Does affordable housing make cities more competitive?

by Remert Erasmus

Thesis presented in fulfilment of the requirements for the degree of Master in Commerce in the Faculty of Economic and Business Science at Stellenbosch University

Supervisor: Dr. Dawid van Lill

December 2021

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2021

Abstract:

This study sheds light on the consequences of unaffordable house prices and how they negatively affect cities' growth prospects. This is done by studying how economic expansion indicators in a city such as income growth and population growth are affected by increases in housing affordability ratios. This study uses a novel incorporation of the house price to replacement cost ratio that indicates high house prices due to lack of supply. By studying how increases in this ratio affect cities' population and income growth, one can observe the economic costs of restrictive zoning laws. The model uses an amenities spatial equilibrium model as a point of departure and the empirical results indicate that a 10% increase in the house price to replacement cost ratio is associated with a 14% decrease in the population growth rate and a 47% decrease in the income growth rate over 5 years when controlling for the capital gain from rising house prices.

Opsomming:

Hierdie studie werp lig op die gevolge van onbekostigbare huispryse en hoe dit die groeivooruitsigte van stede negatief beïnvloed. Dit word gedoen deur te kyk na hoe aanwysers van ekonomiese uitbreiding in 'n stad, soos inkomstegroei en bevolkingsgroei, beïnvloed word deur toenames in bekostigbaarheidsverhoudings vir behuising. Hierdie studie maak gebruik van 'n nuwe huisprys-teenoor-vervangingskosteverhouding wat dui op hoë huispryse as gevolg van ‘n gebrek aan aanbod. Deur te bestudeer hoe toenames in hierdie verhouding die bevolking en inkomstegroei in stede beïnvloed, kan meer lig gewerp word op die ekonomiese koste verbonde aan beperkende soneringswette. Die model gebruik 'n geriefs-geografiese-ekwilibriumsmodel as vertrekpunt en die empiriese resultate dui aan dat 'n toename van 10% in die huisprys-teenoor-vervangingskosteverhouding gepaard gaan met 'n afname van 14% in die bevolkingsgroeikoers en 'n afname van 47% in die inkomstegroeikoers oor ‘n 5 jaar periode as die kapitaalgroei van stygende huispryse in ag geneem word.

Table of Contents:

Table of Figures:

Table of tables:

Introduction:

The affordability of housing has always been a priority for households. Still, there has been an increased concern in recent years due to house prices and rents outpacing inflation in developed economies. Concern by households regarding the affordability of housing is justified as housing expenditure in the United States makes out 40.4% of the official U.S. Bureau of Labor Statistics CPI basket with 31.6% going to shelter itself which includes rental costs and the costs associated with owning real estate and 8.8% going to other expenses such as utilities and furniture (Bureau of Labor Statistics, 2019). This share is even larger for poorer families, and when house and rental prices increase above inflation, the impact is also felt by the average household.

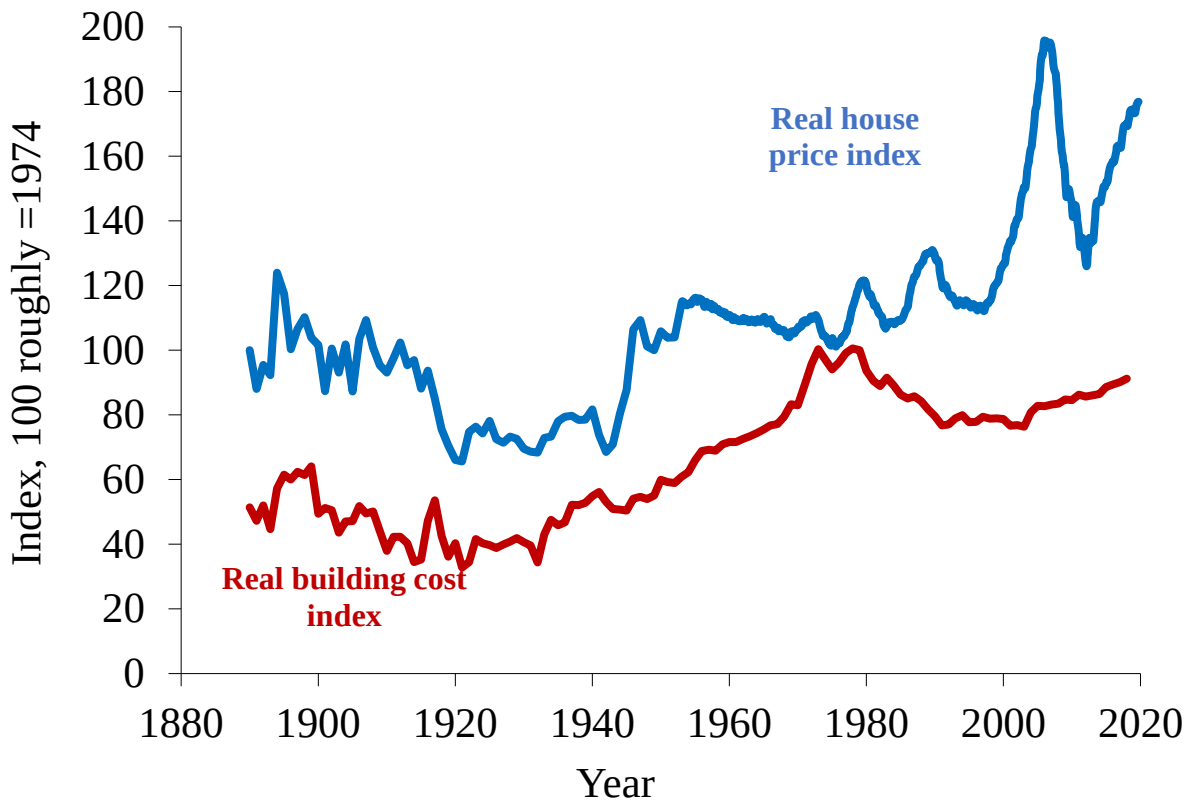
Although the public and the average household are acutely aware of the decline in housing affordability over the past decades, this concern has not been met with decisive action by policymakers and politicians. Some political parties have made it an issue, but this has, for the most part, been a marginal priority that only garners attention after short-term political goals have been met. This is surprising as one would expect a wide-ranging issue such as housing affordability to be a priority. There are two reasons why this issue has not received more attention. The first is that the increase in real house prices and the corresponding decline in affordability are only drawbacks if one does not own real estate. Quigley and Raphael (2004) note that for the 65% of households in the United States that own the house they live in, this rise in their house's real price has substantially lowered the cost of ownership due to the capital gains they have made. However, renters and first time-buyers are faced with the prospect of being excluded from owning a house when affordability levels decrease.

The second reason why housing affordability is not prioritised follows from the difficulties and complexities faced by policymakers and politicians when they want to diagnose the causes of unaffordable housing levels. Policymakers can only practise the appropriate remedies for unaffordable housing once a sound diagnosis of the causes are made. Policymakers have tried to some degree to make progress regarding a diagnosis with the Millennial Housing Commission (2002) in the U.S. and the Redfern Review (2016) in the United Kingdom but have failed to translate any of the findings into substantial policy implementation. To some degree, this is not surprising as economists too do not agree on the diagnosis and the appropriate remedies. This is because the factors at work are varied and can be influenced by the ability of households to borrow, changes in the supply of housing coming to market, changes in housing preferences and changes in consumption patterns of households and changes in zoning laws which have become prolific over the past decades but differ substantially by location.

Academic researchers have started to study the nature and consequences of unaffordable housing only very recently. Although economic geography models have been attempted since the 1970's, research looking at the nature of unaffordable housing levels has only garnered attention post-2000 in the literature. The purpose of this study is to go one step further by looking at the consequences of unaffordable housing. The literature up to this point has developed the formal models for why house prices grow and why cities grow and have incorporated housing costs into the models but have not yet studied the consequences of unaffordable housing levels that follow from an inelastic supply curve for housing.

Where studies in the past have studied how urban growth influence house prices, this study will ask this question in reverse to understand how unaffordable house prices influence the growth of cities. Put another way; does unaffordable housing make cities less competitive? This study will use the amenity model proposed by Roback (1982) as the point of departure and then use two indicators of population and income growth as dependent variables in the empirical study as indicators of a city's growth. The study will depart from other studies by incorporating a novel indicator for unaffordable housing due to restrictive zoning laws which we will call the house-price-to- replacement-cost-ratio. Figure 1 puts the real house price and real building cost into perspective, and by looking at figure 1, one can observe that the real house price has outpaced the cost of construction. Glaeser and Gyourko (2018) argue that this disconnect is due to restrictive zoning laws and that it hampers cities' growth because it's too unaffordable for prospective workers to move there.

Figure 1:Case-Shiller house price versus construction cost.



(Source: The Case-Shiller house price index and the C.S. building cost index).

Using the house price to replacement cost ratio as an explanatory variable and other control variables, we attempt to show that unaffordable housing has a dampening effect on both the income growth and population growth of a city. This idea is finally extended to show that unaffordable housing also serves to dampen a city's capacity to sustain a given economic boom. The theoretical and empirical part of the study will be preceded by a section that explains housing affordability as a concept and why it is important. This will be followed by a section devoted to a review of the literature up to this point.

Lastly, it is important to note that this study will focus on the United States as the geography being studied and not South Africa. The reasons for this are several. Firstly, the United States delivers the highest amounts of urban

variables and largest collection of cities which will make the empirical research more robust. Secondly, the data for the United States has longer time periods (census data for geographies for the original 13 colonies go back to 1790). Finally, although South Africa exhibits differences in housing affordability across geographic location and even different economic compositions of cities themselves, a study of this nature would probably deliver a higher pay-off if one could study cities individually seeing that socio-economic differences inside South African cities can be larger than differences between entire U.S. states. This will however require a novel collection of data which is beyond the scope of this study.

Why housing affordability is important:

Housing affordability as a technical term has over the years been unnecessarily vague, both as a term used in policy planning and in academia. Unlike other terms such as GDP, CPI or the GINI coefficient, housing affordability as a technical term fails to fall back onto its own fixed definition. This can be attributed to the history and development of urban economics and economics of housing as fields of study. Given the important role of housing in a society, it is easy to understand why it has been studied in multiple fields ranging from macroeconomics and finance all the way to sociology and history. Before we define housing affordability for the purposes of this study one will first need to understand the nature of the term in different contexts.

The need to understand the role of housing affordability follows from the role housing itself plays in society. Housing as a standalone concept plays a role in the primary sense of delivering shelter, protecting one's health, and having a place to stay during old age. This is the most basic understanding of housing, and this conceptualisation applies to all societies everywhere. In the secondary sense, housing can be conceptualized in the sense of real estate as an asset class which can also deliver a rental income. This conceptualisation also emphasizes the importance of owning land and how the scarcity of land has major implications on price.

The third conceptualisation of housing incorporates the concept of housing at the societal level. Here, unlike in the first two cases where one unit of housing has implications on the welfare of one household, housing on the societal level has implications on the societal level. It is especially on the societal level where housing affordability as a concept has broader implications and the need for it to be adequately understood and defined arises. On this tertiary level of conceptualisation, the benefits and costs of housing changes when the affordability of housing changes on the societal level. Here, changes will be felt in household finances, cost of aged care, demographic changes, community stability, gentrification, labour markets, crime levels and economic development. All these factors will change according to changes in the benefits and costs of housing, and this balance in turn, will be determined by housing affordability. From this one can see that housing affordability as understood on the societal level has the following ramifications.

Lower growth due to labour market misallocation:

The economic implications of housing affordability are numerous. If either the cost of renting or owning property in a city is too expensive when a job opportunity in a city arises then the high housing costs will serve to constrain the labour market of that city. The ramifications of this will be two-fold. The first leading to a lower population in the expensive city and secondly the result of per capita incomes in that city being suboptimal if one assumes that the marginal contribution to productivity of an educated person would be better absorbed in a highly productive city compared to a city with lower productivity per capita. To put it more simply, total national productivity would be higher if all the software engineers completing their studies in Oklahoma could easily move to high productivity cities such as San Francisco or New York with a high absorption capability instead of being retained in Oklahoma where their absorption in the labour market will be suboptimal and therefore their productivity as well. This is an argument put forward by (Hsieh and Moretti, 2019), where they showed with the use of a spatial equilibrium model that aggregate growth in the United States was 50% lower over the 1964-2009 period due to this form of spatial misallocation.

Household spending:

Expenditure on housing is normally the largest single category of expenditure in most households. This includes the spending on mortgages, rents, property taxes and maintenance and administration expenses. In the United States this amounts to 31.6% of the expenditure of the average household budget. If this percentage becomes too large, then it will affect the rest of the household expenditure negatively. Over the years this percentage has however remained relatively stable. Another, more specific effect of housing expenditure is the debt component associated with mortgages where a rise in interest rates can have significant ramifications for that part of household expenditure that excludes housing expenditure. This is especially relevant in the post-covid-19 interest rate environment where a major component of the current mortgages leads to a high household debt to income component, making households susceptible to an increase in interest rates. This also implies that a 100-basis point increase in interest rates on a variable mortgage issued in 1990 will have a lower effect on household expenditure than a 100-basis point increase on a mortgage issued in 2020 due to the household debt to income ratio being lower in 1990. For this reason, policymakers use debt servicing costs as one of the major indicators of housing costs, as a large part of the household budget can be affected by this (Bank of England. Quarterly Bulletin, 2014: 419).

Cost-push inflation:

The role housing affordability plays in the consumer price basket is also important and often overlooked. As mentioned previously the cost associated with housing, whether it is owning a house or renting a house, is in most countries the largest single expense in the household budget and makes out 40.4% of the official U.S. Bureau of Labor Statistics CPI basket with 31.6% going to shelter itself which includes rental costs and the costs associated with owning real estate and 8.8% going to other expenses such as utilities and furniture (Bureau of Labor Statistics, 2019). The size of housing expenditure is large enough to warrant a separate price index from the standard CPI.

When one excludes the official *CPI-U's shelter component: All items*, one gets an inflation rate of 1.3% over the 2010-2020 period compared to an inflation rate of standard CPI of 1.7% for the same period (Emmons, 2020). This difference can have major implications for monetary policy where countries have inflation targets of say 2%. This housing component can in effect bias the CPI and lead to premature rises in interest rates, especially if one considers that the largest single sub-item of the housing components is owners' equivalent rent of residence which is 19.6% of CPI itself. Stroebel and Vavra (2019) showed that when house prices rise in zip codes that have higher homeownership rates, retail stores will increase their mark-ups at a higher rate than zip codes with a higher share of renters. This is because homeowners seem to exhibit a reduction in demand elasticity when the property they own increases in value (Stroebel and Vavra, 2019).

The relationship mentioned above between real estate prices and CPI pertains to the CPI basket on a national level. When one considers the city-level CPI, and the role local real estate prices play on the local/city level CPI, there must be different considerations. Here one needs to borrow from the Penn effect, which observes that consumer price levels can differ drastically according to locations. This was later expanded upon by the Balassa-Samuelson effect, which better explained how productivity increases in domestic tradeable goods have a spill-over effect on non-tradable goods and especially the price of services, leading to higher consumer price levels overall. The contrast in prices was especially sharp between high and low productivity countries where a service such as a haircut that is non-tradeable differs sharply between countries (Asea and Corden, 1994).

The natural question to ask then is, if a non-tradeable good such as real estate is such a large part of the consumer basket, does it have any effect on the input costs of exported goods and services in a city and does this in turn affect the competitiveness of that city? In other words, do real estate prices deliver some form of cost-push inflation on a city level that makes the prices of tradeable goods and services of that city less competitive to neighbouring cities in real terms? The apparent critique here is that real estate prices are by definition high due to the high productivity of the city, leading to high household income which drives real estate prices higher. This would be true if one ignores the role supply plays in keeping real estate prices balanced on a city level. This has been documented by Glaeser and Gyourko (2018) which showed how cities such as San Francisco which has a low issuance of building permits relative to its housing stock serves as a driving force behind house price growth.

City-level population and income growth:

If it is the case that high real estate prices have a substantial influence on cost-push inflation, then one would expect the growth of an expensive city to fall below its optimal growth. The first form of growth that would be sub-optimal is the population growth rate. This would be due to real estate prices and rent prices being prohibitively expensive for prospective workers to move into the city, which will have a dampening effect on the population growth rate itself. The second sub-optimal growth would be in the form of household income growth rates. Growth in household income can be more nuanced due to two factors at work. The first being the labour matching component where high productivity vacancies match highly qualified workers in a high productivity city. If the labour matching happens at a suboptimal rate, then this will lead to suboptimal household income

growth. This first factor of labour matching affecting household income is thus closely related to the suboptimal population growth rate.

The second factor leading to suboptimal household income growth follows from the role cost-push inflation plays on a city -level economy. Suppose the cost-push inflation in a city is driven by high real estate or rent prices, which in turn follows from a lack of real estate supply. In that case, this will increase commercial rents as well, which will increase the cost of locally manufactured goods and locally delivered services. This will result in a decrease in competitiveness relative to other cities, both in the prices of tradeable goods and tradeable services.

There is also an added factor where companies looking to expand to other cities need to take possible profits into account when workers have to be paid a higher wage in cities where housing is less affordable. Keeping all other factors equal, companies will choose to invest in more affordable cities if they could. In other words, it is possible that housing affordability can determine investment choices of companies. The problem is that those cities with equal amounts of human capital, but different housing affordability levels are usually also biased in terms of regulation. Austin, Texas and Fremont, California might have equal amounts of human capital and Austin might have more affordable housing but how would one know that companies are moving to Austin, Texas for housing affordability instead of the zero-percentage state-level income tax already present in Texas. The overlap between affordable housing and low regulations on state-level makes it more difficult to understand the role of housing affordability for companies' investment decisions and how this will ultimately translate into income growth.

The house as an asset:

Housing affordability has a major role in household wealth where the real estate a household owns is regarded as an asset and fluctuation in the value of this asset's price has itself an effect on other variables. If affordability issues for households prevent access to owning real estate, then it follows that changes in the value of real estate will have little secondary effects in the economy if a large share of households rent their property. One must also note that prices in the residential rental market in the United States are considerably more stable than house prices itself and although rental prices are partly a function of house prices the house price-to-rent ratio can go through large fluctuations (Garriga, Manuelli, Peralta-Alva, 2019)

The secondary effects mentioned here are several. The first has to do with homeowners having access to home equity. According to Agarwal and Qian (2017) lower rates of access to home equity has a negative effect on consumption smoothing in times of household income needs. In effect, access to home equity serves as some form of savings at the household level. This allows households to use that savings during financial need, which allows for some form of self-insurance mechanism on a household level (Agarwal and Qian, 2017). It is important to note that the data used in this study was Singapore-based. Singapore has some of the world's highest home-ownership rates due to an aggressive government initiative to deliver a property-owning society. This study's conclusion is important, but one must keep in mind that in many cities in the United States, the homeownership rate is 40% lower than in Singapore, which may have a dampening effect on the self-insurance mechanism on household level.

The first effect mentioned above relates to home equity in the case of financial need of households. The role of home equity when the value of real estate prices rise has multiple effects on the economy. Lydon and O'Hanlon (2011) studied the role house price growth had on the Irish economy during the pre-2008 housing bubble. They found that rises in home equity followed from rises in house prices and that this leads to an increase in home loans itself, to the extent that at the hight of the property bubble a third of all home loans issued by banks in Ireland were top-up loans where homeowners used the market-related growth in home equity to take out further loans. Those additional loans also had a specific effect on consumption patterns where they found that 90% of the top-up loans were used to spend on property where 66% went towards spending on maintenance and upgrading the same property, and the other 24% was spent on purchasing investment properties. According to Lydon and O'Hanlon (2011), a large part of the rise in house prices is used to grow the already high construction output of the Irish economy and the consumption increase that follows from it. They also found that this rise in home equity leads to a substantial increase in the consumption of durables, such as furniture and vehicles. One significant finding of the study relates to the role that home equity plays in business formation. Lydon and O'Hanlon found that those households that were self-employed used larger shares of their home equity than wage-earning households. Lydon and O'Hanlon (2011) claim that this indicates the important role that home equity plays in increasing business investment and business formation.

Zhou and Carroll (2012) added to the home equity concept when they studied the effect of real estate price increases on the United States' consumption dynamics. According to them, increases in home equity has a role to play in consumption increases. They observed that a dollar increase in real estate prices growth was associated with a 6-cent increase in consumption. This wealth effect was only measured after a 2-year lag of the real estate price increase. Still, this observation is important when one considers that they did not find the same wealth effect when it came to increases in the stock market and pension fund wealth of households. This might indicate there are added wealth effects at work when owning real estate on the household level compared to owning stocks or other financial assets. Zhou and Carroll (2012) argue that this difference in consumption outcomes has different monetary policy implications. Even though both real estate and stock wealth are both household assets, changes in their respective values have different strengths when it comes to real economic impact. This is why Zhou and Carroll (2012) observed that consumption did not significantly fall after the burst of the dotcom bubble but were significantly affected after the housing bubble crash of 2008, meaning that monetary policy makers have to take into account the potency of housing wealth destruction versus that of stock wealth destruction.

Social and neighbourhood planning:

As previously explained housing is one of the largest assets of households in the United States. If it is the case that income and wealth inequality is present, then it follows that real estate prices and levels of affordable housing will mimic these differences in income and wealth as well. This then has additional effects on community composition, crime levels, social issues, and community sustainability. When there are large changes in the affordable housing, this can negatively affect the levels mentioned above. Katz *et al.* (2003) looked at these issues and studied how programmes can be used to increase the level of affordable housing on a local level and what results were delivered when certain housing strategies were effectively implemented.

Katz *et al.* (2003) concluded that one of the most effective ways to allow for affordable housing was by increasing the private economy's capacity to increase supply. Regulations restricting the type of developments to certain minimum lot sizes or restrictions on future population growth rates by zone delivers communities that are normally middle to upper-income families. This leads to property developers only delivering real estate developments that will suit those families in order to maximize profits. When this natural progression towards certain kinds of developments is altered by including an affordable housing requirement, one can directly increase affordable housing supply per development. In other words, there should be an emphasis on targeting a form of inclusionary zoning instead of inadvertently delivering exclusionary zoning (Katz *et al.*, 2003). This also has added social benefits where low-income families are allowed to absorb the social benefits of middle-income neighbourhoods instead of being excluded to a socially and economically marginalized zone due to purpose built affordable housing. One example of this, according to Katz *et al.* (2003) is the Chicago Hope VI project which was a voucher system that allowed low-income families to rent in middle-class neighbourhoods. They found that families that used the vouchers moved to middle-class suburbs and benefitted from the better schools, family stability and more jobs opportunities present in the suburbs. This delivered lower rates of households where children dropped out of school, and a higher percentage of family members had jobs compared to those families not benefitting from the voucher programme in the low-income neighbourhoods.

Defining housing affordability:

The literature studying housing affordability has in the past attempted to define housing affordability but have never arrived at a standardized definition. This can probably be attributed to definitions for housing affordability changing according to the goals of a specific study. In the literature it is however clear that the traditional approach is to define housing affordability based on the ability to pay. Here the house price-to-income ratio comes to mind as the traditional benchmark. This ability to pay approach is also the method used in the Housing Affordability Index compiled by the National Association of Realtors which constructs the index assuming that 25% of the median household income in the United States is used to pay the interest and principal of 80% of the house price value. If the index is higher than a 100, say 120, then the median household would have 20% more than the required income to purchase the median house price.

The Housing Affordability Index gives valuable insights into the general capacity of households across the United States to afford a mortgage. According to the San Francisco Federal Reserve (2003) this index has seen large changes from 150 in 1973 and then declining to 65 in 1982 and then again steadily increasing to 135 in 1994 where it has been maintaining a range of 125-145 ever since. Here it is important to note that although housing affordability has increased substantially since 1982, the index is only focussed on the median affordability of housing. The two obvious shortcomings of the index are that it says nothing about how important the role of lower real interest rates since the 80's is in improving housing affordability. Secondly, the index says nothing about those cities on the extremes of the affordability spectrum such as San Francisco and New York where the introduction of stricter zoning laws have influenced house prices. A third less obvious failure of the index is its ability to explain why net household wealth has substantially decreased for the bracket of 35-44 year-olds in 2013 versus the same group in the year 1983 (Glaeser and Gyourko 2018). In 1983 net household wealth for the 50th

percentile of homeowners for the 35–44 year-old bracket in the United States was $55 900 (2013 dollars). This figure has decreased to $6000 for the same group in 2013. Although housing might seem affordable as a share of income it seems that the picture changes when looking at net wealth and that affordability is more nuanced.

Definitions for housing affordability have largely used a demand side definition such as the house price-to-income ratio. Glaeser and Gyourko (2018) suggests a departure from this definition after they noticed that there was a 0.5 correlation between the Wharton Residential Land Use Restrictiveness Index and house price in 2013. The two authors thus suggest a supply side definition of the house price-to-replacement cost ratio as an indicator of housing affordability. According to the authors and for the purposes of *this study* housing affordability can be defined as the gap between house prices as determined by the market (including the regulatory tax due to zoning laws) and the new construction cost of that same house (assuming a standardized price for land). Again, one must emphasise that this is a departure from the traditional ability to pay definitions mainly because the way of looking at the problem of affordability has it roots in the idea that a well-functioning market can match high demand with a high supply.

The natural question to ask is, what exact ratio of house price-to replacement cost can be defined as 'unaffordable housing'? Glaeser and Gyourko (2003) identifies three segments of the U.S. housing market according to this ratio. The first has a ratio of 0.75 and lower, indicating that the market price for houses are substantially below their replacement cost. These are areas in the Rust Belt such as Detroit that have seen little to no economic growth for decades. The second segment can be defined as those cities where market prices are close to replacement costs and are thus close to 1. This is also the segment where the vast majority of cities in the United States find themselves. The third segment has a ratio substantially larger than 1 and these are cities such as Boston and New York on the East coast and Los Angeles, San Jose and San Francisco on the West coast. A house price-to-replacement cost ratio of 1.4 and larger was used in the Glaeser and Gyourko study in 2003 as an indicator of unaffordable housing and will also serve as the benchmark for the purposes of *this study*. In this study the ratio of 1.4 is used for two reasons, firstly it does make sense that if one assumes that most functioning markets will have a ratio somewhere between 0.9-1.1 and that the maximum premium for amenities paid in a location is not more than 30% of the house price, then one will arrive at a theoretical maximum of roughly 1.4. Secondly, 1.4 is used as a benchmark for unaffordable housing for the sake of academic consistency as this ratio is also used in previous studies.

Lastly one must note that although this study departs from other studies by not linking affordability with the ability to pay approach. This study also considers affordability as the level of affordable housing that is socially maximizing on the national level. To put it another way, although citizens of San Francisco might have the ability to pay for houses in their own city, prospective graduates from Detroit wishing to move to San Francisco might find it 'unaffordable'-according to the definition of this study. This is also inline with Hsieh and Moretti's (2019) conception of the role housing affordability plays in maximising social benefits on a national level.

Literature review:

The literature analysing affordable housing's role in cities' economic development, and the role it plays in making cities competitive is still in its infancy (Quigley and Raphael, 2004). The majority of the literature is focussed on the impact of 'sponsored' affordable housing projects being implemented and the local economic advantages that will follow from it. This analysis method is supported by surveys that are typically led by firms, government agencies, and organizations that already have some form of involvement in housing developments itself. Although there is a data component to this form of study it is clear that there still is a lack of formal economic analysis where there is an emphasis on a multitude of data points and economic variables in order to arrive at a conclusion that is econometrically sound and accurate. This study tries to shine more light on this component of the literature which seems absent.

Note that 'affordable housing' in this literature review refers to it in the sense of either government sponsored affordable housing programmes or in the traditional definition of a low house price-to-income ratio. This is slightly different to the definition of this study largely due to this study using a novel approach.

The literature itself focusses on three factors for the role of affordable housing on local economic development. The first relates to the economic activity associated with the construction of affordable housing programmes and the increase in spending on consumption by the new households. The literature also looks at the improvements to the local fiscus of municipalities with the increase in property revenue and stamp duties on real estate and the added revenue of utility companies. Lastly, there is a focus on how affordability levels serve to attract workers to a city and increase labour matching in the local economy. The literature also explores how high levels of unaffordability can indeed lead to migration away from certain cities.

Survey literature:

Most of the studies done in this regard focus on the economic benefits that follow from the construction of new affordable housing programmes. The National Association of Home Builders (NAHB) studied the effect of constructing 100 units of subsidized housing with the Low-Income Housing Tax-Credit (LIHTC). This study focused on the effects of housing subsidies itself and their effect on a national level as this is a widely implemented federal scheme. The authors concluded that for every 100 units of affordable housing constructed using this scheme one can expect an additional 80 jobs created in construction sector directly and a further 42 jobs added from spending in the economy that follow from the construction workers' wages. There is also the added benefits of 30 jobs created from the ongoing spending that follow from the new households in the neighbourhood (Wardrip, Williams and Hague, 2011).

The Minnesota Housing Finance Agency also studied the added benefits of new affordable housing construction in 2009, which looked at the multiplier effect of additional housing spending in the form of housing construction, improvements, and rental assistance. The authors concluded that the additional $260 million spent on affordable housing resulted in a further $471 million spent by the public and private sector over a two-year period (Wardrip, Williams and Hague, 2011). According to this study, $0.91 of induced spending followed from the first round of

spending $1 on affordable housing. The authors also concluded that economic activity that resulted from the initial $260 million spent on affordable housing led to the creation of 10 700 jobs over a 2-year period. A similar study was conducted in Pennsylvania that looked at the additional economic activity that follows from affordable housing programmes. Unlike the Minnesota study, this study was purely focused on the economic activity that follows from remodelling affordable housing and not new construction. The authors found that for every 1 dollar spent on remodelling, there was an additional $1.28 of induced and indirect spending in the local economy and that for every $1 million dollars spent in this programme an additional 14-20 jobs could be expected (Econsult, 2009).

There also needs to be a discussion on the fiscal effects of affordable housing developments in the literature. Given that subsidies and new construction developments have second-round effects in stimulating local economic activity, one would expect that this activity will also lead to more revenues for local governments on a local and state level. According to NAHB (2010) the construction of 100 LIHTC home generates $827 000 in revenues for local authorities, mostly in the form of permit fees and utility user fees which is more than half of the revenues. The rest consists of smaller portions of business taxes, general sales taxes and other smaller fees and taxes

The literature regarding the role that affordable housing levels play in the labour market focuses on issues regarding workers' decisions, decisions made by employers after companies are already established and decisions made by companies before they employ new workers. In a survey lead by the Urban Land Institute in 2007 on a national level in the US, 1200 workers were asked if they would move closer to work if housing was more affordable. Of those 1200 workers, 57% said they would move, and that number increased to 67% for workers with a household income of less than $50 000 per year, indicating that lower-income families absorb most of the cost regarding the structural inequality that follows from unaffordable housing levels. In contrast to the workers' survey, the same survey asked 300 companies with more than 100 employees if they think affordability levels have an impact on retaining workers. Of those 300 companies, 55% believed that housing affordability was a problem for their employees. Of that 55%, two-thirds thought that this had a negative impact on retaining workers on the entry to mid-level work force (Urban Land Institute, 2007).

The evidence also points to the same results in surveys done for specific cities. McCall (1999) found that when surveying large New York companies that were fast-growing, 64% of the companies believed that unaffordable housing levels were hampering employees' recruitment and retention. A similar result was found by Florida University when they asked members of the Miami Chamber of Commerce if they think unaffordable levels of housing had a negative impact on retaining employees, of which 52% agreed (Wardrip, Williams and Hague, 2011).

As mentioned previously, one of the factors also studied in the literature also emphasises the role housing affordability levels play in companies' decisions when expanding their employee base. Gambale (2009) studied this in a survey completed by the company executives involved in deciding on new workplaces' locations. Housing affordability seemed to play an important role when it came to executives selecting locations. The study focussed on the quality-of-life factors that are important to executives when making these decisions and asked them to rate out of hundred the relative importance of each factor. Housing affordability was rated third behind crime rate and

healthcare facilities at 62% level of importance but was still considered more important than public schools, climate and recreational activities by executives (Gambale, 2009). Even though this information is only in survey form, and one cannot deduce the real amount of jobs created in affordable cities and the amount lost in unaffordable cities, it still lends weight to the direction of jobs movements and affordable housing.

There is also the factor of out-migration from cities where housing levels are unaffordable relative to other cities. The data observing the out-migration from cities is only a surface level observation and should be inspected with more nuanced considerations. There might be other economic factors at play which can bias the amount of out-migration observed. Bluestone, Stevenson, and Williams, (2009) did however undertake a study in this regard and found that over the 2000-2006 period when the US saw some of the largest rises in the level of unaffordable housing, those cities with the highest levels saw the most out-migration. More specifically, for 23 out of the top 35 most expensive cities the authors observed, the most expensive cities such as San Francisco and New York saw out-migration of six percent over the period and growth instead increased in cities that were slightly more expensive than average, cities such as Denver and Minneapolis that saw a 2 to 6 percentage point increase over the period (Bluestone, Stevenson, and Williams, 2009). A similar study was done in 2009 in California by the Centre for Continuing Study of the California Economy which looked at how changes in the ratio of California's house prices relative to the national median resulted in out-migration. The study found that in 2001 California's house price was 60% higher than the national median and California started seeing out-migration in 2004 when that ratio increased to 130%. This ratio increased even further in the period up to 2007; the rate of out-migration increased as well. As mentioned before this might indicate some form of effect on the growth potential of California but the study also notes that it might be the case that one observes some form of cashing out where Californians use their home equity to retire in more affordable states (Centre for Continuing Study of the California, 2009).

The literature discussed above looks at a very local form of economic development and does not emphasise the role of affordable housing levels as such of a specific city compared to the national average and the implications of this in on the economic development. This is because the studies focus on a local housing programme instead of the affordability level as a city-wide concept. This is important to distinguish otherwise one might conclude that a city's economic development is below potential growth due to a lack of government programmes for affordable housing instead of the level of affordability in the city itself hindering future firm-level investment and the absorption of new workers to the city. For this reason, it is important to build and use a formal economic and data-driven model to study this hypothesis, even though there might be studies that support this hypothesis to a small extent initially in surveys. The second part of the literature review discusses these studies.

Spatial models literature:

Formal models that use geographical characteristics to explain economic and especially urban growth were first proposed by Graves in 1976. However, there has been a renewed interest in urban economics due to Paul Krugman winning the 2008 Nobel prize partially because of his work on the New Economic Geography (NEG) (Behrens and Robert-Nicoud, 2009), (Partridge, 2010).

NEG seems to garner most of its appeal in Europe and Asia; however, there's less interest among U.S. regional and urban researchers (Irwin *et al.*, 2010). The lack of interest in NEG models among U.S. scholars is curious as one of the key benefits of the NEG is to endogenize agglomeration economies while allowing for internal migration (Partridge, 2010). In other words, city size and proximity to resources have a special role to play in NEG models. The NEG ought to be more attractive to U.S. researchers due to the United States' high internal migration rates (substantially larger than Europe) and larger labour market flexibility (Partridge and Rickman, 2005).

American researchers prefer to focus on the role amenities play, where individuals move to areas with high levels of natural amenities like warmer winters, distance to oceans, and attractive natural landscapes (Graves 1976, Glaeser, Kolko and Saiz, 2001, Partridge and Rickman, 2003). Although Graves wasn't the first scholar to emphasize amenities, he was the first to formally analyse its effects within the 1970's. (Irwin *et al.*, 2010). Partridge (2010) notes that one would not initially expect American researchers to follow the amenity (quality of life) models (which is a version of spatial equilibrium model where individuals migrate towards higher amenity cities until they equalize), one would instead suspect them to prefer models that explain migration due to strong city level economies. Yet, this doesn't minimize the attractiveness of urban agglomeration economies to American researchers using amenity models, particularly in describing a "special role" for the success of world cities like New York, Boston and San Francisco (Glaeser and Maré, 2001; Glaeser 2007).One important difference between NEG and the perspective of Glaeser and most U.S. urban amenity model economists like him, is that the they attribute a special role to human capital, data spill overs, marketplace pooling; and innovation (Partridge, 2010). However, a third perspective is held by Michael Storper and Allen Scott (S&S)—who are two leading contemporary scholars in the field of urban and regional economics. S&S believe NEG proponents regarding agglomeration economies' essential role as a cause of growth although they do not apply this to any formal models (Partridge, 2010).

Moreover, like Glaeser and his colleagues (and core NEG models), they see a special ability of successful/international cities like New York, San Francisco and Boston to utilize agglomeration economies (Scott, 2007; Storper and Manville, 2006). Yet, what stands out regarding S&S (2009) is that they disagree from Graves and his followers (including Glaeser) in their claim that amenity migration isn't a key issue behind American regional growth patterns (Partridge, 2010). According to Partridge (2010), S&S argue that individuals solely move to "nice places" with appropriate employment (S&S, 2009). They argue that agglomeration economies and amenity equalization forces take hold only after employees are interested in a specific location due to jobs. There is thus three differing views of how growth takes place as projected by the leading scholars of Urban Economics: Paul Krugman (NEG), Philip Graves (amenity-led growth), and Michael Storper and Allen Scott ("jobs" not "people" or amenities) (with the latter being more of an alternative than an expansion of the amenity-led growth argument) (Partridge, 2010).

Partridge (2010) attempts to compare these three models' accuracy by studying their accuracy in the United States over 40 to 60 year periods and finds that the Graves amenity led growth model is decisively more accurate than the other two. One of the most conclusive studies done in recent years is by (Sasser, 2010) which followed

individuals that moved from one state in the United States to another with data retrieved from the Internal Revenue Service. This allowed the author to use the difference in unemployment rates, per capita income, and housing affordability between the origin and destination state as explanatory variables to predict each state's out-migration rate. The results show that all three of these variables are very important in explaining migration, and they also experience changes in their strength as the national unemployment rate changes. This approach also follows from the amenity/special equilibrium model first proposed by Graves. Due to these observations, this study will build on this by using an amenity led model where differences in amenity levels serve to drive migration.

Chakrabarti and Zhang:

When it comes to the research studying the role affordable housing plays on cities' competitiveness, a more empirical and econometric approach is put forward by (Chakrabarti and Zhang, 2015). Their approach is very similar to this study where several city-related explanatory variables are collected across the largest metropolitan areas in the U.S. so that one can utilize this in a regression to study a dependent variable that will measure the 'competitiveness' of cities. In their case this dependent variable was employment growth over a 10-year period. The authors use a spatial equilibrium model first proposed Graves (1976) and expanded by Roback (1982) where there is an emphasis on the role of amenities in driving migration.

Chakrabarti and Zhang (2015) found that at first, it might seem that when one studies the correlation between employment growth and the unaffordable levels of housing, it will show that there is a positive relationship between employment growth and already high levels of unaffordable housing. In this study, it is important to note that the explanatory variable they use for unaffordable housing is the median house price-to-income ratio for a specific geographical area. By simply studying the surface level correlation between these two variables, it is clear that both have an effect on the other, which will lead to endogeneity in the model, which will bias the coefficient. In other words, a simple OLS regression that uses the house price-to-income ratio as an explanatory variable and employment growth as the dependant variable will only be of value if one assumes that employment growth has no feedback effect on the house price-to-income ratio. This is not the case as employment growth can be seen as a demand shock, which will affect the price of houses in the process.

Chakrabarti and Zhang (2015) try to eliminate endogeneity in their study by setting up a theoretical framework which follows from the model used previously by Roback (1982) which explains that housing affordability levels in cities are determined by each city's amenity levels where one can regard amenities as climate, crime, pollution, entertainment or access to the coast. The use of real household income is also included in the model, but the term 'amenity' model is used due to these models having a history of incorporating quality of life indices where all these variables are merged. Their model also explains that cities with high levels of unaffordability will also experience slower employment growth due to the land rents in that city being high because they are already at the inelastic part of the land supply curve.

The focus on the amenity level's role in determining the level of housing affordability might seem questionable at first but is an important part of their approach. Chakrabarti and Zhang (2015) realize that one cannot simply use an OLS regression to regress the house price-to-income ratio on employment growth because endogeneity is

present. Instead, they use an instrumental variable (IV) in the place of the house price-to-income ratio that serves as a proxy for the house price-to-income ratio. The authors argue that if amenities affect the house price-to-income ratio and the house price-to-income ratio in turn affects the employment growth, then one can use the amenity level as an instrumental variable to explain employment growth in a regression. In their case they use the mean January temperature as the instrumental variable for the house price-to-income ratio as there is a positive correlation between winter temperature and the house price-to-income ratio.

Their empirical strategy is to run two fixed effect regressions with employment growth as the dependent variable and the house price-to-income ratio and the mean January temperature (IV) as the explanatory variables, respectively. They also control for the log of the city population and the percentage of the city with a bachelor's degree. They use the fixed effect approach as this will give a more accurate picture of causality where one can study the within-city changes over time instead of merely comparing the different cities with a regression that does not include city-fixed effects.

The results of their study indicate that when they use the house price-to-income ratio as the explanatory variable in a city-fixed effect regression, it shows that the employment growth rate over a 10 year period was 9.8% lower when there was a one-unit increase in the house price-to-income ratio (roughly one standard deviation increase). For the fixed effect regression that used the IV (mean January temperature); it showed that employment growth over a 10-year period was 10.3% lower (Chakrabarti and Zhang, 2015). The obvious problem with this last IV approach is that the mean January temperature allocated to each city can't vary very much, which is a problem for a fixed effect approach. The authors try to use the temperature of two different 30 year means for each city to make the regression workable. However, it seems that the IV is still weak and this is also present in the small F-statistic when they try to see if the year-to-year correlation between the winter temperature and house price-to-income ratio is strong enough.

One other consideration one must make when using the mean January temperature as an IV to describe employment growth is that the winter temperature becomes drastically higher as one nears the coast in the US. This third unobserved variable which is the proximity to the coast, can influence both employment growth in the form of port access and the mean January temperature itself. This will call into question the validity of the IV. It is therefore necessary to find a better explanatory variable.

As discussed above, the major variable used to indicate the level of housing affordability in a city is the house price-to-income ratio. Those cities with a higher ratio will spend more of their income on housing and less on other goods while those cities with a lower ratio will have more to spend on goods. The benefit of this ratio is that it incorporates the demand side of the factors influencing the house price. If one were to merely work with house price as an indicator of affordability, one would exclude that fact that house prices are determined by income. Using the house price-to-income ratio, one considers that house prices are high due to demand. There is however an omission of the supply-side factors in determining house prices. To what extent is the numerator (house price) determined by supply. Glaeser and Gyourko (2003) studies this question in a very unique way by developing a novel ratio that incorporates the supply side. This is done by creating a measure of the theoretical replacement cost for a standard house in each city which incorporates the cost of land, construction costs (wages and materials)

and profit. The replacement cost for each is the theoretical cost of a building if one assumes that land cost is no more than 20% of the total cost of construction. By incorporating a set amount of land value one can then develop a house price to replacement cost ratio for each city; where a high ratio would indicate that house prices are too high for the replacement cost, indicating that house prices might instead be high due to other reasons such as a lack of supply that follows from planning restrictions in the city. This approach builds on Tobin's q ratio which was developed as a ratio of a firm's market value versus replacement cost but can also be used as a measure of regulatory constraints if the ratio is above 1.

Theoretical model:

In the models discussed above, several approaches incorporate the supply of land as a variable in determining city growth. However, this inclusion of the supply of land is included to make the models work and does not serve as the primary variable being studied in the model. The level of housing affordability in these models is used even less as a variable. When used, it is often the dependent variable being studied and not as an independent variable that determines city growth. The model proposed by Roback (1982) does however offer a viable theoretical model that explains the linkages between the supply of housing, incomes, and amenities for cities in a shared national economy where workers can migrate easily between cities. This model initiated what is known as the amenity models in the field of Urban Economics and was initially intended by Roback to deliver some form of quality-of-life index for each city. Chakrabarti and Zhang (2015) expands on the Roback model by letting the house price-to-income ratio play a larger part in the model. This approach will serve as the theoretical framework for this study. The core of the model must allow one to explain how increases in unaffordable housing influences the growth of cities.

The departure point of the Roback model is that each city is endowed with a certain level of amenities, where one can think of amenities as the climate (mean January temperature), the level of crime, proximity to the ocean and commuting times in minutes, we define this as a. In this city everybody is a worker and each worker receives an income in the form of a wage w and a non-wage income m and then uses both incomes to consume a plot of land s (which is used for housing but is different form the structure used for shelter), and a composite good x. The price of this composite good is determined on the international market, and we normalize it to 1 so that it can be the numeraire that serves as the base value we use to index the market with. There is also r which is the land rent. The land rent, amenities and wages in this model is different for each city. This then allows for a utility maximization problem for each worker being:

$$\text{Max } U(x, s, a),$$

$$\text{Subject to: } x + rs = w + m, \qquad (1)$$

If one then assumes that workers can freely move between cities with zero cost, and that the land rent in the equation is collected in the model and distributed to every worker in the form of non-labour income; one will arrive at an indirect utility function for each worker in a state of equilibrium:

$$V(w, r, a) = u. \quad (2)$$

For V one sees that utility u will increase with wages w and amenities a, and it will decrease with land rent r. Put another way $V_w > 0$, $V_a > 0$, and $V_r r < 0$.

After establishing the equilibrium state for workers, one wants to do the same for firms in this economy. Preferably these firms have to incorporate the labour of workers and other inputs such as land when they produce the composite good x (which is standardized to 1). When firms use workers and labour to produce the composite good one can write the production function as $f(n, d)$, where n is the number of workers and d is the quantity of land used in production. Chakrabarti and Zhang (2015) note that one can also include capital in the production function of the Roback model if you wish, but one will have to assume a fixed capital to labour ratio of say one machine per worker. If you incorporate capital in this fixed manner and you also assume that the price of capital is not determined on the local city level one will find that the production function can still be written as $f(n, d)$ in the model. Consequently, this shorter form of the production function is chosen for the sake of simplicity in the model. This then allows us to define firms' cost functions when they use workers and land to produce the composite good x. If one assumes constant returns to scale for the production function f, one can write the cost function of a firm when they produce one composite good as:

$$C(w, r) = 1 \quad (3)$$

Similar to workers, firms can enter and exit the market and move between cities with zero cost in this model. Equation 3 also shows that C increases with wages and land rents, implying $C_w > 0$ and $C_r > 0$.

By combing the variables used in the model above one can also deliver a formal definition for housing affordability which we will define as the house price to labour income ratio h (income in this model is equal to wages). However, one must first define the cost of housing, which in this case will be defined as the cost of the physical structure b that is constructed and the price of land rs. Note however that we define b as part of the composite good x, meaning just like the composite good, the structure is constructed on the international market and in equilibrium $b \ll x$ always. This means that a worker spends $b+rs$ of his wage on housing. This then allows us to define the house to income ratio as the following, where a higher h implies that housing is less affordable:

$$h = (b + rs)/w. \quad (4)$$

We now want to establish how changes in cities' amenity levels change the wages and land rents of cities. If we can do this, we can then show how amenities affect changes in h. In order to do that we differentiate equations 2 and 3 with respect to amenities a and then solve for dw/da and dr/da. By then substituting the signs of the V's and C's we established above into equations 5 and 6 we can now show the role of amenities on wages and land rents:

$$dw/da = V_a C_r/(V_r C_w - V_w C_r) < 0, \quad (5)$$

$$dr/da = V_a C_w /(V_w C_r - V_r C_w) > 0. \quad (6)$$

These two equations show that those cities with higher amenities will have workers that receive lower wages, furthermore, lands rents will be higher in cities with more amenities. The implications for firms are that they will need to pay higher land rents in higher amenity cities but will have to balance this out by paying lower wages for workers to still break even, keeping the cost of production *x* unchanged in the process. Workers will accept these lower wages as long as they can be compensated with an increase in amenities.

When considering the nature of demand for land one can state that *d(rs)/dr* > 0 which means that if the demand for land is inelastic workers will always pay more for land (rs) when land rents increase. If demand behaves in this way and equations 5 and 6 show that wages will decrease, and land rents will increase when there is an increase in amenities; one can extent this to equation 4 where an increase in land rents and a decrease in wages will increase h because it is a ratio. From this logic, one can then show that *dh/da* > 0 which states that higher amenities mean higher house price-to-income ratios. This is the first of the two major conclusions one can draw from this model.

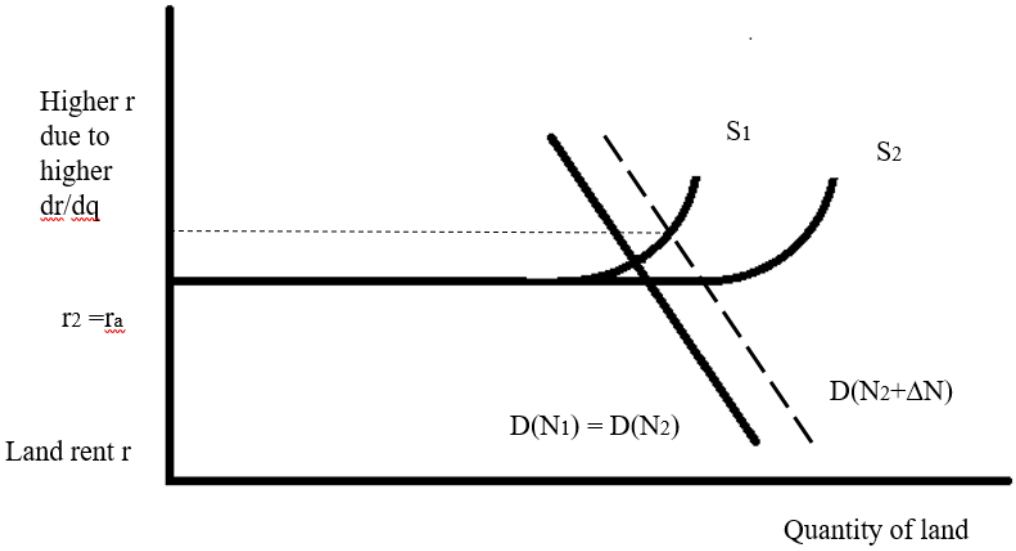
The following part of the model will focus on the second conclusion. Here we want to establish what will happen to changes in population (workers) ΔN if a city has an inelastic supply of land where *dr/dq* is sensitive and what will happen to ΔN if a city has an elastic supply of land where *dr/dq* is insensitive. In order to do this, we have to start by explaining how the land surrounding a city will influence the land rent in a city.

Suppose we assume that cities need to purchase agricultural land at the opportunity cost of that land for famers in order to expand cities and we further assume that each city is to some extent constrained by the supply of the surrounding agricultural land. In that case, we can then say the expansion of additional urban land can only take place with additional costs. We can write this inelasticity of land supply for a city *i* with the following function:

$$r = \begin{cases} r_a\,, & if\ \ q \leq \overline{q_i} \\ r_a + (q - \overline{q_i})^{\rho}, & if\ \ q > \overline{q_i} \end{cases} \qquad (7)$$

The function shows that urban land rent *r* will be equal to agricultural land rent r_a if the quantity of land required by the city *q* is less than the maximum quantity of land that the agricultural sector can provide if all of agricultural land were to be converted to urban land. This is symbolized by $\overline{q_i}$. If however the city needs more land than the agricultural sector can provide, then urban land rents will be increased with $r_a + (q - \overline{q_i})^{\rho}$, where $\rho > 1$ due to the supply of land becoming inelastic beyond the point where $q > \overline{q_i}$. This means that the sensitivity of land rents to land supplied *dr/dq*, is different at different points along the land supply curve and that *dr/dq* is higher when there is a lack of supply. This can be illustrated in figure 2 if one compares the higher *dr/dq* on the supply curve S1 with the lower *dr/dq* on supply curve S2 where the two populations are still equal.

Figure 2: Land rent prices and inelastic supply.



(Source: Expanded from Chakrabarti and Zhang, 2015)

By looking at figure 2 we can now ask what will happen if there is an increase in population in both cities symbolized by ΔN, leading to an increase in the demand of land as this new population needs a place to stay. It is clear from figure 2 that when this population increase plays out in both cities one and two, which start out with the same population, the increase in rents on the inelastic supply curve S1 will be higher due to the increased sensitivity of *dr/dq* already being present on S1. This shows that land rents will increase more for a given increase in population if a lack of supply is present in the city.

However, one must remember that this is an amenity model that emphasises the role that variables such as wages, land rents and amenities play on the utility of workers and that changes in population ΔN will also affect wages and land rents. These different paths can be illustrated by equation 8:

$$(V_r - V_w C_r / C_w)(\frac{dr}{dN})\Delta \mathrm{N} = \Delta \mathrm{u} \qquad (8)$$

To derive this equation, we assume that the price of the composite good is unaffected by the increase in population because the price is determined on the international market. We also differentiate equations 2 and 3 with respect to *N* and then we substitute for *dw/dN*. From equation 8 we see that a change in *N* will change *r* and when *r* changes it affects both a worker's utility through V_r , and through wages when firms cut wages to keep unit costs stable with $-V_w C_r / C_w$. The interpretation of equation 8 can be simplified if we assume that $V_r - V_w C_r / C_w$ remains

constant in each city[1] .Given this stability and the fact that indirect utility Δu is the same for each city due to cities reaching their equilibrium both before the increase in population and again after the increase, albeit in their own unique way. One can now see that the remaining variable that is unique to each city is its sensitivity for changes in land rent to population changes *dr/dN*. From equation 8 we can now see that a given city with a high *dr/dN* will be accompanied with a small change in population ΔN if utility changes Δu is constant. Put another way, population increases will be lower in a city with a built-in propensity to increase land rents by a lot due to the lack of land supplied in the city (inelastic supply curve).

This extension of Roback's amenity model done above by Zhang and Chakrabarti (2015) has two important conclusions that need to be discussed. The first is that cities with higher amenity levels also have higher lands rents. This is because workers are willing to receive lower wages and spend more on housing as long as they have higher amenity levels. This is true even if there is ample land supply. If two cities both have ample land supply, but one has high amenity levels, the city with the high amenities will have a higher house price-to-income ratio. It is important to note here that land rents in this first point are not high due to lack of supply but due to amenities and one cannot go one step further and claim that high amenities lead to unaffordable housing which in turn leads to low employment growth.

The second conclusion from this model is that high land rents that follow from a lack of supply leads to slower population and employment growth. When there is an increase in population in a city with a lack of land, the proportional increase in land rents is so high that it inhibits workers from moving there because it is not worth their while to move there. This means that there is a sub-optimal absorption of workers in the city with an inelastic supply of land.

These are two distinct conclusions, and each is arrived at in a unique way. Here it is important to mention that Chakrabarti and Zhang (2015) combine these two as a departure point to regress changes in employment growth on mean January temperature in a fixed-effects regression. In other words, they show that employment growth is low in cities with high amenity levels. However, this is a leap too far because the low employment growth is probably due to a lack of land supply. The reason they use mean January temperature is because they substitute it in the place of the house price-to-income ratio as an instrumental variable. This is because there is a correlation between US cities with warm winters and unaffordable housing. Even though warm winters explain unaffordable housing levels quite well in their regression, using it as an instrumental variable to explain low employment growth might explain other dynamics at work.

As mentioned previously the model being discussed is an amenity model and the main priority of the scholars researching this model is to find the best variables that explains the utility of workers in each city so that they can explain why people migrate from one city to the other. If there is one city with all the variables that increase its utility, workers will move there up to the point of equalizing out utility across the entire country. Researchers

[1] Ckakrabarti and Zhang (2015) notes that one can derive this assumption by imposing conditions on the indirect utility function and the cost function. For example, we may assume the following: (1) in the indirect utility function V, a is additively separable from r and w; (2) V is linear in r and w; and (3) C is linear in r and w. Together these assumptions imply that (Vr – VwCr /Cw) is a constant everywhere.

using this model have shown that one of the variables serving as a major predictor in population growth is winter temperature. There is however a problem with using winter temperature as a standalone predictor of employments growth. Kemeny and Storper (2012) caution that when scholars want to predict the rise and growth of the Sunbelt States due to warmer winters, it does not take full account of the decline of manufacturing in the interior of the country where winters are colder. Kemeny and Storper (2012) argue that if manufacturing has declined in the interior of the US over the past 50 years and there was lower employment growth due to that in the interior and that employment growth was stable in the less industrialized South, then winter temperature would falsely predict employment growth because it is actually picking up the fact that manufacturing heavy cities happened to be positioned in the interior of the country. The authors argue that winter temperature fails to explain why employment growth over the past decade was highest in Alaska and North Dakota. This they argue is due to the oil boom seen in these states and the fact that firms choose to employ more workers there. In other words, the ultimate predictor, according to them, is the pressure on households finding employment and not amenities or weather.

This claim by Kemeny and Storper (2012) should be taken into account when doing empirical research but does not justify the elimination of winter temperature in regressions, it instead shows that manufacturing employment should play a role alongside winter temperatures when we explain employment growth.

Empirical strategy:

To fully study the role of unaffordable housing levels on the competitiveness of a city one has to choose the appropriate dependent variables to measure competitiveness. This study will use two variables to measure this competitiveness, more accurately the growth in competitiveness. This will be in the form of population growth and median household income growth. This is because we assume that growth in population and incomes are the traits associated with an increase in a city's competitiveness. In this regard cities are similar to firms; when a firm becomes more efficient this increase in efficiency will normally manifest itself in the form of a growing labour force and higher growth rates in wages relative to its competitors (assuming full employment in the economy), implying it is more competitive.

Besides the economic intuition, these two dependent variables indicating competitiveness are also associated by the wider public, municipal managers and politicians as indicators of a "booming" city. Even though competitiveness can be measured by other means such a constructed composite indicator, it is not a given that changes in a composite indicator will be clearly useful for both the public and economic interpretation of competitiveness. Furthermore, a composite indicator such as this will also not be relatable to the variables used in other studies because it will be too specific to one study. The fact that these variables are also major components in spatial equilibrium and amenity models in the field of Urban Economics underscores their appropriateness for the purposes of this study.

The empirical study will regress and discuss the following two equations in order to study the role of affordable housing on the competitiveness of cities.

$$\Delta N_{i,t} = \beta_0 + \beta h_{i,t-1} + \lambda X_{i,t-1} + \tau_t + u_i + \varepsilon_{i,t} \quad (9)$$

The dependent variables in equation 9 is the growth in population for city *i* in period *t*, $h_{i,t-1}$ is the key independent variable measuring housing affordability in city *i* in period *t*-1. A higher *h* will indicate unaffordable housing; this is because *h* will either take the form of the house price-to-income ratio or the house price to replacement cost ratio. Decisions on which to choose will be made as the study progresses. $X_{i,t-1}$ will represent the vector of control variables for the previous period. This will be variables such as median householde income, city population and the share of the population with a bachelor's degree or higher. τ_t is the year fixed effect; u_i is the unobserved heterogeneity for a given city *i* and $\varepsilon_{i,t}$ is the error term.

$$\Delta I_{i,t} = \beta_0 + \beta h_{i,t-1} + \lambda X_{i,t-1} + \tau_t + u_i + \varepsilon_{i,t} \quad (10)$$

Equation 10 is very similar to equation 9; the only difference being that the dependent variable is the growth in median household income for a given city *i* for period *t.*

In order to accurately estimate equations, 9 and 10, one needs to incorporate a fixed-effects model. If one were to use a simple OLS regression, the estimated coefficients for β would be biased for both equations 9 and 10. This preference for a fixed effect model over a simple OLS regression for a study like this is explained by Chaktrabarti and Zhang (2015) when they conducted a similar study. According to Chaktrabarti and Zhang (2015) if one were to consider a simple OLS regression of equation 9, several factors lead to a biased coefficient β. This is because our model assumes that when there is a shock, the adjustment process will be immediate, and the new equilibrium will be attained quickly. This is however problematic as this is not the case in practise where something such as an employment shock in a city will play out over several years. By using this data, when the system is still out of equilibrium, one will get a biased coefficient. When one uses a simple OLS regression in a period where there is a positive economic shock such as oil being discovered in San Francisco, it will lead a duel increase in population and land rents and thus house prices, if we assume that San Francisco has an inelastic supply curve. Even though the population might migrate away from San Francisco in later periods, the OLS coefficient will be positive. Due to endogeneity over this period. This is illustrated in figure 3. This same bias is present if there is a supply shock in land. If one collects data over a period after a supply shock, say San Francisco only allows single level residential developments, and after the shock house prices are higher but then house prices move lower as the population moves to other cities to reach a new equilibrium. During this process, one will also observe a positive and biased coefficient which is illustrated in figure 4.

Figure 3: Demand shock for land with inelastic supply.

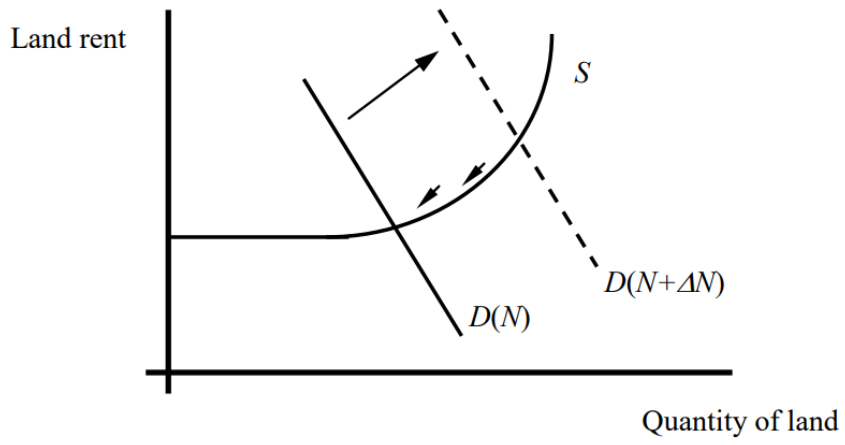
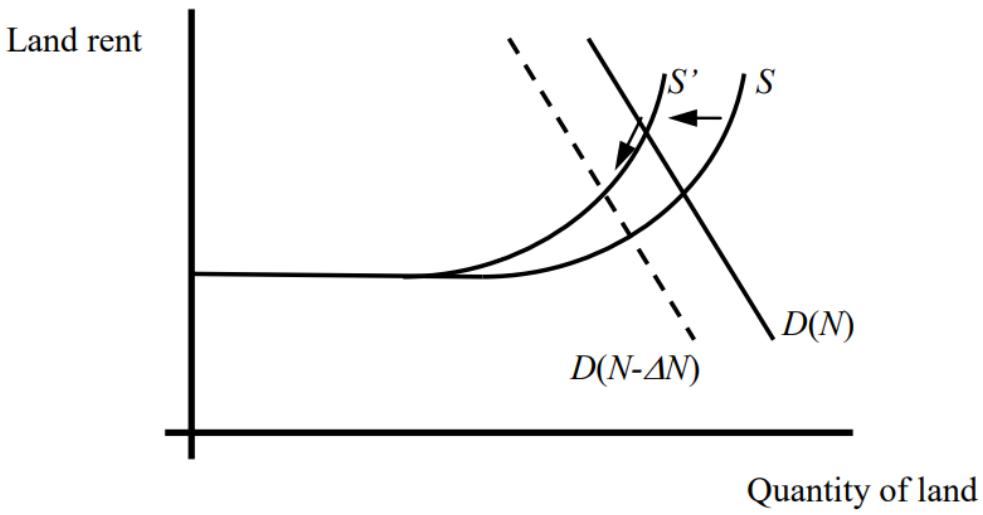


Figure 4: Supply shock for land with inelastic supply.



(Source: Chakrabarti and Zhang, 2015).

The bias present in the simple OLS regression can be addressed by incorporating specific changes. The first is to use the explanatory variables from the previous period, *t-1*. This is done because we assume that when growth is not anticipated in the dependent variable in time *t*, it will not affect the explanatory variable in time *t-1*. This allows for the removal of endogeneity to a certain degree.

Secondly, we use city fixed-effects in the model to receive a within-city estimation for say, the effect of unaffordable housing on population growth. By using a fixed-effects model instead of a simple OLS regression, we remove the unobserved heterogeneities u_i from the model for each city, giving us less biased coefficients.

This study's empirical strategy will thus use a fixed-effects model as a point of departure where there are location (city) fixed-effects in the main equations 9 and 10. This then removes the unobserved heterogeneity term u_i from equation 9 and 10 to then deliver equations 11 and 12 for a fixed-effects regression. As the study progresses, the

explanatory variables' appropriateness in equation 11 and 12 will be considered, and then the fixed-effects model itself will be changed to incorporate time-invariant variables such as winter temperature.

$$\Delta N_{i,t} = \beta_0 + \beta h_{i,t-1} + \lambda X_{i,t-1} + \tau_t + \varepsilon_{i,t} \quad (11)$$

$$\Delta I_{i,t} = \beta_0 + \beta h_{i,t-1} + \lambda X_{i,t-1} + \tau_t + \varepsilon_{i,t} \quad (12)$$

The first explanatory variable to consider is the housing affordability ratio $h_{i,t-1}$, where a higher $h_{i,t-1}$ will indicate less affordable housing. The majority of previous studies have used a house price-to-income ratio as an indicator for housing affordability. This study however considers an alternative ratio for $h_{i,t-1}$, which will be the house price to replacement cost ratio $hrc_{i,t-1}$This will allow for a variable that considers the restrictiveness of zoning laws, which in turn raises house prices due to a lack of supply. This is in contrast to the standard house price-to-income ratio $h_{i,t-1}$, which we will now call $hi_{i,t-1}$,.

Housing prices differ substantially across cities. This variability is however the result of a multitude of factors at work in determining the level of house prices for each city. When using house prices to explain the affordability of housing in cities one of the main challenges is to establish if house prices are high due to zoning restriction which lowers supply in the housing market or are they high due to an increase in demand as the disposable income of consumers increase. In order to isolate the housing demand from housing supply Glaeser and Gyourko, (2003) developed a very useful metric to indicate to what extent house prices are high due to supply restrictions. It builds to some extent on Tobin's q-factor where current prices of a good are compared to the replacement cost of that good. Glaeser and Gyourko (2003) extend this idea to housing prices to compare the current market value price of housing versus the current replacement cost of that same house. The denominator is constructed using the per square feet cost of constructing a certain kind of house (in this case, an economically single-level detached house of 2000 square feet). Each city's cost of construction per square feet is different due to variables such as wages, transportation and materials cost being different on the individual city basis. By multiplying the individual square footage cost by a standardized floor area of 2000 feet, one gets the hypothetical cost of replacement that also assumes that the land price is equal to 20% of the construction cost. This value needs to be adjusted slightly to include the profit margin and operations cost to deliver a more realistic replacement cost value. A 17% increase is used by Glaeser and Gyourko (2003) to incorporate the profit margin and the cost of operations and is called the marginal profitable production cost (mppc). By constructing a ratio of market housing prices to mppc, one can better observe if house prices are high due to demand or supply. If the house price to mppc ratio is high, it would indicate that house prices are high not due a demand increase but due to restrictions in supply that normally follow from zoning and urban planning restrictions in neighbourhoods and cities (Glaeser and Gyourko, 2003). If the house price to mppc ratio is high in say a city with little supply restrictions one would expect to see an increase in housing supply as it becomes profitable to construct the new supply at the mppc level and sell it at the market house price level. If this continues for long enough in a city with little supply restrictions the numerator will decrease and the house price to mppc ration will stabilize. If there are however strong zoning restrictions in place it would not be captured by the mppc value and instead it will spill over into the market price of housing. Even if

housing developers increase supply at the mppc value they would include other costs from zoning restrictions which ultimately force them to sell at the market house price instead of the mppc price to break even.

There are two benefits with using the $hrc_{i,t-1}$ as an indicator for the restrictiveness of a city's zoning laws. The first is that there is improved incorporation of construction cost *b* from the theoretical model which in turn helps to isolate the cost of land rs that is high due to zoning restrictions, where $hrc_{i,t-1}$=market house price/replacement cost, and restrictions on zoning laws will increase the price of land *rs* which will in turn raise the market house price numerator. This is shown in equation 13:

$$hrc_{i,t-1}\text{=market house price/1.17*}(b+rs)\text{ where rs=0.2*b} \qquad (13)$$

The second benefit of using $hrc_{i,t-1}$ is that it better explains the causes of unaffordable housing, or rather that part of unaffordable housing that causes the population of a city to emigrate. From the theoretical framework that is illustrated in figure 2 one can see when a city is already at the inelastic part of the supply curve where the house price-to-income ratio is high, the future growth in the house price-to-income ratio is expected to be higher than the growth seen for a city that is on the elastic part of the supply curve. This is due to the unaffordable city's house price being more sensitive to a proportional increase in demand versus an affordable city. Table 1 shows that this sensitivity due to inelasticity can be better explained by $hrc_{i,t-1}$ instead of $hi_{i,t-1}$. This shows that when cities saw a growth in their unaffordability over a 5-year period that growth can be better explained due to restrictive zoning laws than unaffordability in the previous period. This is shown in the first column where, as expected, when cities experience increases in their own house price to replacement cost ratio, then growth in unaffordability in the following period seems to increase. This is highlighted with the positive coefficient for $hrc_{i,t-1}$ in the first column and the higher R^2. Due to these considerations, one can make a case for the $hrc_{i,t-1}$ instead of $hi_{i,t-1}$ to be used as an explanatory variable to predict both the population growth and income growth of equations 9-12.

Table 1: Predictors of house price-to-income growth

VARIABLES	Growth rate of (D.log):	
	Column 1 House price-to-income ratio	Column 2 House price-to-income ratio
House price to replacement cost ratio in the previous 5 years	0.113***	
	(0.008)	
House price-to-income ratio in the previous 5 years		-0.065***
		(0.011)
Constant	0.021***	0.110***
	(0.001)	(0.015)
Observations	2,529	2,529

R-squared	0.090	0.015
Number of metropolitan areas	281	281
Type of regression Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1	City -Fixed effects	City -Fixed effects

One must also consider that there is an added issue with the $h_{i,t-1}$ ratio if it is the case that cities with higher incomes have a higher percentage of citizens that can make a down payment on a mortgage which will increase the demand for houses leading to an increase in prices. This will then skew the $hi_{i,t-1}$ ratio towards higher-income cities outside their expected house price-to-income range and it will then be biased towards down payment rates instead of purely unaffordability.

Control variables:

The second part of the empirical strategy is to include appropriate control variables for term $X_{i,t-1}$ in equations 11 and 12. Similar to the unaffordability term $h_{i,t-1}$ the regressions use the lagged term for all the variables in vector $X_{i,t-1}$ so that there is some removal of endogeneity which will better explain the effects that follow from the control variables on that of the dependent variables of income and population growth. If we were to use the $X_{i,t}$ term instead, there would be an increased risk of picking up endogeneity between the dependent variables and the explanatory variables in such a regression.

The control variables do however form a major part of this study and including them just on the basis of econometric aesthetics would be premature. This is why each of the control variables and the history and literature behind their use will be discussed in this study's interpretation section. The control variables themselves have formed part of major studies in the literature of urban economics and the role explanatory variables such as winter temperature, population size, median household income, share of the workforce with a bachelor's degree or more, share of the workforce in manufacturing; play in the growth of a city's population and income, are important and can even be counterintuitive.

The control variables in the fixed effects regression will then be expanded to include the time-invariant variable of winter temperature as this has a major influence on the growth in population and has had differing degrees of importance for different eras (Glaeser, Ponzetto and Tobio, 2014). This will necessitate a random-effects model in order to include the time-invariant component. This will be done at the cost of exposing the regression to endogeneity by including the u_i term but is done for two reasons: to see if winter temperature corresponds with the literature in explaining population growth and to use it as a steppingstone for a Hausman-Taylor regression, which will help remove the endogeneity from a random-effects model with a time-invariant component. The random-effects equation for population and income growth is as follows and includes the u_i term and the time-invariant vector of mean January temperature Y_i :

$$\Delta N_{i,t} = \beta_0 + \beta h_{i,t-1} + \lambda X_{i,t-1} + \theta Y_i + \tau_t + u_i + \varepsilon_{i,t} \qquad (14)$$

$$\Delta I_{i,t} = \beta_0 + \beta h_{i,t-1} + \lambda X_{i,t-1} + \theta Y_i + \tau_t + u_i + \varepsilon_{i,t} \quad (15)$$

This part of the random effects study and its explanatory power will then highlight the need to search for a better but still missing explanatory variable for population growth, which has also been highlighted as a problem in previous studies and serves to critique amenity and spatial equilibrium models in the literature (Kemeny and Storper, 2012). The search for a better explanatory variable to explain population growth is then attempted with the inclusion of the lag of population growth itself as an explanatory variable. The need for this inclusion is discussed in the interpretation section but concludes that in order to fully understand population growth in cities one must understand how economic "booms" play out in cities over different epochs and that the innate characteristics (amenities) of cities only partially explain population growth.

Data:

The empirical analysis is conducted on the metropolitan statistical area (MSA) level for the largest cities in the United States by using the weighted average (based on population share) of the data of each county that falls within the same metropolitan area. The data spans over the 1940 to 2017 period but the data from 1972-2017 is used as the primary dataset for the regressions as this period contains additional variables allowing for a panel dataset of 5-year intervals between the period 1972-2017. In this section there is a description of the data sources and variables used in for the regression output.

Dependent variables:

U.S. metropolitan statistical area population growth: The data is sourced from the U.S. census for the years 1940-2000 by using the County Data Books which contain demographic, social and economic data from both the U.S. Census Bureau and private organizations. The electronic version of this is contained in the ICPSR table 2896 which is made available by the Inter-University Consortium for Political and Social Research. The population used for 2017 was sourced from the American Community Survey (ACS) which replicate the data for the Census but utilizes a smaller sample at the county level. The variables of 2017 that correspond to those in the 1972 to 2000 period were then added to the panel dataset. The dependent variable is then adjusted by taking the first difference of the logarithm of population to deliver a population growth rate over a 5-year period for the panel dataset. All the regressions that have population growth rate as the dependent variable will take this format.

U.S. metropolitan statistical area median household income growth: Similar to the population growth data, data was retrieved from the County Data Books and combined for the same metropolitan area for the years 1940-2000 and the income data for the year 2017 was added from the 2017 ACS to correspond with the 1972-2000 panel dataset. The dependent variable is then adjusted by taking the first difference of the logarithm of income to deliver an income growth rate over a 5-year period for the panel dataset. All the regressions that have income growth rate as the dependent variable will take this format.

Metropolitan statistical areas are multi-county units, defined using the 1999 Census definitions of primary metropolitan statistical areas (PMSA) and New England county metropolitan areas (NECMA). The metropolitan-

area population data are combined by using a weighted average from the county-level data using a consistent definition of geography over time, so changes in metropolitan-area boundaries do not affect any of the dependent or explanatory variables. The main panel dataset consists of 281 metropolitan/city areas for the years 1972-2017 and a larger amount of metropolitan areas for the OLS regressions as the regression normally contains two to three variables.

Independent variables:

All of the explanatory variables excluding the mean January temperature will be adjusted by taking the logarithm of the variables to show the effect of a 1% increase in the explanatory variable on that of the dependent variable. Most of these variables will also be in the lagged format where the explanatory variable in the regression symbolizes the variable 5 years ago.

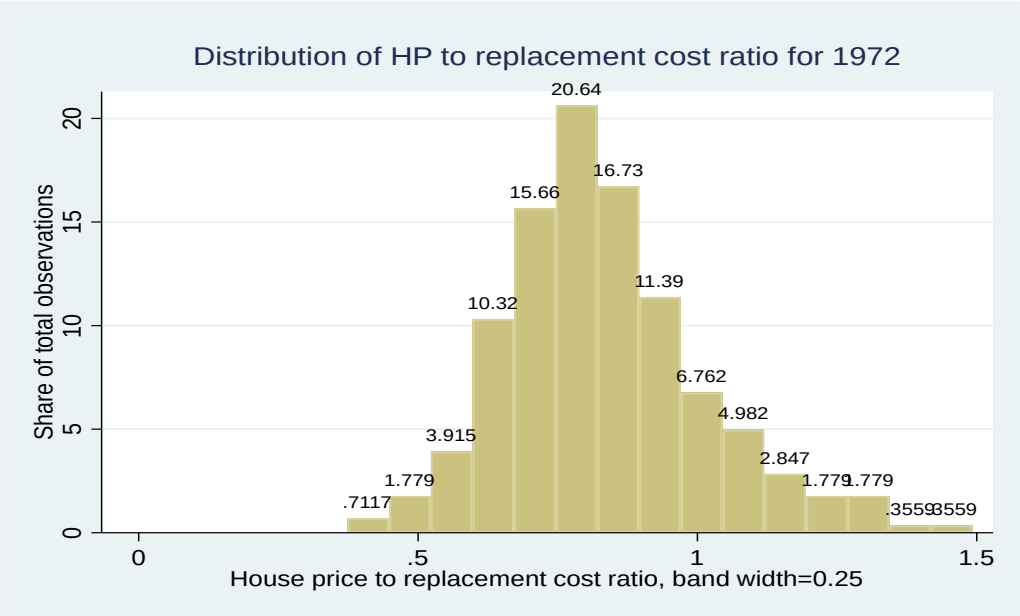
Two main independent variables for housing affordability:

U.S. metropolitan statistical area median house price to median household income ratio: The data for both house prices and incomes are sourced from the County Data Books for the 1940-2000 period and the data for 2017 was added from the ACS. All the data are nominal values and are unadjusted for inflation as we are interested in the ratio and not the real value. The median house price was however adjusted for each year as the data did not correspond with the Case-Shiller house price index. After the adjustment of each year's data according to the Case-Shiller index (nominal), one removes the bias in the house price data that follows from house surface areas increasing drastically over the 1940-2017 period. This is because the Case-Shiller index records the sale price of the same houses being sold multiple times over long periods which keeps quality constant. This adjustment is a level adjustment done by year, and one still retains the difference between one city and the next for a given year which is crucial for panel data. The house price is then divided by income to deliver the wanted ratio.

U.S. metropolitan statistical area median house price to replacement cost ratio: Similar to the previous variable, the house price data was retrieved from the same sources and adjusted in the same manner. The construction cost data was retrieved from the Simple Est unit price generator from Design Cost Data which delivers the nominal cost of construction per square foot of a single story and free-standing residential house of 2000 square feet. This allows for a useful comparison of construction cost for the same house in each city and changes by survey year. Due to construction wages and material prices being higher in more expensive cities such as San Francisco the price per square foot is higher but one still observes that house prices deviate substantially from the cost due to high land prices and restrictive zoning laws. The data for 1940 is constructed by using the construction cost data in each city for the year 1972 and by adjusting the level by using the Case-Shiller construction cost index we arrive at the 1940 level. Although the difference between cities for the year 1940 is unadjusted it is still of some value given that the year 1940 is only used in OLS regressions. After having a cost of construction per square foot for each city, we multiply the value by 2000 square feet and then increase the value by 20% to include the cost of land and this new value is then increased with 17% to include the profit margins and operation costs of getting the property to market. This then delivers the theoretical replacement cost of a house in a city with low zoning restrictions where land supply is very elastic. By dividing the house price with the replacement cost, we get the

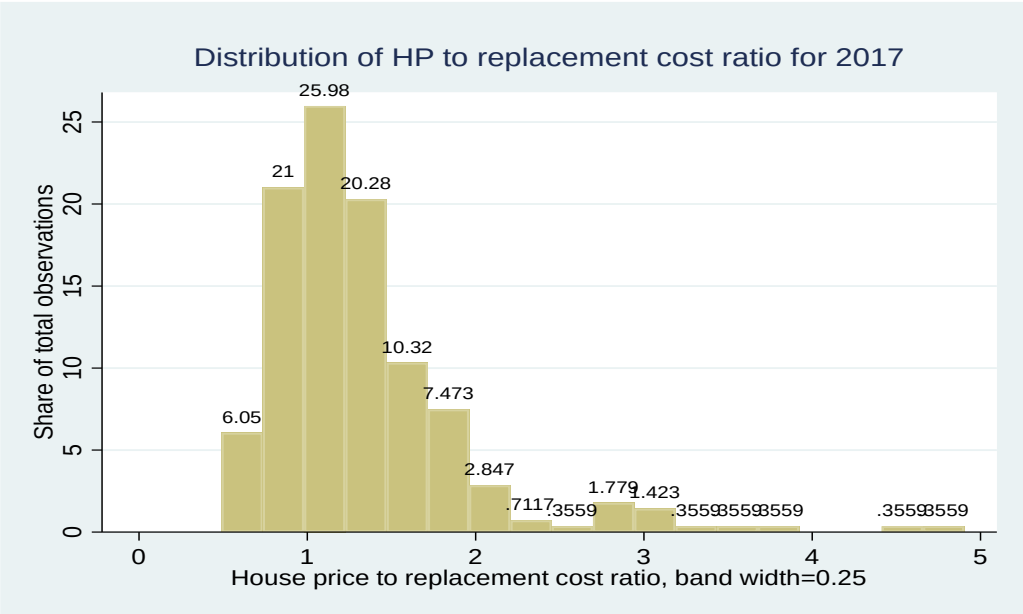
wanted ratio. This is the same method and percentage used by Glaeser and Gyourko (2003) to arrive at a house price to replacement cost ratio in their study. The change in this ratio is illustrated in figures 5&6 where one can observe the distribution of the house price to replacement cost ratio for years 1972 and 2017. Here there is clear evidence that the mean of the ratio has increased from 0.83 in 1972 to 1.33 in 2017 and that a larger share of outliers can be found in the right tail, showing that cities like San Francisco with a ratio of 4.90 and San Jose (Silicone Valley) with a ratio of 4.54 in 2017 find themselves in markets where the supply of land is very inelastic.

Figure 5.



(Source: Calculated by the author)

Figure 6.

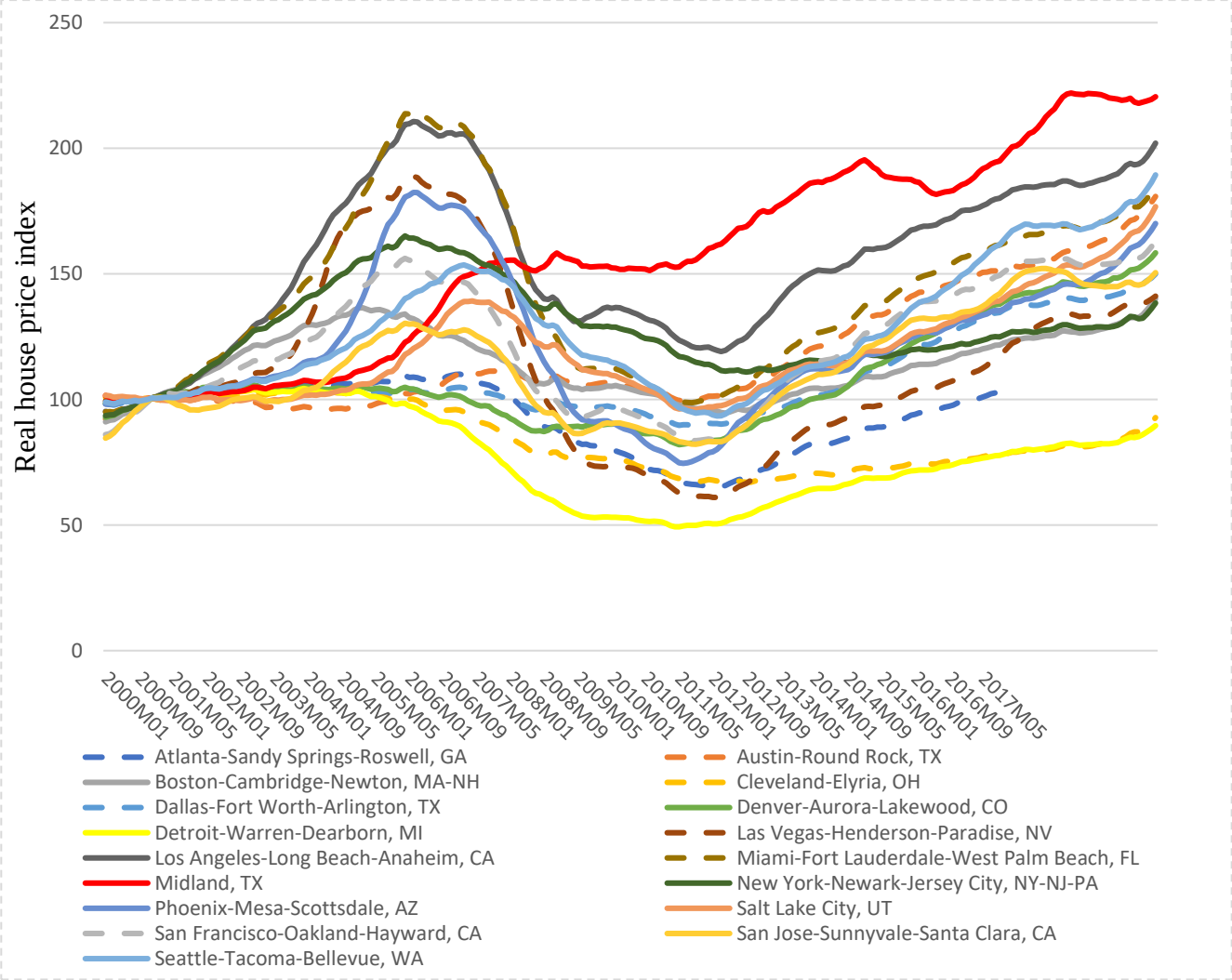


(Source: Calculated by the author)

One can also see from figure 7 how the real house price index has grown for different cities in the U.S. The index base of 100 is equal to November 2000 and delivers a better picture for the post-2000 period that saw the highest

growth rates in real house prices, as shown in Figure 1. With closer inspection, one can observe caveats of growth by looking at the house price graphs. One can observe that Rust Belt cities such as Cleveland and Detroit have seen negative real house price growth and that cities with warm winters such as Miami, Los Angeles, Las Vegas and Phoenix experienced immense increases in price during the housing market boom. In that same boom period, we see that other city with elastic housing supply such as Atlanta, Dallas and Austin did not see the same increases in house prices even though they saw rapid population growth in this period. Lastly, one can observe Midland, Texas, which is home to the shale boom in the U.S., showing the highest growth even though it finds itself in a state with an elastic housing supply.

Figure 7: Real house price index for several cities, 2000-2017.

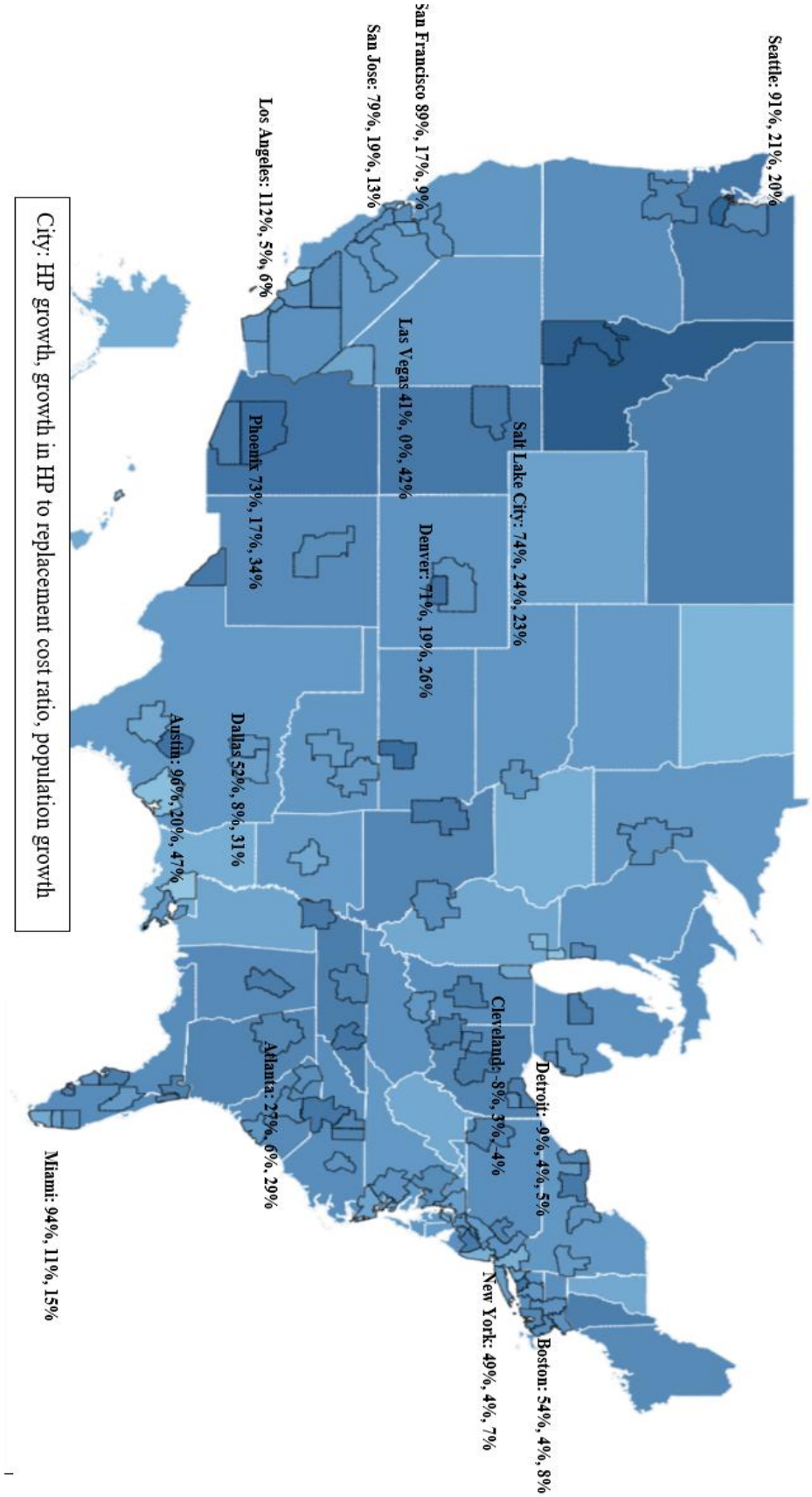


(Source: Calculated by the author)

Figure 8 adds to figure 7 by combining the real house price growth, growth in the house price to replacement cost ratio and the population growth over the 2000-2017 period. From the map, one can observe the low population growth in recent years in unaffordable cities such as New York, Boston and all the cities on the West coast. One can also observe that population growth was higher in warmer cities with elastic supply such as Atlanta, Dallas, Austin and Las Vegas. Figure 8 can also be interpreted in a different way if one observes that cities such as San Francisco and San Jose experienced low population growth yet high increases in real house prices and cities such

as Dallas and Atlanta experienced much higher population growth rates yet lower increases in real house prices. This can be interpreted as an indication that different zoning laws between cities are responsible for the observed elasticity in house prices between cities. House prices will thus react differently given the same influx of population based on the zoning laws associated with a given city.

Figure 8: Map of real house price growth, growth in the HP to replacement cost ratio and population growth.



(Source: Calculated by the author)

Control variables:

All of the control variables were retrieved from the County Data Books 1940-2000 and then combined in the same manner as the dependent variables to deliver metropolitan area-level data. Data for all the control variables for the year 2017 was retrieved from the ACS and added to the dataset. These control variables are the following: percentage of the workforce with a bachelor's degree or higher, percentage of workforce working in the manufacturing sector, percentage of workforce working in the professional services sector, which include finance, consulting, information technology, education and healthcare; the percentage of households that own the house they live in and the time-invariant variable of mean January temperature in degrees Fahrenheit where a lower temperature indicates the severity of the winter.

Empirical results:

The empirical results analysis is done by using MSA level data for cities in the United States. This is done by merging county level data with the same metropolitan id number to form metropolitan area data that encompasses the entirety of a specific city's economic activity. This is done for the largest cities in the United States which allows for the largest possible sample to fit the empirical model. The empirical result analysis will start with the discussion on the individual control variables, starting with the role of population size and growth and then followed by the incorporation of the housing affordability ratios. The discussion will then end where it started with population but now there will be a revision with the emphasis put on a city's ability to sustain an economic expansion as an important but under-emphasised issue in the empirical data.

Population:

For the purposes of this study population as a variable is used both as a dependent variable and as an explanatory variable at different stages, as mentioned before population is used as an explanatory variable because it is one of the main indicators for economic growth and is observable during the economic boom of a city and it is used as an explanatory variable because it indicates city size and determines agglomeration effects. This discussion of population as a variable has to start with the nature of population size and growth in the United States over the very long term.

One of the strongest relationships seems to be present with the persistence of city population size over long periods of time. If cities were large in 1940, the chances are good that they will be large in 2017. This same argument also holds for their population rank. This is illustrated by regressing the log of population in 2017 on the log of population in 1940:

$$\log(\text{Pop in 2017}) = \underset{(0.439)}{4.72^{***}} + \underset{(0.037)}{0.70^{***}}.\log(\text{Pop in 1940}) \qquad (16)$$

The most striking observation is the high coefficient of 0.7, which shows that over the post-war era up until 2017, the initial population seemed to be highly dependent on the population size in 1940. Equally striking is the high R^2 of 0.543, which corresponds to a correlation of 0.737 in the 299 observations. As Glaeser, Ponzetto and Tobio

(2014) points out, it is important to note that this merely points to the persistence of population size over the long-term and not necessarily to the subsequent growth rate of population. It is not necessarily the case that larger cities deliver subsequent larger growth rates in population. This will be discussed shortly. First one has to look at shorter time periods in the post war era to understand if the persistence shown by equation 16 is also present in these shorter periods. The first period is between 1940 and 1972, which saw the highest economic growth. It includes the so-called Golden Fifties and Sixties where income growth was strong and excludes the era that saw the decline of manufacturing cities such as those in the Rust Belt that started in the 1970's. This era also roughly includes the Bretton Woods era, which precedes the high inflation of the 70's and 80's.

The second subgroup is the era between 1972 and 2000. This is the era that saw the decline of manufacturing in the Rust Belt and the pivot towards a services-oriented economy. By 1972, agriculture's role in the economy is also much smaller compared to what it was in 1940, and the shift from rural to urban populations has largely taken place. This era also saw rises in real house prices that were significantly larger in 2000 compared to what they were in 1972. The third group takes two important changes into account. This incorporates the years 2000 to 2017 which saw the final shift to the services economy led by the ICT industry and an even stronger rise in the real price of housing.

Table 2: Log of population between years (OLS).

	Log of population by year			
VARIABLES	2017	1972	2000	2017
log of Pop in 1940	0.6992***	0.8561***		
	(0.0372)	(0.0215)		
log of Pop in 1972			0.9341***	
			(0.0182)	
log of Pop in 2000				1.0165***
				(0.0063)
Constant	4.7167***	2.3479***	1.2095***	-0.0778
	(0.4389)	(0.2531)	(0.2260)	(0.0810)
Observations	299	299	299	299
R-squared	0.5427	0.8391	0.8950	0.9882
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				

Table 2 shows this long-term persistence for different periods of the post-war era. It is important to note that this persistence seems to increase for periods closer to 2017, as is shown by the increasing coefficients and R^2's. At first this seems obvious as the periods are shorter compared to that of the first column but note the increase in

both coefficients and $R^{2\prime}s$ from column 2 to 3 as the two columns have roughly the same periods. This increase might be indicative of the permanent shift from the rural population to the urban population. Simply put, if rural populations are more indifferent about city locations than urban populations, then a smaller rural population source (such as that in 1972) will make it more difficult for cities to change their subsequent population rank.

$$\text{Diff. in log (Pop in 2017)} = \underset{(0.439)}{4.72^{***}} - \underset{(0.037)}{0.30^{***}}.\text{log (Pop in 1940)} \qquad (17)$$

As previously mentioned, one must distinguish between population size persistence and population growth following from population size. The data does not indicate that larger cities deliver subsequent higher population growth rates. Equation 17 in fact shows that larger cities saw lower population growth rates. Put another way; equation 17 shows that smaller cities grew faster. This is why it is important to work with metropolitan units and not merely urban county data, as county units would bias the data due to higher growth in suburban areas. When metropolitan units are used, it incorporates the population spill-over from the inner city to the suburbs during growth periods, otherwise it would overestimate the role of population growth in lower population areas such as the suburbs. What is however not clear from equation 17 is the mechanics of this relationship between population growth rates and initial population. Do smaller cities grow faster because larger cities are "saturated" somehow or is there some type of mean reversion in population, similar to the mean reversion one sees with household incomes, where cities with initially lower household incomes see higher subsequent growth in household income? This question will be answered by incorporating different variables to explain this observation better.

The interpretation of equation 17 can however be improved by focussing on shorter periods. Namely those of 1940-1972 and from 1972 to 2017. This is illustrated in table 3.

Table 3: Growth in log of population given initials levels (OLS).

	Growth in log of population	
VARIABLES	Column 3 1940-1972	Column 4 1972-2017
Log of population in 1940	-0.1439***	
	(0.0215)	
Log of population in 1972		-0.0789***
		(0.0235)
Constant	2.3479***	1.5042***
	(0.2531)	(0.2923)
Observations	296	296

R-squared	0.1284	0.0351
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1		

From table 3 it is clear that in both periods there was a negative relationship between population growth and initial population levels- similar to that of equation 17. Column 4 does however indicate that this relationship weakens dramatically from 1972 onwards. Both the coefficient and R^2 of column 4 seems to be weaker. This weakening of the role of initial population levels can be interpreted in a number of ways. One version is that in column 3 smaller cities saw higher growth rates due to the influx of the rural population to cities. Hypothetically, if the same number of people from rural areas migrated to cities, then a city with a population of a hundred thousand would see a higher growth rate than a city with a million. If by 1972 the majority of the rural population has moved to cities, this bias towards smaller cities will weaken. The second hypothesis would be that by 1972 either real house prices or the cost of living in larger cities have become prohibitively expensive, so much so that it would dampen the population growth of larger cities. Larger cities would in effect be "saturated". The second hypothesis can however be tested by incorporating other variables such as income.

Glaeser, Ponzetto and Tobio (2014) does however note that the population growth rate itself persists over shorter periods of time and have shown for most decades over the 1790-2000 period there is a 0.5 and higher correlation coefficient between the population growth of a decade and the decade preceding it. This correlation has also markedly increased in the post-war period. In contrast to this, Glaeser, Ponzetto and Tobio (2014) also showed that over the very long term there was actually a negative correlation between the 1860-1930 period and the 1930-2000 period. According to Glaeser, Ponzetto and Tobio (2014), cities experience their economic booms over different epochs and while there is economic momentum over the short-term structural changes in the city's economy will ultimately prevail.

Household Income:

The same observation for household income is seen in cities. Similar to the mean reversion of population growth, income seems to mean revert. The mechanics of this mean reversion might however be due to totally different drivers. First one must look at the long-term observations for initial income and subsequent growth. This is shown by table 4 which shows that from 1972 onwards income growth had a weaker relationship with the initial income level. This relationship seems to be much stronger for column 5 which spans the 1940 to 1972 period. One might conclude that this observation is very similar to the weakening seen for population growth in table 4 and speculate that there must be a strong relationship between population growth and income growth. This might be the case, but it is also important to notice the higher constant in column 5 in table 4 compared to the constant in column 6.

Table 4: Growth in median household income given initial levels (OLS).

	Growth in household income	
VARIABLES	Column 5 1940-1972	Column 6 1972-2017
Household income in 1940	-0.9583***	
	(0.0087)	
Household income in 1972		-0.2835***
		(0.0451)
Constant	8.7809***	4.4097***
	(0.0634)	(0.4099)
Observations	296	296
R-squared	0.9757	0.1133
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1		

The hypothesis that lower-income cities are inherently more competitive than higher-income cities and that this difference can explain the higher subsequent growth might be premature. As mentioned, the income growth in the first period is higher than the growth in the second period, as shown by the higher constant of column 5 in table 4. The weakening of the coefficient and R^2 from column 5 to 6 might be due to another effect. As technology spreads and the nation's average income rises, the percentage growth of lower-income cities would be due to them starting from a lower base. This would be similar to a dry harbour where all the boats are moored on dry land experiencing a rising tide. Those boats with shallow hulls will be the first to rise and they will also see a higher percentage rise compared to those boats with deeper hulls. This analogy explains why the coefficients in both columns are negative, but it also might explain why column 5's coefficient is stronger. If this mean reversion of income is entirely due to the rising tide effect is unclear. Others have argued that the diffusion of technology can explain the convergence in incomes throughout an economy (Barro and Sala-I-Martin, 1990). If low-income cities are inherently more competitive in the future, one would see the persistence of income growth over longer periods in alternative regressions. Equation 18 tests this hypothesis where the growth of household income over the 2000-2017 period is regressed on growth in income over the 1972-2000 period. The coefficient for this regression is very small and only significant at the 13% level. The R^2 is also very close to zero and the correlation between the two periods is 0.08. This is not consistent with the hypothesis that lower-income cities are inherently more competitive. Suppose lower income cities are associated with higher subsequent income growth. In that case, there should be at least a significant relationship between two periods of growth as this inherent competitiveness would not die out immediately and instead show some persistence over the two periods. To better explain the mean reversion observed in population and income seen in the previous paragraphs, one can look at the labour market composition.

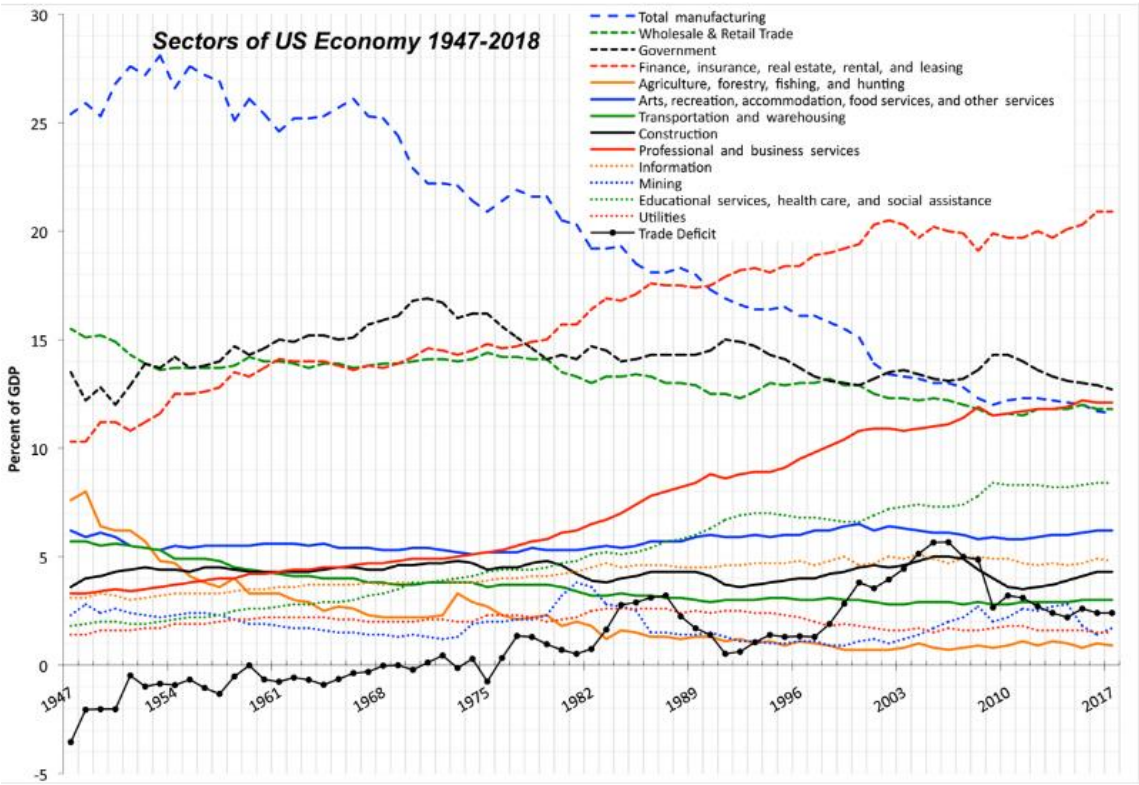
Diff. in log (Income from 2000-2017) = 0.30***+ 0.07.Diff in log (Income from 1972-2000) (18)

(0.046) (0.066)

Manufacturing:

One of the key factors influencing the population and income growth of cities from 1940 onwards seems to be the role the labour market. More specifically, how the composition of the labour force in cities determine the subsequent growth in incomes and population. One of the biggest changes that took place in the era since 1940 was the sharp decline in manufacturing employment, especially starting in the 1970's. This decline is shown by figure 9 which indicates the decline in manufacturing as a share of GDP already starting in the late sixties. The employment picture seems to follow the same trajectory. By looking at Wayne county's population growth in our own data, which encompasses a manufacturing-heavy city such as Detroit, one can see the decline taking form in the 1970's. Population growth for Wayne county from 1960-1970 was 0%, 1970-1980 was -12.3% and from 1980-1990 was -9.7%.

Figure 9: Percentage share of GDP for several sectors of the U.S. Economy.



(Source: Kossik, 2018)

Using:

Gross Domestic Product by Industry: Fourth Quarter and Annual 2017.

Gross-Domestic-Product-(GDP)-by-Industry BEA Data, Value Added, 1947-2015: up to 71 Industries.

United States Current Account to GDP 1980-2017, Trading Economics.

MERCHANDISE IMPORTS, EXPORTS, AND TRADE BALANCE: 1790-2006 (percent of GDP).

In contrast to this decline is the growth in the service economy. As indicated by figure 9 this surge in growth was higher for those service sectors that incorporated finance, insurance, real estate and rentals and business and professional services. The share of educational services and healthcare also saw significant growth, albeit at a lower rate. Note that the three above mentioned components of the services sector do not start their respective growth paths exactly when manufacturing starts declining. In other words, the decline of manufacturing is not necessarily linked to the immediate substitution of the service sector. Glaeser, Ponzetto and Tobio (2014) notes that in the first half of the 20th century one of the largest manufacturing regions in the United States was not in the Rust Belt but in fact New York City where there was an enormous garment industry. Like the Rust Belt, New York City saw a decline in manufacturing employment due to the decline in the local garment industry. This decline was not however immediately replaced by the finance industry one associates with New York, but instead there was prolonged stagnation and somewhat decline in the population of the city. From 1950-1960 population growth was -1.4%, 1960-1970 was +1.5%, 1970-1980 -10.4% and 1980-1990 +3.5%. New York City could only attain it's 1950 population by the year 2000 (Glaeser, Ponzetto and Tobio, 2014). The authors also note that the 1970's was especially hard for New York when there was the largest exodus before the years of the financial boom. Figure 9 thus highlights the important role the composition of the economy plays in determining the growth of a given city's economy. Thus, one can say that the economic composition of a city will affect the future growth trajectory of said city given its specific ability to adapt to technological change. For this reason, the economic composition of a city needs to be taken into account as an important variable in any econometric analysis.

By looking at initial employment levels of the manufacturing and service sectors in table 5 it is clear that higher initial levels of manufacturing and service employment are associated with lower subsequent growth. It would however be a mistake to conclude that high initial levels are in themselves detrimental for later growth in the sector. Even though the coefficients are very similar, the mechanics behind it might be totally different. In the case of manufacturing, high initial levels in column one of table 5 are associated with lower subsequent growth in manufacturing due to high manufacturing cities having a larger share to lose as the national economy transitions away from manufacturing overall. However, column 2 can be the opposite; where those cities that had the lowest initial levels of services employment had the most to gain as the economy transitions into a services lead economy. Both arguments can theoretically explain why both coefficients are similar even though the reasons behind them might be different.

Table 5: Growth in services and manufacturing share (OLS).

VARIABLES	Growth in (1972-2017):	
	Prof. serv. share	Manufacturing share
Prof. Serv. share in 1972	-0.719***	
	(0.017)	
Manufacturing share in 1972		-0.716***
		(0.016)
Constant	0.464***	0.042***
	(0.009)	(0.004)
Observations	311	311
R-squared	0.848	0.871
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1		

Table 6 shows that there are indeed different factors at play and that table 5 has insufficient explaining power when looking only at initial levels of employment share. Table 6 shows both the growth in income and population as dependent variables, regressed on the initial levels and growth in manufacturing and the professional services part of the economy. The most significant observation of table 6 is the contrast between the services share and the manufacturing's share on growth in income and population. From columns 1 and 3 one can see that services are associated with higher income and population growth. Columns 2 and 4 shows the opposite picture. In this case, initial manufacturing levels and the growth in manufacturing have a negative relationship with growth in income and population. One must also note the larger R^2's of manufacturing in columns 2 and 4 which might indicate that manufacturing heavy cities saw stronger declines in income and population versus cities with little manufacturing. Table 6 does however have multicollinearity in the columns which one would expect following from the observations of the above table 5. This does have a weakening impact on the significance levels of the coefficients but does not necessarily lessen the contrast between the manufacturing and services sector's role on income and population growth rates as illustrated by table 6.

Table 6: Growth in income and population (OLS).

VARIABLES	Growth in (1972-2017):			
	Log of income	Log of income	Log of population	Log of population
Prof. Serv. share in 1972	0.657***		1.396***	
	(0.144)		(0.424)	
Manufacturing share in 1972		-1.040***		-2.743***
		(0.170)		(0.465)
Growth in prof. serv. share from 1972-2017	0.759***		0.646 (1.)	
	(0.185)		(0.543)	
Growth in manufacturing share from 1972-2017		-0.764***		-0.960 (2.)
		(0.222)		(0.607)
Constant	1.427***	1.967***	-0.234	0.987***
	(0.091)	(0.017)	(0.266)	(0.047)
Observations	311	311	311	311
R-squared	0.063	0.199	0.097	0.333

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
1.Significant at 23.5% level
2.Significant at 11.5% level

Education:

Given the positive relationship between the service sector and income and population growth, one has to consider the role education plays in the growth of the service sector. Studies such as Glaeser, Ponzetto and Tobio (2014) have shown that cities with higher educational attainment outperform those cities with lower educational attainment. This does make sense in theory as table 6 has shown that service sector led cities outperform others in income and population. If it is the case that that one of the requirements for a service sector led economy is a better educated workforce, then one would expect some correlation between the services share part of the labour force and the educational attainment. There does however remain an unanswered question on whether highly educated populations lead to income and population growth-where they would catalyse job creation. Or do highly educated populations follow initial job creation and merely amalgamate around established sectors?

What is clear is that educational levels have risen over the years following 1940. To better understand this growth in education levels, one can look at table 7. This table has the growth in the share of the workforce with bachelor's

degrees and higher for different periods, as the dependent variables and the initial share of the population with bachelor's degrees and higher as the independent variables. At first glance, one can see the strong positive relationship between initial education levels and subsequent growth. This shows that educated populations tend to amalgamate around cities where the population has already educated populations. One must also observe that even though the coefficient in column 1 is double that of column 2, the constant of column 2 is ten times larger than that of column 1. This is due to a more educated population in 1972. This difference between columns 1 and 2 of table 7 might be similar to the rising tide analogy of income reversion in table 4. This might lead to a weakening of the coefficient of column 2 in table 7, and consequently, it is not clear if educational levels had less of a role to play from 1972 onwards. To better answer this question, one will need to incorporate more variables in the regression-which is done at the end.

Table 7: Growth in bachelor's degree share given initial levels.

	Growth in bachelor's share	
VARIABLES	1940-1972	1972-2017
Bachelor share in 1940	1.027***	
	(0.1441)	
Bachelor share in 1972		0.510***
		(0.0649)
Constant	0.01***	0.121***
	(0.0069)	(0.0073)
Observations	296	296
R-squared	0.143	0.166
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1		

Fixed effects regression:

The fixed effects regression of equation 11 is compared to the normal OLS regression that incorporates the same control variables. The important distinction to note is that the OLS regression results are very much exposed to endogeneity as previously explained. This is why we will compare the results of the fixed effects regression with that of the normal OLS regression. This is done in table 8 where the dependent variables are the growth rates in population and the median household income over a 5-year period.

Table 8:OLS and fixed-effects regressions of population and income growth from 1972-2017.

	Growth rate of (D.log):			
VARIABLES:	Column 1. Population	Column 2. Income	Column 3. Population	Column 4. Income
Log of income in the previous 5 years	-0.055***	-0.267***	0.001	-0.322***
	(0.003)	(0.002)	(0.002)	(0.003)
Log of population in the previous 5 years	0.001	0.009***	-0.156***	0.022***
	(0.001)	(0.001)	(0.003)	(0.006)
Log of bachelor's or higher share in the previous 5 years	-0.012**	0.107***	0.036***	0.154***
	(0.006)	(0.005)	(0.003)	(0.007)
Log of the share of the workforce in manufacturing in the previous 5 years	-0.005***	0.002**	-0.000	0.003***
	(0.001)	(0.001)	(0.000)	(0.001)
Log of the share of the workforce in professional services in the previous 5 years	0.090***	-0.128***	-0.050***	-0.115***
	(0.010)	(0.009)	(0.007)	(0.014)
Log of homeownership rate in the previous 5 years	0.087***	0.120***	0.100***	0.159***
	(0.010)	(0.010)	(0.018)	(0.034)
Log of the house price to replacement cost ratio in the previous 5 years	0.060***	0.059***	-0.014***	0.143***
	(0.004)	(0.004)	(0.004)	(0.008)
Constant	0.269***	2.458***	1.645***	2.791***
	(0.048)	(0.045)	(0.087)	(0.166)
Observations	2,516	2,516	2,516	2,516
R-squared	0.325	0.893	0.749	0.946
Number of metropolitan areas	281	281	281	281
Regression type	OLS	OLS	City-fixed effects	City-fixed effects
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				

Columns 1 and 2 show the OLS coefficients and columns 3 and 4 have all the same variables but now it includes the city-fixed effects. By looking at the coefficients of 1 and 2 most of the variables seem to follow from the economic intuition previously discussed. Income and population growth seem to be lower if the incomes are

already high in the previous period and cities with higher initial educational attainment levels are associated with higher income growth. The initial population levels are only significant for income growth in column 2 but should be interpreted with caution as the coefficient is very small. The coefficients for manufacturing share and professional services share should also be interpreted with caution, although there seems to be high significance for professional services compared to the manufacturing share. We can also see that the coefficient for the percentage of homeownership in a city is also significant and positive for explaining income and population growth and this observation seems to persist for all the future regressions as well. Indicating that there is clear evidence that high homeownership rates are beneficial for income growth and population growth. Columns 3 and 4 show that a 1% increase in the homeownership rate increases population and income growth with 10% and 16% respectively. The most important coefficient in columns 1 and 2 is that of the house price to replacement cost ratio which in this case is positive. This is counter to the economic intuition where liberal zoning laws and affordable housing would be beneficial for population and income growth. This observation in the OLS regressions of columns 1 and 2 necessitates the move towards a fixed-effects regression as highlighted in the empirical strategy as there is too much risk of endogeneity in the OLS results of column 1 and 2. What the positive coefficient for the ratio might indicate is that successful cities (with their high ratios) have high income and population growth, but we already know this and this does not properly answer the structural change in affordability issue. This is why a city fixed-effects regression is needed.

For columns 3 and 4 the coefficients are more in line with what one would expect from the theoretical model where stricter zoning laws would lead to lower population growth due to the inelasticity of housing supply. Following from the theoretical model one can now see that as expected when dr/dN is high in a city with restrictive zoning laws it will lead to a small change in population ΔN if utility changes Δu are constant. Column 3 shows that a 10% increase in the house price to replacement cost ratio is associated with 14% decrease in the population growth rate. This is the major observation of column 3 which is complemented by the other control variables. From these control variables, one can now see that higher levels of education and homeownership seem to benefit population growth but higher initial levels of population and higher levels of the workforce in professional services saw lower population growth, which does make sense if one thinks of the lower population growth seen in the past where cities like San Francisco, New York and Boston all have high populations and a higher share of workers in the professional services although moving there is prohibitively expensive. The other variables of manufacturing share and initial income levels do not seem to be significant in explaining cities' population growth. However, one must mention that income would probably be significant if one were to use the incomes adjusted for each city's living cost, which some authors have successfully done although there is still some debate regarding the importance of this variable to explain migration between cities (Kemeny and Storper, 2012).

Column 4 shows the fixed effects regression for income growth and although the role of income growth as an explanatory variable for population growth might be debatable it does seem as if the explanatory variables that explain income growth itself seems to be all very significant and delivers a high R^2. Similar to the OLS regression, one sees that income growth is lower in cities with initially high incomes and higher when educational attainment is high. The results also indicate that cities with high initial level of workers in the professional service sector

leads to lower income growth and that manufacturing is significant, but the coefficient is too small to interpret. This falls in line with the previous discussion of professional services' share. One of the reasons behind the negative coefficient is the economy transitioning into a services-led economy. In other words, low professional service share cities had the most to gain in incomes when the economy transitioned into the services sector and cities with initially high levels had the least to gain. In order to fully understand the role of the composition of the labour force one will need to do a more detailed analysis beyond the scope of this study due to the multitude of relationships between labour force composition and economic growth. For one, the professional services sector has seen some of the highest employment and wage growth, leading one to expect a positive coefficient.

One clear observation is the beneficial effects of cities with a higher educational attainment. By expanding column 4 to include an interaction term that combines educational attainment and income, one can see that higher levels of education seems to deliver better income growth when it is placed in an already high income, high productivity city. This is shown in column 6. This observation corresponds to the results of (Jara-Figueroa *et al.*, 2018) which showed the benefits of highly skilled workers being in close economic proximity to one another delivering large positive spill-over effects.

Table 9: Growth in income from 1972-2017.

	Growth rate of (D.log):	
VARIABLES	Column 5 Income	Column 6 Income
Log of income in the previous 5 years	-0.479***	-0.267***
	(0.010)	-0.009
Log of population in the previous 5 years	0.023***	0.021***
	(0.006)	-0.006
Log of bachelor's or higher share in the previous 5 years	0.085***	-0.108***
	(0.007)	-0.039
Log of the share of the workforce in manufacturing in the previous 5 years	0.003***	0.003***
	(0.001)	-0.001
Log of the share of the workforce in professional services in the previous 5 years	-0.070***	-0.052***
	(0.013)	-0.016
Log of homeownership rate in the previous 5 years	0.295***	0.149***
	(0.033)	-0.034
Log of the house price to replacement cost ratio in the previous 5 years	-0.047***	0.136***
	(0.013)	-0.008
Log median house value in the previous 5 years	0.218***	
	(0.013)	
Interaction term for bachelor's share and income		0.023***

		-0.003
Constant	2.683***	2.260***
	-0.162	-0.182
Observations	2,516	2,516
R-squared	0.949	0.947
Number of metropolitan areas	281	281
Regression type	City fixed effects	City fixed effects
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1		

The positive coefficient of the house price to replacement cost in column 4 also need to be inspected further. To what extent do restrictive zoning laws really benefit income growth. If one expands column 4 to include the homeownership rate in column 5 the picture does seem to change dramatically. We already know that high rates of homeownership are associated with higher income growth. But what if restrictive zoning laws drive up the prices of the real estate held by those homeowners? Column 5 shows a different picture with the inclusion of the median house price as an explanatory variable. Now one can see that the coefficient on the replacement cost ratio is actually negative and that the coefficient on house price is positive, as well as the homeownership rate. The coefficients in column 5 show that a 10% increase in the house price to replacement cost ratio is associated with 47% decrease in the income growth rate when we control for the capital gain experienced by homeowners. This observation might suggest that it is not restrictive zoning laws that benefit incomes but rather higher house prices together with higher homeownership rates, indicating that incomes benefit from a larger property-owning class. One must however consider that multicollinearity might be present between the house price to replacement cost variable and that of the house price variable. Because this is a fixed-effects regression we can inspect this by regressing the growth in first on the growth in the latter. From equation 19 one can see that this relationship does not seem large enough to invalidate the significance of the explanatory variables in column 6.

$$\text{Log(house price to replacement cost ratio)} = \underset{(0.711)}{-4.051^{***}} + \underset{(0.006)}{0.350^{***}}.\log(\text{Median house price}) \qquad (19)$$

$$R^2 = 0.541$$

We can now go further with the inclusion of the time-invariant explanatory variable of mean January temperature to see if it corresponds to the findings of other studies which found it to be important in explaining the higher population growth rates observed in the Southern Unites States since the 1970's due to the winters being warmer (Glaeser, Ponzetto and Tobio, 2014). This requires the fixed effects regression to be changed to a random-effects regression in table 10.

Table 10: Growth in population from 1972-2017.

	Growth rate of (D.log):	
VARIABLES	Column 7 Population	Column 8 Population
Log of income in the previous 5 years	-0.028***	-0.003
	(0.002)	(0.002)
Log of population in the previous 5 years	-0.051***	-0.143***
	(0.002)	(0.003)
Log of bachelor's or higher share in the previous 5 years	0.032***	0.036***
	(0.005)	(0.003)
Log of the share of the workforce in manufacturing in the previous 5 years	-0.000	-0.000
	(0.000)	(0.000)
Log of the share of the workforce in professional services in the previous 5 years	0.011	-0.044***
	(0.009)	(0.007)
Log of homeownership rate in the previous 5 years	0.112***	0.108***
	(0.018)	(0.018)
Log of the house price to replacement cost ratio in the previous 5 years	-0.039***	-0.019***
	(0.005)	(0.004)
Log of mean January temperature	0.003***	0.004***
	(0.000)	(0.001)
Constant	0.480***	1.328***
	(0.083)	(0.091)
Observations	2,516	2,516
R-squared	0.628	
Rho: Fraction of variance due to ui	0.805	0.989
Number of metropolitan areas	281	281
Regression type	Random effects	Hausman-Taylor regression
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1		

By comparing column 7 with the fixed effects regression of column 3 we can see that higher winter temperatures benefit population growth rates, where a 30 degrees Fahrenheit /20 Celsius increase in the winter temperature is associated with a 9% increase in the population growth rate over 5 year period. This 30-degree Fahrenheit increase in winter temperature is roughly the same as the difference between New York City and Miami. The other explanatory variables correspond roughly with the output of the fixed-effects regression. One can see that the role of manufacturing and the professional services' share still seem insignificant. As stated, the random-effects model is exposed to endogeneity due to the inclusion of the unobserved heterogeneity term u_i but this can however be

addressed with the use of a Hausman-Taylor model which is a slight adjustment on the random effects model. The Hausman-Taylor model allows one to choose those variables that one thinks is exposed to endogeneity in a random-effects model and then one can create new instrumental variables for those endogenous variables by demeaning them. For the regression output in table 10 we compare the Hausman-Taylor model with the random-effects model to see if there are added benefits. Note that the endogenous variables that were used are the bottom six rows of table 10. From column 8 one can see that the changes that follow from the Hausman-Taylor model are small. Most of the variables are only marginally altered except for winter temperature which saw a slight increase on the coefficient and the halving of the coefficient for the house price to replacement cost ratio. It is however unclear if this expansion of the Hausman-Taylor model is superior to the random effects model as the share of the error term explained by the unobserved heterogeneity term has increased. What this model however does point to is the necessity to look for another explanatory variable that can better explain the population growth rate.

The incorporation of the growth in population from the preceding 5-year period in our fixed effect regression seems to play a substantial role in explaining the growth of population. This incorporation builds on the idea that we discussed in the control variables section where we observed momentum in the growth of population in a city and it will sustain itself over long periods. From the normal OLS regressions from that section we know that population growth is highly predicted by the growth in the preceding 10 year period but we also know from Glaeser, Ponzetto and Tobio (2014) that over the very long term when the periods are expanded to 70-year periods from years 1860-2000, there is a negative relationship, indicating that population growth sustains itself but dies out over the longer period.

This idea does however allow for a regression where we can study if this capacity of a city to expand and experience momentum can be dampened by restrictive zoning laws. In other words, if we assume that that population growth in one period can stimulate growth in such a way in a city's economy that it will actually induce additional growth in the following period. This can be due to high periods of growth coinciding with a higher expenditure on investment and capital formation which will add to growth in the following period, or it can be that when growth is high the mere fact that it is above the national average is enough to sustain growth in the following period. Whichever theory one choses for why growth sustains itself over periods longer than 10 years, we do know that it is there and seems to indicate some form of economic momentum. The important question to ask is if this built-in capacity of a city to absorb this momentum of growth coming its way in the second period can be dampened by restrictive zoning laws and unaffordable housing. Table 11 attempts to answer this question by using both a fixed-effects and normal OLS regression. For both regressions, we use an interaction term for the population growth rate in the previous period and the growth rate of the house price to replacement cost ratio. Although one can normally expect OLS and fixed regressions to deliver quite different results, the output for both columns are very similar and show that past population growth highly predicts growth in the following period but there is a dampening effect if that growth was accompanied with higher growth in the restrictiveness of the zoning laws which is highlighted by the large and negative interaction term for both columns. This adds to the arguments that stricter zoning laws serve to restrict population growth. More specifically, cities will be less able to convert

the current economic booms on their doorstep into growth in the following period if unaffordable housing stands in the way.

Table 11: Growth momentum affected by unaffordable housing.

VARIABLES	Growth rate of (D.log):	
	Column 9 Population	Column 10 Population
Growth rate of population in the previous 5 years D.log	0.746***	0.858***
	(0.003)	(0.004)
Growth rate of the house price to replacement cost ratio in the previous 5 years D.log	0.033***	0.065***
	(0.003)	(0.005)
Interaction term of the population growth rate and the replacement cost ratio growth rate	-0.214***	-0.748***
	(0.021)	(0.037)
Constant	0.007***	0.001***
	(0.000)	(0.000)
Observations	2,248	2,248
R-squared	0.985	0.978
Number of metropolitan areas	281	281
Type of regression	City Fixed effects	OLS
Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1		

Conclusion:

This study sheds light on unaffordable housing levels in cities in the United States and the consequences that follow from this by studying specific indicators of competitiveness associated with expanding cities. Although the public in the United States have for years known about rising levels of unaffordability in certain cities, policymakers and politicians have been slow to address this issue due to the difficulty of understanding the true causes of unaffordability. Economist too, have only recently started to study unaffordable levels of housing and only a handful have studied the economic consequences. This study sheds light on this issue by using a spatial equilibrium model where migration between cities follow the differences in amenities between cities. This model allows for the combination of three critical variables where the first two; population and income growth serve as indicators of an expanding city and the third variable of the house price to replacement cost ratio serves as an

indicator for the restrictiveness of zoning laws leading to unaffordable house prices. Where in previous studies the house price-to-income ratio was used as an indicator of unaffordability and was closely related to the demand side of house prices, this new ratio is a novel approach to study what a lack of supply does to the growth prospects of a city.

These three variables are then combined with other control variables to study the effect on population and income growth. If the variables have a negative effect on these two dependent variables, then we conclude that it hampers the competitiveness of a city. The study comes to the following conclusions. The first is that, as the amenities model predicts, an increase in the house price to replacement cost ratio has a negative effect on population growth. If the ratio increases by 10%, then the population's growth rate over a 5-year period will be lowered by 14%. Secondly, the results initially show that an increase in the house price to replacement cost ratio benefit incomes but after controlling for the advantages of higher house prices for homeowners one finds that a 10% increase in the ratio is associated with a 47% decrease in income growth. This large number should however be cautiously interpreted as it is difficult to truly separate it from the capital growth homeowners experience but does serve as a preliminary indication of the negative effects on income growth. The study also finds that high homeownership rates are beneficial to income growth and population growth.

The study also attempted to study the effect of unaffordable house prices on cities' ability to sustain an economic expansion. If one assumes that cities sustain their population growth over a period longer than 5 years during an economic expansion, one can then observe from the results that growth in unaffordability will lessen the population growth in the following 5 years. This indicates that growth in unaffordability serves to lower the capacity of a city to sustain its population growth into future periods.

In addition to the main conclusion the study also attempted to study observations and caveats made by previous urban economics studies. In line with previous studies, this study finds higher education is worth more when people find themselves in highly educated cities and that there is some form of mean reversion for population size and income.

The evidence presented in this study is suggestive, and not definitive, but sheds valuable light on housing affordability across the United States and the consequences this holds for the growth of cities. This is done in the larger field of urban economics; a field that is rapidly expanding but is also not uniform in its approach to answering important and complex questions. Continuing this path in urban economics should be encouraged as there are valuable insights for policy recommendations, both in the United States and other geographies where a lack of housing supply plays a role in cities.

Reference list:

Agarwal, S. and Qian, W. (2017) 'Access to home equity and consumption: Evidence from a policy experiment', *Review of Economics and Statistics*. doi: 10.1162/REST_a_00606.

Asea, P. K. and Corden, W. M. (1994) 'The Balassa-Samuelson Model: an Overview', *Review of International Economics*. doi: 10.1111/j.1467-9396.1994.tb00040.x.

Barro, R. J. and Sala-I-Martin, X. (1990) 'Economic growth and convergence across the United States', *National Bureau of Economic Research*.

Behrens, K. and Robert-Nicoud, F. (2009) 'Krugman's papers in regional science: The 100 dollar bill on the sidewalk is gone and the 2008 Nobel Prize well-deserved', *Papers in Regional Science*. doi: 10.1111/j.1435-5957.2009.00241.x.

Bureau of Labor Statistics (2015) 'Chapter 17 - The Consumer Price Index', *BLS Handbook of Methods*.

Chakrabarti, R. and Zhang, J. (2015) 'Unaffordable housing and local employment growth: Evidence from California municipalities', *Urban Studies*. doi: 10.1177/0042098014532854.

Federal Reserve Bank of San Francisco. (2003) 'How is the Housing Affordability Index calculated?'

Garriga, C., Manuelli, R. and Peralta-Alva, A. (2019) 'A macroeconomic model of price swings in the housing market'. *American Economic Review*, *109*(6), p.2036-72.

Glaeser, E. and Gyourko, J. (2003) 'The Impact of Building Restrictions on Housing Affordability', *Economic Policy Review*.

Glaeser, E. and Gyourko, J. (2018) 'The economic implications of housing supply', in *Journal of Economic Perspectives*. doi: 10.1257/jep.32.1.3.

Glaeser, E. L., Kolko, J. and Saiz, A. (2001) 'Consumer city', *Journal of Economic Geography*. doi: 10.1093/jeg/1.1.27.

Glaeser, E. L., Ponzetto, G. A. M. and Tobio, K. (2014) 'Cities, Skills and Regional Change', *Regional Studies*. doi: 10.1080/00343404.2012.674637.

Graves, P. E. (1976) 'A reexamination of migration, economic opportunity and quality of life', *Journal of Regional Science*. doi: 10.1111/j.1467-9787.1976.tb00954.x.

Hsieh, C. T. and Moretti, E. (2019) 'Housing constraints and spatial misallocation', *American Economic Journal: Macroeconomics*. doi: 10.1257/MAC.20170388.

Irwin, E. G., Isserman, A. M., Kilkenny, M. and Partridge, M. D. (2010) 'Acentury Of research on rural development and regional issues', in *American Journal of Agricultural Economics*. doi: 10.1093/ajae/aaq008.

Jara-Figueroa, C., Jun, B., Glaeser, E. L. and Hidalgo, C. A. (2018) 'The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1800475115.

Katz, B., Turner, M. A., Brown, K. D., Cunningham, M. and Sawyer, N. (2003) 'Rethinking Local Affordable Housing Strategies: Lessons From 70 Years of Policy and Practice', *The Brookings Institution*.

Kemeny, T. and Storper, M. (2012) 'The sources of urban development: Wages, housing, and amenity gaps across American cities', *Journal of Regional Science*. doi: 10.1111/j.1467-9787.2011.00754.x.

Lydon, R. and O'Hanlon, N. (2011) 'Housing equity withdrawal, property bubbles and consumption', *Journal of the Statistical and Social Inquiry Society of Ireland*.

Partridge, M. D. (2010) 'The duelling models: NEG vs amenity migration in explaining US engines of growth', *Papers in Regional Science*. doi: 10.1111/j.1435-5957.2010.00315.x.

Partridge, M. D. and Rickman, D. S. (2003) 'The waxing and waning of regional economies: The chicken-egg question of jobs versus people', *Journal of Urban Economics*. doi: 10.1016/S0094-1190(02)00501-6.

Partridge, M. D. and Rickman, D. S. (2005) 'Regional cyclical asymmetries in an optimal currency area: An analysis using US state data', *Oxford Economic Papers*. doi: 10.1093/oep/gpi022.

Quigley, J. M. and Raphael, S. (2004) 'Is housing unaffordable? Why isn't it more affordable?', *Journal of Economic Perspectives*. doi: 10.1257/089533004773563494.

Roback, J. (1982) 'Wages, Rents, and the Quality of Life', *Journal of Political Economy*. doi: 10.1086/261120.

Sasser, A. C. (2010) 'Voting with their feet: Relative economic conditions and state migration patterns', *Regional Science and Urban Economics*. doi: 10.1016/j.regsciurbeco.2010.02.001.

Storper, M. and Manville, M. (2006) 'Behaviour, preferences and cities: Urban theory and urban resurgence', *Urban Studies*. doi: 10.1080/00420980600775642.

Stroebel, J. and Vavra, J. (2019) 'House prices, local demand, and retail prices', *Journal of Political Economy*. doi: 10.1086/701422.

Wardrip, K., Williams, L. and Hague, S. (2011) 'The Role of Affordable Housing in Creating Jobs and Stimulating Local Economic Development', *Insights from Housing Policy Research*.

Zhou, X. and Carroll, C. D. (2012) 'Dynamics of wealth and consumption: New and improved measures for U.S. States', *B.E. Journal of Macroeconomics*. doi: 10.1515/1935-1690.2403.