing affordability have also tried to identify and deescalate the severities of the sources of the problem.<sup>5,8,16</sup> While many have achieved success in identifying criteria thought to cause housing affordability, new challenges have raised alarm over the definition of affordable housing and how results differ depending on how it is measured.<sup>17</sup> These are considered critical success factors.

According to a study published in the journal *Building and Environment* there are 13 critical success factors (CSFs) relevant to ensuring sustainable affordable housing.<sup>1</sup> These CSFs stem from a survey of affordable housing experts who responded to a questionnaire with 30 unique success factors identified from their review of existing literature in 2019. Responses were compiled and researchers were able to rank the experts’ agreement with each other on each unique success factor and found underlying patterns. Among those patterns they found the main obstacle was “political will and commitment to affordable housing” but the second highest consideration was the “formulation of sound housing policies.” As shown through records of housing affordability indices and reviews of the literature, forming sound housing policies is heavily influenced, and some might even say dependent on, the accuracy of the estimates themselves. These results have reaffirmed through various surveying and policy simulation modeling methods in subsequent studies.<sup>9,17</sup>

Experts in housing policy and urban studies tend to calculate housing affordability indices as the “average share of households paying more than 30 percent of their income for shelter.”<sup>6</sup> According to the National Association of Realtors, the Joint Center for Housing Studies at Harvard, and other peer-reviewed sources, this is the most widely accepted metric for estimating housing affordability.<sup>3,14,18</sup> The Joint Center for Housing Studies also states that when household spending exceeds this threshold, it is reasonable to expect that they do not have enough to afford other necessities. This is an important distinction from other sources of existing literature, which tend to have a broader definition such as “the relationship between housing and its users”<sup>3</sup> and those interested in examining the interdependencies of housing and health.<sup>2,19</sup>

Based on studies of global housing markets it is clear that in order to develop a better, more robust metric that is useful to both consumers and policymakers in understanding and addressing housing affordability, it is necessary to build on existing standards and data sources.<sup>1,12,13</sup> There are well-understood relationships between housing affordability and housing cost, transportation, cost of living, and other variables and these should be used to determine an ideal measurement approach.<sup>1,15,19</sup> Despite the exponential growth of research on the measurement and concept of housing affordability, there remains concerns over the reliance of conventional methods to inform and reform weaknesses in policy instruments.<sup>3,4,9</sup> Current literature reiterates the need for better methods of conceptual analysis on and physical estimations of housing affordability to bring new insights to the field in a transparent, less data intensive manner.

### 0.3 Methodology

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
# Read in Data  
df <- read.csv("https://raw.githubusercontent.com/palmorezm/msds/main/698/Data/compiled.csv")  
# Change Data Types to Numeric  
df[4:length(df)] <- sapply(df[4:length(df)], as.numeric)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
df <- df %>%
  dplyr::select(-X)
```

We define affordable housing as the potential of the median household in a given area to finance or purchase a house at or above that area’s average house cost threshold wherein the median household has enough income leftover to afford other commonly purchased expenditures. As mentioned, there are many perspectives regarding what affordable housing means. To cover the topic at a reasonable depth, and to consider multiple affordability strategies, we develop 7 housing affordability indices (HAIs) based on the literature and available data. There is consensus that focus should be on creating metrics that are widely applicable, require relatively little computational power, and build on existing data. Ideally, these metrics would also be transparent for the public and policymakers alike. We adhere to these principles when developing our own adjusted HAI metrics.

Our original data comes from publicly accessible institutions including The United States Bureau of Economic Analysis (BEA) and their partnership projects with the Bureau of Labor Statistics (BLS) along with a mixed set of parameters created by the U.S. Census Bureau. These datasets each contain 1 or more of 15 essential variables that we use to calculate relevant HAI statistics. Our final dataset contains 43 variables with 28 created from separate computations used in the HAI adjustments. This final dataset and its original data sources, merging, and joining processes, are linked in the appendix for reference and open-source reproducibility.

We focus our efforts on localizing the HAI and utilizing relevant information at the level of metropolitan statistical areas (MSAs). As defined by the Census Bureau, an MSA is “one or more counties that contain a city of 50,000 or more inhabitants or contain a Census Bureau-defined urbanized area (UA) and have a total population of at least 100,000 (75,000 in New England).” It also noted that MSAs contain a population nucleus as its core with a “high degree of economic and social integration with that core.” In 2019, approximately 83% of the U.S. population lived in these urbanized metropolitan areas with an average population density of 283 people per square mile. The national average population density was about 94 for the same year. Urbanization and the growth of MSA has steadily increased in population and size (to include more counties) for over a century. Because of this, we can be confident that our methods capture at least 83% of the U.S. population but do not capture the full spread of its area.

These statistics show distinct gaps between where people are concentrated and where people are spread out and result in vastly different challenges to housing affordability outside of metropolitan areas. We note this but do not adjust for the differences in our 7 adjusted HAI estimates due to calculations that incorporate nationally smoothed values. These would not serve micropolitan or other smaller statistical areas. Outside of MSA (where at least 83% of the population resides), the HAI is less flexible.

Mathematically, we can write out these 7 HAI estimates as equations. All methods use a fixed 30-year interest rate of 3.5% which according to major lending services, is currently closest to the market average. In the interest of making our HAI’s widely applicable, transparent, and useful to a broader audience, we include exact specifications of each new version of HAI and build on the national standard.

These methods can be reduced to the following equations:

NAR Method

$$\left( \frac{M_i}{\left( \frac{M_p \times 0.8 \left( \frac{IR}{12} \right)}{1 - \left( \frac{1+IR}{12} \right)^{360}} * 48 \right)} \right) \times 100$$

Where  $M_i$  is the median household income for each MSA,  $M_v$  is the median home value for the same MSA, and  $IR$  is a fixed interest rate. We make a slight adjustment to this to create a real wage using regional price parities for each MSA provided by BEA.

RPP Adjusted Income (Real Wage) HAI

$$\left( \frac{M_i - \left( \left( \frac{R_{all}}{100} \right) \times M_i \right)}{\left( \frac{M_p \times 8 \left( \frac{IR}{12} \right)}{1 - \left( \frac{1+IR}{12} \right)^{360}} * 48 \right)} \right) \times 100$$

With  $R_{all}$  representing the regional price parities for each MSA of the entire basket of goods and services surveyed by the BEA and the BLS. The rest of the equation went unchanged effectively allowing us to estimate the HAI with a ‘real wage,’ that is, the income leftover after adjusting for cost of living per MSA.

Rent Adjusted HAI

$$\left( \frac{M_i - \left( \left( \frac{R_{rent}}{100} \right) \times M_i \right)}{\left( \frac{M_p \times 0.8 \left( \frac{IR}{12} \right)}{1 - \left( \frac{1+IR}{12} \right)^{360}} * 48 \right)} \right) \times 100$$

In this case, we isolate the most integral category of the regional price parities of MSAs, which is rent. These results could be interpreted as part of the route to homeownership since renting a place is often necessary before purchasing with a mortgage or other financial commitment outside of leasing.

IPD Projected HAI

$$\left( \frac{\left( \frac{I_{rw} + I_{rt}}{2} \right) \times \frac{IPD}{100}}{\left( \frac{M_p \times 0.8 \left( \frac{IR}{12} \right)}{1 - \left( \frac{1+IR}{12} \right)^{360}} * 48 \right)} \right) \times 100$$

The Implicit Regional Price Deflator (IPD) for MSAs is a “regional price index derived as the product of two terms: the regional price parity and the U.S. PCE price index” in which “growth rate or year-to-year change in the IRPDs is a measure of regional inflation.” We use this growth rate to project where the average between real personal income ( $I_{rw}$ ) and rent adjusted incomes ( $I_{rt}$ ) would be and create an HAI as if housing prices were stagnant.

Raw HAI

$$\left( \frac{M_i}{\left( \frac{M_p \times 0.99 \left( \frac{IR}{12} \right)}{1 - \left( \frac{1+IR}{12} \right)^{360}} * 48 \right)} \right) \times 100$$

This Raw HAI considers how affordable housing would be if the individual were able to have 99% of the value of the home financed at a fixed interest rate for 30 years.

Outstanding Debts HAI

$$\left( \frac{\left( \frac{I_{rw} + I_{rt}}{2} \right) - \sum (D_{mv}/h), (D_{ed} * h), (D_{il} * h), (D_{cc} * h), (D_{oc} * h)}{\left( \frac{M_p \times 0.99 \left( \frac{IR}{12} \right)}{1 - \left( \frac{1+IR}{12} \right)^{360}} \times \frac{100}{30} \times 12 \right)} \right) \times 100$$

For the outstanding debts HAI only consumer debt is considered. No mortgages or home equity loans are included in consumer debt. These debts are averages for the American individual and as such are adjusted to fit household dollars. There are 5 kinds of debt in our formula, each of which are representative of the median for their category. They include motor vehicle debt at roughly \$20,000 per household, education debt (which includes student loans) at \$9,664 for each individual in a household under 40 years of age, debts from installment loans at \$9,609 per individual on average, credit card debt at \$3,500 per individual, and all other debts from medical expenses, loans against pensions, life insurance, payday loans, and more lumped into one category at about \$10,000 per individual. The average household size (h) is 2.53 based on the 2020 Census Bureau’s decennial census. We expand the loan allowance here increasing the lending leniency to 30% of monthly personal income for the average household to compensate for the additional debts and give as much tolerance as possible with current lending practices.

## Lenient Lending HAI

$$\left( \frac{M_i}{\left( \frac{M_p \times 0.97 \left( \frac{IR}{12} \right)}{1 - \left( \frac{1+IR}{12} \right)^{360}} \times \left( \frac{100}{60} \right) \times 12 \right)} \right) \times 100$$

To stretch lending practices even further, a concept which has been argued would improve housing affordability while increasing market instability, we adjust the median value of the mortgage loan to account for 97% of the value of the home. This indicates that the individual would only need to pay 3% of the home value for a down payment. Then, we reevaluate the qualifying income of the HAI to allow greater flexibility in how much the average consumer could pay to 60% of their monthly income being spent on their monthly mortgage payment.

Within each of these HAIs we have a particular purpose and tactic to better calculate housing affordability. In all cases, our interest rate is held constant to attribute changes to the formula rather than state and federal loan rates. The NAR method serves as our control, and it follows the same formula as the national standard setup by National Association of Realtors. Since the housing data collected contains estimated valuations by homeowners (which is not an exact estimate of home values), we need a base sample to build upon and measure deviations from. Unfortunately, the national association of realtors does not allow data sharing or open-source use without randomization of the MSA from their respective HAI's. This is not valuable to us since this would remove the opportunity for the public and policymakers to identify their locality and thereby eliminate the potential for improvements.

The remaining HAI methods rely on adjustments to median household income with the same qualifying income as the NAR HAI except when calculating the Outstanding Debts HAI and Lenient Lending HAI. The first adjustment simply creates a real wage estimate for the locality by adjusting for the cost of living. This integrates a median dollar adjustment derived from the national average for a market basket of goods and services including prices for transportation, food, rent, and other categories. We isolate rent in the Rent Adjusted HAI and adjust the median household income proportionally for each MSA. Next, we use the implicit regional price deflator for each MSA to create an IPD Projected HAI that incorporates wage growth with a theoretically stabilized housing market. Our Raw HAI weighs the influence of down payments on affordability by increasing the expected loan amount to 99% of the median home value. Lastly, we identify the average household debts from the literature to form an Outstanding Debts HAI and form a new Lenient Lending HAI by hypothetically easing lending restrictions.

Importantly, the data used to form these HAI estimates are collected and updated at regular frequencies and averaged (by median) across 5-year intervals. This limits volatility within the data, making it less biased and more robust. Our study spans from 2010 through 2019, or one decade. Traditionally, the 'cap' of household income that could be reasonably spent on housing was 30% of household income. We lower this in all but two of our calculations where it makes intuitive sense to reflect present lending practices. We also assume that every household is capable of making a 20% down payment. We note that this is highly unlikely and is why government programs exist to assist individuals with lowering the up-front costs of purchasing a home. Nevertheless, because it is a standard of the industry and lending institutions, we focus on changes to income that could affect housing affordability.

To elucidate upon the effects of each HAI method we model their distributions across income and population by MSA. The relationship between the population and our newly formed median HAI affordability model is such that we can state half of the population is below the HAI value and half is above. We take advantage of this relationship and available literature to estimate the proportion of the population that is unable to afford housing. We also compare rates of change in personal and real wage income to the change in housing affordability at various HAI levels. This allows us to use linear regression techniques to model changes by HAI method and predict where their metrics are heading as well as approximate how many people could afford at house in the future.

```
# All HAI Estimates
df.fin <- df %>%
  mutate(IR = 0.035,
         PMT = MEDVAL * 0.8 * (IR / 12) / (1 - (1 / (1 + IR / 12))^360)),
```

```

    QINC = PMT * 4 * 12,
    HAI = (MEDINC / QINC) * 100) %>%
mutate(ADJALL = (MEDINC - ((RPPALL / 100)* MEDINC)),
    AINCALL = MEDINC + ADJALL,
    # Real Wage HAI:
    HAIRW = (AINCALL / QINC)*100) %>%
mutate(ADJRNT = (MEDINC - ((RPPRENT / 100)* MEDINC)),
    AINCRNT = MEDINC + ADJRNT,
    # Rent Adjusted HAI:
    HAIRNT = (AINCRNT / QINC)*100) %>%
mutate(ADJIPD = (AINCRNT + AINCALL / 2)*(IPD/100),
    # IPD Projected HAI:
    HAIIPD = (ADJIPD / QINC)*100) %>%
mutate(PMTRAW = MEDVAL * 0.99 * (IR / 12)/(1 - (1/(1 + IR/12)^360)),
    QINCRAW = PMTRAW * 4 * 12,
    # Raw HAI at 20% Down Payment:
    HAIRAW = (MEDINC / QINCRAW) * 100) %>%
mutate(HHSIZE = 2.53,
    DEBTMV = (20000/HHSIZE),
    DEBTED = 9664*HHSIZE,
    DEBTIL = 9609*HHSIZE,
    DEBTCC = 3500*HHSIZE,
    DEBTOC = 10000*HHSIZE,
    DEBTS = DEBTMV + DEBTED + DEBTIL + DEBTCC + DEBTOC,
    AQINC30 = PMT * (100/30) * 12,
    AINCDBT = ((AINCRNT + AINCALL / 2) - DEBTS),
    # HAI Adjusted for Average American Household Debts
    HAIDBT = AINCDBT / AQINC30) %>%
mutate(PMT3DP = MEDVAL * 0.97 * (IR / 12)/(1 - (1/(1 + IR/12)^360)),
    AQINC60 = PMT * (100/60)*12,
    # Lenient Lending Practices HAI
    #(60% of monthly income on mortgage is acceptable with 3% DP)
    HAILEN = (AINCALL / AQINC60)*100)

```

## 0.4