Trickle-down housing economics

Charles G. Nathanson*

September 20, 2023

Abstract

I evaluate the equilibrium effects of supplying high-quality housing on house prices and populations throughout a metro area. Between 2000 and 2019, metros that lack high-quality construction witness "trickle-down" growth in house prices throughout the quality distribution, gentrification in middle-quality housing, and homelessness. I fit an equilibrium model of the Los Angeles housing market to Census data and find a causal effect of high-quality construction on these outcomes. High-quality construction relieves house prices for the poor through equilibrium effects, but when it comes through reducing the low-quality housing stock, it can make the poor worse off.

*Kellogg School of Management, Northwestern University. Email: nathanson@kellogg.northwestern.edu. Phone: (847) 467-5141. First draft: November 15, 2018. I thank David Albouy, Jonathan Dingel, Edward Glaeser, Daniel Hartley, Therese McGuire, Timothy McQuade, and seminar/conference participants at BYU, CEPR, FRB Philadelphia, HULM, Kellogg, LSE, MIT, Notre Dame, SED, Stanford, Wisconsin, and Wharton for helpful comments and Anthony A. DeFusco for conversations that inspired this paper. All financial support for this research comes from Northwestern University.

Newly built housing units typically command higher market values than existing units, and high-income households are more likely to occupy new construction than low-income households (Rosenthal, 2014). Nonetheless, the benefits of new construction can "trickle down" to low-income households in market equilibrium (Sweeney, 1974). In a city with a fixed population, building 100 new high-quality housing units frees up 100 units for the poor by inducing households throughout the city to move into nicer housing (Braid, 1981). When the city's population is not fixed, migration attenuates this trickle-down effect as incoming households occupy some of the new housing. The extent to which migration attenuates the benefits of market-rate construction for households in lower quality units is an empirical question. In this paper, I seek to answer this question by evaluating the effect of high-end construction on the populations of low- and middle-income households in the same metro and on the house prices paid by these households.

I first bring new reduced-form facts about the effect of construction on market equilibrium in the cross-section of metros between 2000 and 2019. I develop a methodology that allows me to assign housing units in 2019 to deciles of the metro's 2000 housing quality distribution. Following Gyourko, Mayer and Sinai (2013), I explore the effect of *not* building housing by focusing on "superstar" metros—those with abnormally high house price appreciation and low housing stock growth. Relative to other metros, superstars witness less high-quality construction but a comparable amount of low-quality construction. Nonetheless, differences between the metros are not confined to the top deciles: superstars have strong house price appreciation throughout the quality distribution, and they experience a particularly strong growth in the college share (gentrification) in middle deciles. Homelessness also increases differentially in superstar metros, while the number of persons per household does not. These facts suggest that a lack of high-quality new units in superstars "trickles down" to the rest of the housing market as rich college-educated households choose medium-quality housing. This choice, in turn, drives up the prices of low- and medium-quality housing and pushes lower-income households without a college degree into low-quality housing, homelessness, or out of the metro entirely.

To evaluate the effect of construction in a causal framework, I then fit an equilibrium model of a superstar metro, Los Angeles, to data from the US Census. In the model, households who differ in income and education choose whether to live in Los Angeles, and if so, the quality of their housing unit. I pick the distributions of house prices and household income in the model to match the empirical housing market equilibrium in 2019, and I select labor demand shocks, amenity changes, and construction levels to match changes in wages, house prices, and housing stock in Los Angeles relative to non-superstar metros from 2000 to 2019. I calibrate the elasticity of migration into Los Angeles in response to changes in house prices and wages using estimates from Diamond (2016). Using the model, I explore a counterfactual in which construction in Los Angeles resembles that of an average non-superstar metro through additional high-quality housing. According to my estimates, this additional construction reduces excess house price growth in Los Angeles across the quality distribution, from an average of 15.0% to 9.4%. It also attenuates gentrification in the middle of the quality distribution. In terms of providing housing to the poor, 100 additional housing units cause 20 (resp. 33) households to move out of the bottom quintile (resp. half) of the housing market into nicer housing and induces poor households to move into the metro. Therefore, the benefits of new construction trickle down to poorer households, although migration attenuates much of this effect. I also evaluate the relative equilibrium effect of building low- versus high-quality housing. Adding one unit of low-quality housing improves the welfare of the poor 40-90% more than one additional unit of high-quality housing because it lowers the equilibrium price of low-quality housing more sharply. However, when skill-biased amenity spillovers are sufficiently strong, new low-quality units make the some rich households in the metro worse off by lowering the metro's amenities.

Quantifying the benefits of new construction for the poor is important for informing housing policy debates. Although the US government spends billions of dollars annually in housing assistance to the poor through the low-income housing tax credit and housing choice vouchers, these programs have a limited effect on the housing costs paid by poor households (Eriksen and Rosenthal, 2010; Collinson and Ganong, 2018). Direct government supply of low-income housing has well-known costs, such as crime (Aliprantis and Hartley, 2015). According to my estimates, government policy that encourages the construction of one additional high-quality unit is 50-70% as effective for the poor as a policy that supplies one additional low-quality unit. Therefore, encouraging market-rate construction may be an equally, if not even more, effective policy for addressing soaring housing costs for the poor than existing policies targeting them directly. At the same time, my results caution against urban renewal projects that tear down low-quality units in order to build high-quality units. Unless the total number of new units exceeds the count of torn-down units by at least 40%, such a project may make poor households worse off purely through equilibrium effects (that is, even ignoring moving costs).[1]

[1] According to Diamond, McQuade and Qian (2019), rent control in San Francisco induces landlords to convert older properties to newer, more expensive housing. If the number of new units does not exceed the count in the torn down structures, rent control may make the city's poor worse off through the equilibrium effects I am estimating.

The paper proceeds as follows. I review related literature, and the papers whose I estimates I use to quantify my model, in Section 1. In Section 2, I describe the data I use, and I present the reduced-form facts about the evolution of housing markets between 2000 and 2019 in in Section 3. I lay out the model in Section 4 and describe how I fit the model to the data for the Los Angeles metro area in Section 5. Section 6 contains the estimated counterfactual results of different types of construction, and I conclude in Section 7.

1 Literature review

1.1 Related papers

My empirical results add to the literature estimating the role of housing supply restrictions on the growth of house prices and populations across US metros since the 1980s (Glaeser and Gyourko, 2003; Glaeser, Gyourko and Saks, 2005; Saiz, 2010; Gyourko, Mayer and Sinai, 2013; Ganong and Shoag, 2017; Molloy, Nathanson and Paciorek, 2022). While most of this literature focuses on average effects, I take a non-parametric approach by examining changes in house prices and populations throughout the quality distribution.[2] This approach allows me to demonstrate, for instance, that the relative decline in non-college households in superstar cities is strongest in middle-quality housing. My results on vacancy and homelessness, to my knowledge, do not appear in prior work in this area.[3]

My empirical work relates to Epple, Quintero and Sieg (2020), who also develop a methodology for measuring growth in the housing stock by quality for a metro area. I innovate by using Census data to construct house price and rent distributions for all 384 CBSAs in the US. These data allow me to measure housing growth by quality level for significantly more metro areas than Epple, Quintero and Sieg (2020) analyze using the American Housing Survey. Our methodologies differ as well. While Epple, Quintero and Sieg (2020) infer quality using a utility model, I measure quality by comparing counts of housing units in consistent geographic areas over time.

I also contribute to a growing literature estimating the spillovers of new construction on housing market equilibrium in nearby areas (Schwartz et al., 2006; Baum-Snow and Marion, 2009;

[2] Gyourko, Mayer and Sinai (2013) show that superstar metros experience comparably high house price growth at both the mean and tenth percentile of the house value distribution. I extend this result across the other nine deciles and by using a housing cost measure that combines prices and rents.

[3] The result that homelessness increases more in metros with faster house price growth is consistent with evidence showing that cash assistance to the poor decreases homelessness (Evans, Sullivan and Wallskog, 2016; Locks and Thuilliez, 2023).

Diamond and McQuade, 2019; Damiano and Frenier, 2020; Pennington, 2021; Li, 2022; Anagol, Ferreira and Rexer, 2023; Asquith, Mast and Reed, 2023). A related literature looks at house price effects of demolitions (Almagro, Chyn and Stuart, 2022; Blanco, 2023). In contrast to much of this literature's focus on the surrounding neighborhood, my paper examines effects on the entire metro area, with my empirical results providing a reduced-form guess of these effects. Mense (2023), who exploits rainfall as an instrument, likewise finds that new construction lowers rents throughout a metro area.

Early theoretical work on the metro effects of different types of housing supply focuses on qualitative insights (Sweeney, 1974; Braid, 1981). Recent papers using quantitative models find only small effects of increasing housing supply on house prices (Anenberg and Kung, 2020), the welfare of city residents (Favilukis, Mabille and Van Nieuwerburgh, 2023), and gentrification (Couture et al., 2023). I find large effects of expanding the supply of high-quality housing on all three outcomes. Unlike these other papers, I link the quality distribution of new construction in my model to the data.[4] I discuss the quantitative differences with respect to these papers in more detail in Section 6. My paper complements work by Almagro, Chyn and Stuart (2022), Anagol, Ferreira and Rexer (2023), Mast (2023), and Mense (2023), who concentrate as I do on quantifying the effects of changes in the housing stock on house prices throughout a metro area. Relative to that work, my paper is unique in calibrating the effect of cross-metro migration using previous results in the literature on migration elasticities.

My theory adopts the techniques of assignment models, in which differences in the attributes of goods collapse into a single quality index. Recent work applies these models to housing markets (Määttänen and Terviö, 2014; Landvoigt, Piazzesi and Schneider, 2015; Epple, Quintero and Sieg, 2020). I extend this literature by nesting a housing assignment model in spatial equilibrium, which allows me to evaluate how cross-metro migration attenuates the trickle-down benefits of new construction for the poor. In this respect, my paper is similar to Davis and Dingel (2020), but they study the sorting of households and firms across metros while I analyze the equilibrium effects of housing supply. Määttänen and Terviö (2014) and Landvoigt, Piazzesi and Schneider (2015) also do not examine the effect of new construction, focusing instead on changes in the income distribution and credit constraints, respectively.

[4]In my model, I abstract from important properties present in the models of these other papers: Anenberg and Kung (2020) estimate a rich model of neighborhood choice, Favilukis, Mabille and Van Nieuwerburgh (2023) nest a stochastic lifecycle income problem inside a model of New York City, and Couture et al. (2023) feature non-homothetic preferences for urban amenities that generate gentrification in response to income shocks. The absence of these model attributes from my paper may also explain why those papers find smaller effects of construction than I do.

1.2 Empirical results used in the model

To calibrate household preferences, I draw on existing work by Diamond (2016) in which utility depends on consumption of a national good, consumption of a local good that includes housing, metro amenities, and an idiosyncratic term representing a household's desire to live in a particular metro area. She estimates the parameters of this utility function separately for households with and without a college degree using data on US metros between 1980 and 2000. These parameters deliver the propensity of households to move into a metro in response to changes in that metro's wages, housing costs, or amenities. Relative to her work, the key innovation I make is to model housing as an indivisible good available at different quality levels, instead of as a homogeneous good that can be consumed in arbitrary amounts. This innovation allows me to describe the equilibrium consequences of construction for different quality segments of the housing market.[5] Another change is modeling income as continuously distributed among metro households, allowing me to describe equilibrium effects of construction separately for poor versus rich. In Diamond (2016), metro households with the same education all have the same income.

My model features a local labor market, which I calibrate using consensus results from work in labor economics. Following the literature on the returns to education (Goldin and Katz, 2008; Card, 2009), I adopt a constant elasticity of substitution production function over college and non-college labor. I allow labor productivity to increase in the metro's population, consistent with work in urban economics estimating agglomeration effects of metro density onto productivity (Combes and Gobillon, 2015). Productivity also rises in the share of the metro's population with a college degree, consistent with Lucas (1988), Moretti (2004*a*,*b*), and Gennaioli et al. (2013).

In an extension to my baseline results, I allow the level of metro amenities to depend on the ratio of college to non-college households. The specification follows Diamond (2016), who estimates that amenities such as school quality and lack of crime are larger in metros with a greater share of college households. Bayer, Ferreira and McMillan (2007) similarly find an increased willingness to pay for housing in a neighborhood with a greater share of college-educated residents. In this extension, I find that constructing low-quality housing can make high-income, college-educated households worse off. A negative spillover of construction onto the welfare of metro residents is novel relative to the classic papers on this topic (Sweeney, 1974; Braid, 1981).

[5] This change means I cannot plug her estimated parameters directly into my model. I discuss how I use her parameters in more detail in Section 5 and Online Appendix C.

2 Data

The data are 5% samples of the US that come from the US Census Bureau. I use data from the long-form census for 2000 and from the 2015–2019 American Community Survey for 2019. In this section, I outline how I use these data; summary statistics and additional detail on the data build appear in Online Appendix A.

The data are available in two formats. In the first format, the Census Bureau aggregates data to the level of a block group, which is akin to a neighborhood and contains approximately 1,000 people. I access these tables through the National Historical Geographic Information System, or NHGIS (Manson et al., 2022). For each block group, b, I observe the number of housing units by the year the units was built, y (e.g., 1950–1959), and the form of occupancy, o (owner-occupied, renter-occupied, or vacant). I denote this count by n_{byot}^{age}, where $t \in \{2000, 2019\}$ is the year of the Census survey. I also observe housing unit counts by housing cost, j, which is house value for owner-occupied units (e.g., \$100,000–\$124,999) and gross monthly rent for renter-occupied units (e.g., \$1,000–\$1,249). I denote these counts n_{bjot}^{cost}. Finally, I observe the counts of persons in housing units and not in housing units, i.e., in group quarters.

In the second data format, the Census Bureau releases the micro data from their surveys. Data are available for each person, each household, and each housing unit. The finest geographic identifier available is a public use microdata area (PUMA), a region containing at least 100,000 people. I access these microdata through the Integrated Public Use Microdata Series, or IPUMS (Ruggles et al., 2023). I use data on the year each housing unit was built, the cost of each occupied unit, whether the head of the household has a college degree, and household income. For persons in group quarters, I use data on income, age, employment status, and school attendance.

When analyzing metro areas, I adopt the 2018 definitions of core based statistical areas (CBSAs). I map block groups to CBSAs using crosswalk files from the US Census Bureau and NHGIS. Hence, each metro represents the same physical region whether I analyze it in 2000 or 2019. I also employ crosswalks from the US Census Bureau, NHGIS, and the Missouri Census Data Center to map block groups to PUMAs in the same year and 2000 block groups to 2019 tracts, which are areas of about 4,000 people that nest block groups.

3 Motivating facts

3.1 Methodology for assigning deciles

Following the literature on housing assignment models cited in Section 1, I assume the existence of a unidimensional quality index for housing in each CBSA. If housing unit A is of higher quality than housing unit B, then all households prefer to live in A over B. The quality index captures physical attributes, such as structure and lot size, as well as locational attributes, such as proximity to parks and the frequency of neighborhood crime. As shown in the previous literature, housing units of higher quality command higher prices in market equilibrium. Therefore, to measure the quantity of housing in each quality decile, it suffices to sort housing units by their market prices and then form deciles. I form these deciles in three steps.

In the first step, I impute the joint distribution of housing cost and year built within each block group. In the IPUMS data, I observe this joint distribution of cost and year built at the PUMA level, and I observe the marginal distributions of each variable at the block-group level in NHGIS. To impute the joint distribution at the block-group level, I begin by assuming that cost and year built are independent within each block group, and I then make a minimal perturbation so that the implied joint distribution at the PUMA level matches the data. Formally, for each PUMA, occupancy o, and survey year t, the imputed counts $\widehat{n}_{bjyot}$ for block groups b in that PUMA satisfy:

$$\{\widehat{n}_{bjyot}\} \in \arg\min_{n_{bjyot}} \sum_{b,j,y} \left| n_{bjyot} - n^{ind}_{bjyot} \right|$$

subject to the constraints

$$\begin{aligned} \sum_{j} n_{bjyot} &= n^{age}_{byot}, \quad \text{for each } (b,y) \text{ pair} \\ \sum_{y} n_{bjyot} &= n^{cost}_{bjot}, \quad \text{for each } (b,j) \text{ pair} \\ \sum_{b} n_{bjyot} &= n^{puma}_{jyot} \quad \text{for each } (j,y) \text{ pair,} \end{aligned}$$

where n^{ind}_{bjyot} is the count when cost and year built are independent within block group b and n^{puma}_{jyot} is the joint count at the PUMA level.[6]

[6]The marginal distributions for cost and year built at the PUMA level are close, but not exactly equal, when I compute them in IPUMS rather than NHGIS. For the imputation procedure to find a solution, these marginals must coincide exactly. Therefore, I make a minimal perturbation to the joint distribution of cost and year built in IPUMS so

In the second step, I annualize housing costs observed in NHGIS in order to rank both owner-occupied and renter-occupied units. I annualize rents by multiplying by 12, and I annualize home values by multiplying by a user cost specific to each PUMA and survey year. I choose this user cost so that, conditional on household characteristics, annual housing costs are equal on average for owner-occupiers and renters. To calculate this user cost, I define a variable p_i in the IPUMS data equal to home values for owner-occupants and annual rents for renters. For each PUMA and year, I then estimate the user cost in a GMM specification with the following moment conditions:

$$0 = E\left(\text{ownerocc}_i(\phi p_i - \beta X_i)\right)$$
$$0 = E\left((1 - \text{ownerocc}_i)(p_i - \beta X_i)\right),$$

where X_i consists of a dummy for college degree interacted with deciles of the income distribution in that PUMA and year. The estimated value of ϕ is the user cost I use to annualize home values in the NHGIS data for block groups in the corresponding PUMA and survey year. I denote the annualized cost by p_{bjot}.[7]

In the final step, I assign the imputed housing units for each block group, cost, and year built to housing cost deciles by CBSA. In each survey year, I pick cost cutoffs so that each decile in the CBSA contains the same number of housing units built before 2000. Due to teardowns, the number of such units in 2019 is less than the number of such units in 2000. To make the deciles comparable across survey years, I therefore reweight the units in 2019 to match the distribution by year built in the same Census tract in 2000.[8] I then pick cost cutoffs for 2019 so that the number of reweighted housing units are the same in each decile for a given CBSA. Given the way the weights are defined, this number also equals the count of housing units in each decile in the same CBSA in 2000.

Two advantages of this methodology are worth noting. First, it assigns all housing units—owner-occupied, renter-occupied, and vacant—to different deciles. Given the gradual transition from owner-occupied to renter-occupied that occurs on average as housing units age (Rosenthal,

that the implied marginal distributions match those from NHGIS, and I use this perturbed joint distribution as n_{jyot}^{puma}.

[7] In the NHGIS data, I do not observe housing costs for rented single-family homes on 10 or more acres in 2000 and for vacant units and rented units without cash rent in 2000 and 2019. To calculate housing costs for the entire housing stock, I assign housing costs to these units in NHGIS using the distribution of housing costs for units in the same block group that are built in the same year and, in the case of rented units, are also renter-occupied. If there are no such other units, I use the distribution of housing costs for the same year built in that CBSA.

[8] Matching this distribution exactly is not always possible, as 0.5% of housing units in CBSAs in 2000 do not correspond to a unit in 2019 with the same tract and year built. To account for these units, I increase the weights of units with the same year built in other tracts within the same CBSA.

2020), coverage of all housing units is important for accurately measuring changes in the housing stock over time within quality deciles. Second, the methodology produces separate measures of new construction and teardowns as contributors to changes in the housing stock. Teardowns are measured at a fine geographic level of a Census tract using information about the decade at which housing units are built. One limitation is that I assume that, conditional on tract and year built, the quality of housing that is torn down is the same as that which remains. Another limitation is abstracting from depreciation and renovations occurring between the two survey years.[9] Despite these drawbacks, my methodology provides novel evidence on the evolution of metro housing markets by quality deciles from 2000 to 2019.

3.2 Defining superstar metros

To present motivating facts on the effects of not building housing in expensive metros, I focus on "superstar" metros. As defined by Gyourko, Mayer and Sinai (2013), these are metros that witness strong housing demand growth over several years but in which relatively few new housing units are built, presumably due to geographic or regulatory constraints on housing supply. To define superstar metros in the data, I adopt the definition in Gyourko, Mayer and Sinai (2013). In particular, I measure the annualized growth rates of the housing stock and of the inflation-adjusted mean owner-occupied housing value between 2000 and 2019. A metro is a superstar if the sum of these growth rates exceeds the median among CBSAs and the ratio of the price to the quantity growth rate falls in the top decile of all CBSAs.

Before analyzing the differential evolution of the housing market in superstar metros by quality decile, I first present summary statistics at the CBSA level by superstar status. The results appear in Table 1. Between 2000 and 2019, there are 15 superstar metros; they comprise 16% of total housing stock in CBSAs in 2000. Cumulative housing stock growth is 12 p.p. lower, and real mean home value appreciation is 33 p.p. larger, in superstar metros. The growth in the housing cost measure defined above, which covers both renter- and owner-occupied units, is likewise more than twice as high in superstar than non-superstar metros. Income growth is also higher in superstar metros for both college and non-college households. Finally, there is a sizable decline in the number of non-college households in superstar metros relative to non-superstars, while the

[9]Abstracting from net depreciation affects my results to the extent that it differs in superstar and non-superstar metros. It is difficult to know where net depreciation is higher. Factors constraining housing supply in superstar metros could decrease renovations. Alternatively, these constraints could incentivize renovations by discouraging new construction exclusively. When interpreting the results, I abstract from distinctions in net depreciation by superstar status.

growth in college households is similar across both groups of CBSAs.

3.3 Results

In Figure 1, I report the average change in the housing stock by decile across CBSAs from 2000 to 2019. I measure new units as those built on or after 2000, and I calculate teardowns using the change in the number of units built before 2000 between the two survey years. Teardowns represent a significant component of housing supply, equal to 14.2% of net housing unit growth. This ratio is significantly higher in the bottom deciles (82% in decile 1) than the top (5-6% in deciles 8-10), so it is important to account for teardowns as I measure housing supply separately across the quality distribution. Net housing stock growth is stronger in top deciles than the bottom, consistent with previous findings that new construction commands higher prices and is larger than existing housing (Rosenthal, 2014; Molloy, Nathanson and Paciorek, 2022). However, there is still substantial housing stock growth of about 7-9% even in the lowest deciles.

I report the marginal effect of superstar status on prices and quantities for each decile in Figure 2. As shown in Panel A, the lack of new housing in superstar metros is concentrated in the top quality deciles. In particular, about 50% of the diminished housing supply in superstars comes just from deciles 9 and 10, while housing stock growth is nearly the same between superstars and non-superstars in the bottom three deciles. In Panel A of Figure A1 of the online appendix, I normalize these estimates by the average housing stock growth in each decile and find that normalized growth in deciles 4-10 is uniformly lower in superstars than non-superstars while roughly the same in deciles 1 and 2. Therefore, the quality tilt of missing supply in superstars is due both to the general tendency of construction to have high quality and to the equality of low-quality housing supply between the two groups of metros.

Given that the housing missing from superstars is predominantly high quality, comparing superstars to non-superstars gives a reduced-form guess of the equilibrium effects of *not* supplying high-quality housing. In the remaining panels of Figure 2, I report these reduced-form effects on house prices and the populations of households with and without a college degree across the quality distribution.

As shown in Panel B, housing cost growth in superstars exceeds that in non-superstars across the quality distribution. In particular, real housing cost growth in the lowest two deciles is 16-17 p.p. higher in superstars, even though the supply of such housing is similar in both groups of metros. This result suggests that housing decile submarkets are not fully segmented, and that the lack of high-end construction in superstars trickles down to low-quality prices. This trickle-down mechanism can occur when high-income households choose lower quality housing due to the lack of high-quality construction, and these choices raise the prices of lower quality units.

In Panels C and D, I provide evidence consistent with this trickle-down mechanism by decomposing the population change by education. College households are richer and experience faster income growth than non-college households from 2000 to 2019 (see Table 1), so examining trends by education can reveal whether richer households displace poorer households in the lower segments of the housing market in superstar metros. I impute the college degree status of households in NHGIS using the share of households with a college degree in the same PUMA and housing cost bin in IPUMS. To abstract from initial differences in college shares across deciles, I normalize the difference in college (resp. non-college) households between 2000 and 2019 by the initial count of occupied units in each decile in 2000.

As shown in these panels, college population growth in superstar metros is lower in the top four deciles but not in the bottom six, suggesting that some college households who do not get to live in high-end housing switch into lower quality housing. In contrast, superstar metros witness a net loss of non-college households in deciles 2–9, suggesting that the influx of college households into low-quality housing pushes non-college households out of superstar metros. The combined effect of these forces is a strong relative growth of college vs. non-college households in the middle of the quality distribution. In Panel B of Figure A1 of the online appendix, I show that the relative growth of college households in superstar metros is strongest in deciles 4-6 and 8. Therefore, the lack of high-quality housing in superstar metros seems, in reduced form, to lead to gentrification in the middle of the quality distribution.

An alternative effect of the lack of new housing in superstar metros is an increase in the crowding of people into housing units. To examine crowding, I measure average persons per unit and occupancy rates at the CBSA level and regress the logs of these variables on a 2019 dummy, a superstar dummy, and their interaction. As I show in columns 1 and 2 of Table 2, there is no differential increase in persons per unit or occupancy rates in superstar metros between 2000 and 2019, although superstars have higher initial rates of persons per unit and occupancy in 2000. Therefore, the dominant margins of adjustment for persons in a superstar metro seem to be moving into lower quality housing units or moving out of the metro entirely.

In column 3 of of Table 2, I provide suggestive evidence that homelessness increases differentially in superstar metros. I cannot measure homelessness directly in IPUMS, but I can capture persons living in homeless shelters and adult group homes by selecting adults in group quarters who are non-institutionalized, non-employed, and not students.[10] The homeless rate, which I define as the number of homeless divided by the number of housing units, grows 25% more in superstar than non-superstar metros between 2000 and 2019. This finding is consistent with the trickle-down mechanism, as it suggests that some households who are priced out of low-quality housing enter homeless shelters instead of leaving superstar metros. Column 3 of Table 2 implies a net decline in homelessness in CBSAs between 2000 and 2019, which is consistent with the drop in point-in-time counts of the homeless between 2007 and 2019 (Meyer, Wyse and Corinth, 2023).

These results suggest a mechanism in which a lack of high-quality housing in superstar metros leads rich households to occupy lower quality housing, in turn increasing house prices and driving low-income households out of the metro or into lower quality housing and homelessness. The purpose of the model is to quantify this channel by isolating the effect of construction, as opposed to changes in productivity and amenities, in driving the differential trends in a superstar metro relative to other areas in the US.

4 The model

In this section, I present a static model of a metro area with different qualities of housing and households who differ by income and education.

4.1 Environment and preferences

There are $J+1$ distinct qualities of housing, q_j, where $j \in \{0,...,J\}$ and q_j increases in j. The lowest quality, $q_0 > 0$, corresponds to a homeless shelter, while $q_1,...,q_J$ represent standard housing units. The measure of housing of quality q_j in the metro is $h_j > 0$ and its market price is p_j.

There are two types of agents: households and rentiers. Rentiers are endowed with the metro's housing stock and have utility that is a linear function of a composite tradeable non-housing consumption good c_T, whose price I normalize to 1. They take house prices as given and choose how much housing to sell and how much to consume subject to a budget constraint.

[10] In the ACS, non-institutionalized group quarters include college dormitories, military facilities, workers' living quarters, homeless shelters, religious group quarters, adult group homes, and adult residential treatment centers (U.S. Census Bureau, 2012, 2014). Excluding students and the employed seems likely to restrict to the final four categories. I calculate 792,321 homeless persons in 2019 under this definition, which is comparable to the count of 600,000 in Meyer, Wyse and Corinth (2023).

Households differ in their education, $e \in \{L, H\}$, with L standing for non-college- and H representing college-educated households. Households also differ in the amount of labor, z, that they can supply to the labor market. They decide whether to live in the metro or in some external location that I refer to as the reservation locale. I denote the total measure of households of education e and labor endowment z by $\widetilde{n}_e(z)$; this distribution is atomless and has full support on $(0, \infty)$. I denote the measure of households choosing the metro by N, and I let N_e and Z_e represent the measure and labor endowment, respectively, of households of education e in the metro.

A representative firm in the metro combines non-college and college labor to produce the tradeable good, c_T, according to the production function

$$F(Z_L, Z_H) = ((A_L Z_L)^\rho + (A_H Z_H)^\rho)^{\frac{1}{\rho}}, \tag{1}$$

where $0 < \rho \leq 1$. The firm maximizes profits by setting wages to the marginal products of labor:

$$w_e = ((A_L Z_L)^\rho + (A_H Z_H)^\rho)^{\frac{1}{\rho}-1} A_e^\rho Z_e^{\rho-1} \tag{2}$$

for each e. The productivity shifters for non-college and college labor, A_L and A_H, depend on an exogenous factor as well as characteristics of the the metro's population:

$$A_e = \widetilde{A}_e N^{\gamma_N} \left(\frac{N_H}{N}\right)^{\gamma_H}, \tag{3}$$

where $\gamma_N \geq 0$, $\gamma_H \geq 0$, and $\widetilde{A}_e > 0$. This specification accommodates positive spillovers of population density on productivity when $\gamma_N > 0$ and of the college share on productivity when $\gamma_H > 0$.

The amenities enjoyed by each education group, a_e, likewise depend on an exogenous factor as well as the metro's population:

$$a_e = \widetilde{a}_e \left(\frac{N_H}{N_L}\right)^{\gamma_{a,e}}, \tag{4}$$

where $\gamma_{a,e} \geq 0$ and $\widetilde{a}_e > 0$. When $\gamma_{a,e} > 0$, amenities increase with the metro's college share.

Each household in the metro chooses tradeable consumption, $c_T > 0$, non-tradeable consumption, $c_{NT} > 0$, and housing quality, q_j, to maximize utility:

$$U = \beta_e \Big((1 - \xi_e) \log c_T + \xi_e \log c_{NT} + \omega_e \log q_j \Big) + \log a_e + \epsilon, \tag{5}$$

subject to the budget constraint $c_T + r c_{NT} + p_j q_j \leq w_e z$, where r is the price of non-tradeable consumption and ϵ is a taste for living in the metro that is specific to that household. This taste, ϵ, is distributed independently as a type I extreme value across households. Because the coefficients in the utility function depend on education e, this specification allows preferences to differ between non-college and college households.

Following Diamond (2016), I assume that the price of the non-tradeable good, r, is proportional to local house prices. Because there is no single price of housing in my model, I posit that the price of the non-tradeable good is proportional to the geometric average of local house prices:

$$r = \widetilde{r} \prod_{j=1}^{J} p_j^{\frac{1}{J}}, \tag{6}$$

where $\widetilde{r} > 0$ is a constant.[11]

I do not explicitly model the utility from living in the reservation locale and instead assume an exogenous indirect utility from this choice that is specific to each household. I denote this reservation utility by $v_e^0(z) + \epsilon^0$, where ϵ^0 is distributed independently as a type I extreme value across households.

4.2 Equilibrium

Equilibrium consists of house prices $p_0, ..., p_J$ such that the total measure of households choosing each quality of housing q_j equals an amount of such housing rentiers are willing to sell.

To characterize equilibrium, I first describe how the metro's house prices determine the populations of households choosing the metro, $n_L(z)$ and $n_H(z)$. Households whose income does not exceed the smallest house price cannot afford to consume anything in the metro, so $n_e(z) = 0$ in these cases. Otherwise, a household's indirect utility from the metro is calculated by choosing non-housing consumption, c_T and c_{NT}, to maximize utility. I denote this indirect utility by $v_e(z) + \epsilon$, where

$$v_e(z) = -\xi_e \beta_e \log r + \log a_e + \beta_e \max_{j | p_j < w_e z} \Big(\log \big(w_e z - p_j \big) + \omega_e q_j \Big).$$

The household chooses the metro when this indirect utility exceeds the utility from the reservation locale: $v_e(z) + \epsilon > v_e^0(z) + \epsilon^0$. Given standard results on type I extreme value distributions, the

[11] This functional form obtains when competetive firms produce the non-tradeable good with a Cobb-Douglas technology combining the standard housing units in equal proportion.

resulting population choosing the metro equals:

$$n_e(z) = \frac{\widetilde{n}_e(z) \exp v_e(z)}{\exp v_e(z) + \exp v_e^0(z)}. \tag{7}$$

Hence, the population choosing the metro is larger when the utility from that choice is greater.

To finish characterizing equilibrium, I describe how the metro's populations, $n_L(z)$ and $n_H(z)$, determine house prices. I simplify the analysis by making two assumptions about the equilibrium being studied. The first is that the metro's population is greater than the measure of standard housing units, but less than the measure of all housing units including homeless shelters:

Assumption 1. $\sum_{j=1}^{J} h_j < N < \sum_{j=0}^{J} h_j$.

Assumption 1 implies the following Proposition (all proofs appear in Online Appendix B):

Proposition 1. *In equilibrium: $p_0 = 0$, housing demanded by households equals h_j for $j > 0$, and a household's housing quality weakly increases in z for each e.*

Intuitively, because the rentiers holding the housing stock engage in perfect competition, households in equilibrium occupy the N highest quality units of housing available. Given Assumption 1, all standard housing units are therefore fully occupied, while some excess homeless shelters remain. As a result, the equilibrium price of this lowest housing quality, p_0, equals 0. The result in Proposition 1 that households sort on income across housing qualities holds due to Cobb-Douglas utility over housing and non-housing consumption.[12]

The second assumption about the equilibrium being studied is:

Assumption 2. *Some households of each education e choose each quality q_j.*

While Proposition 1 already states that some households choose each quality in equilibrium, Assumption 2 further implies that both non-college and college households choose each quality. An assumption is necessary because the two groups of households may hold different preferences for housing versus non-housing consumption.

Given Assumption 2 and the sorting described in Proposition 1, there exist endowment cutoffs $z_{e,j}$ for each e and $j > 0$ such that a household of education e and labor endowment $z_{e,j}$ is indifferent between choosing qualities q_j and q_{j-1}. This indifference implies the equation:

$$\left(w_e z_{e,j} - p_j\right) q_j^{\omega_e} = \left(w_e z_{e,j} - p_{j-1}\right) q_{j-1}^{\omega_e}, \tag{8}$$

[12] Määttänen and Terviö (2014) show that for more general utility functions, this sorting condition holds whenever the marginal rate of substitution from housing to non-housing consumption increases in non-housing consumption.

for each e and $j > 0$. Proposition 1 further implies that:

$$\sum_{j'=j}^{J} h_{j'} = \sum_{e\in\{L,H\}} \int_{z_{e,j}}^{\infty} n_e(z)dz, \tag{9}$$

for each $j > 0$. Eqs. (8) and (9) provide $3J$ equations in $3J$ unknowns consisting of house prices and the endowment cutoffs. I show in Online Appendix B.2 that a unique solution exists. When the endowment cutoffs, $z_{e,j}$, in this solution strictly increase in j, then an equilibrium satisfying Assumption 2 exists.

4.3 Comparative statics

To predict how endogenous variables change in response to perturbations to housing stocks, h_j, and productivity and amenity shifters, $\widetilde{A}_e$ and $\widetilde{a}_e$, I calculate comparative statics around a given equilibrium. I can calculate these comparative statics using information solely about the metro—as opposed to data on the reservation locale—under the assumption that the share of each type of household in the economy choosing the metro is small:

Assumption 3. *For each e and z, $n_e(z) \ll \widetilde{n}_e(z)$.*

Assumption 3 seems reasonable for studying a single metro in a large economy such as the US.

Applying Assumption 3 to Eq. (7), which determines the metro's population, yields an equation relating net migration to changes in non-tradeable prices, amenities, wages, and house prices:

$$\partial \log n_e(z) = \left(1 - \frac{n_e(z)}{\widetilde{n}_e(z)}\right) \partial \log v_e(z) \approx -\xi_e \beta_e \partial \log r + \partial \log a_e + \frac{\beta_e w_e z}{w_e z - p_j} \partial \log w_e - \frac{\beta_e}{w_e z - p_j} \partial p_j, \tag{10}$$

where j is the quality index of the housing that a household of education e and labor endowment z chooses, and the operator ∂ denotes the change in response to exogenous changes to the housing stock and productivity and amenity shifters. The elasticity of the city's population with respect to the non-amenity variables scales with β_e, so I refer to β_L and β_H as migration elasticities.

To complete a system of equations for comparative statics, I aggregate Eq. (10) across households to obtain $\partial \log N$, the change in the metro's total population, as well as $\partial \log N_e$ and $\partial \log Z_e$, the changes in the aggregate population and labor endowment of each education group. Differentiating Eqs. (2)–(4) and (8)–(9) completes the system of equations. By inspecting the coefficients in this system of equations, I prove the follow proposition:

Proposition 2. *First-order changes to outcomes in the metro in response to shocks to the housing stocks, h_j, and the productivity and amenity shifters, $\widetilde{A}_e$ and $\widetilde{a}_e$, depend only on these aspects of the initial equilibrium: house prices, p_j, populations by education, N_e, the distributions of income by education, $f_e(y)$, and the minimal incomes among households of education e choosing each housing quality, $y_{e,j}$.*

Therefore, to calculate comparative statics for a given metro, I need only fit the variables listed in Proposition 2. In Section 5.1, I describe how I complete that exercise.

5 Fitting the model to the data

To fit the model to the data, I focus on the Los Angeles-Long Beach-Anaheim, CA metro area. In addition to being the second largest superstar metro (see Table 1), Los Angeles overlaps exactly with a set of PUMAs, allowing me to select all persons in this metro area in IPUMS.

I fit the model in two steps. First, I fit the model equilibrium in Section 4 to the observed micro data for 2019. In this step, I target 2019 and not 2000 because the data on home values are binned in IPUMS in 2000 but not in 2019. Then, I choose shocks to housing supply, productivity, and amenities to match the 2000–2019 evolution of the Los Angeles housing market relative to non-superstars.

5.1 Fitting the 2019 equilibrium

To fit the 2019 equilibrium observed in the data, I select a sample from IPUMS with observations in the 87 PUMAs making up the Los Angeles metro area. Because I require housing cost data for households, I drop 2% of renters who do not pay cash rent. Among persons in group quarters, I keep those who seem likely to be in homeless shelters based on the methodology described in Section 3. Summary statistics for the variables I use appear in Table A5 of the online appendix.

To estimate house prices by quality level, p_j, I first annualize housing costs by multiplying gross rents by 12 and house values by ϕ, the PUMA-specific user cost described in Section 3. I then form percentiles by this housing cost and set p_j to the average housing cost in percentile j. I set the housing stock in percentile j, h_j, equal to the number of households in that percentile, which is approximately constant across percentiles. I set the price of housing for the homeless, p_0, to 0 to be consistent with Proposition 1. I observe the number of households with each education, N_L and N_H, directly in the data.

In the data, households within education group do not perfectly sort based on income across housing percentiles, even though Proposition 1 predicts such sorting. To resolve this discrepancy, I assume that the income observed in IPUMS may differ from the long-run income a household uses to make housing choices, and I further assume that this difference is orthogonal to the household's education and housing quality choice.[13] Therefore, the average income in IPUMS for households of education e choosing housing percentile j gives an unbiased estimate of the average latent income in this group. I estimate the latent income distributions, $f_L(\cdot)$ and $f_H(\cdot)$, by targeting these income averages as well as the share of college households choosing each housing quality level, a total of 303 moments.

To conduct this exercise, I make a parametric assumption that each latent income distribution is double Pareto-lognormal, a four-parameter family that Reed (2003) and Reed and Jorgensen (2004) propose to characterize income distributions. I denote the vector of parameters identifying each distribution $f_e(\cdot)$ by θ_e. I jointly estimate a vector $\theta = (\theta_L, \theta_H, \zeta)$, where $\zeta = \omega_L/\omega_H$ is the relative weight on housing consumption for non-college versus college households. Together with the estimates for house prices, p_j, housing stocks, h_j, and populations, N_e, this vector θ uniquely determines the income cutoffs, $y_{e,j}$.[14] Therefore, each vector θ provides model-generated moments corresponding to the 303 empirical moments listed above. I choose the vector θ best fitting these moments using two-step GMM.

As shown in Figure 3, the estimated model closely fits the empirical moments. I plot average incomes by household education and housing quality in Panel A and the shares of households choosing each housing quality with a college degree in Panel B. I estimate that $\zeta = 1.760$, which indicates that non-college households value housing versus non-housing consumption more than college households. This finding is consistent with the smaller incomes of non-college households within each housing percentile. The college share in each percentile lies strictly between 0 and 1, consistent with Assumption 2.

5.2 Fitting the 2000–2019 change

I choose shocks to amenities, productivity, and the housing stock in the model to match the evolution of the Los Angeles housing market between 2000 and 2019 relative to non-superstar metros.

[13]The difference between income in IPUMS and the long-run income that governs housing choice may arise due housing adjustment costs that lead households not to change their housing choices when income changes (Chetty and Szeidl, 2007).

[14]To guarantee that the candidate parameter vector, θ, uniquely determines income cutoffs, $y_{e,j}$, that increase in the quality index, j, I restrict to a subset of the parameter space as described in Online Appendix C.1.

To conduct this exercise, I first select values for the remaining parameters, drawing on previous estimates in the literature. The values and their sources appear in Panel A of Table 3. I calculate preference parameters using results from Diamond (2016), who estimates a city choice model with a utility specification similar to Eq. (5). I use her estimates for the elasticity of migration with respect to wages to pin down the migration elasticities in my model, β_L and β_H, and her estimate for the share of income spent on local goods to determine the relative importance of non-tradeable consumption in my model for each education group, ξ_L and ξ_H. I calculate ρ, the parameter governing the elasticitity of substitution between college and non-college labor, using consensus estimates of this production function from the literature (Card, 2009). I similarly take γ_N, the elasticity of productivity with respect to metro population, from consensus estimates (Combes and Gobillon, 2015). Finally, I derive γ_H, the productivity elasticity with respect to the metro's college share, from estimates of this spillover in Moretti (2004*b*). Further details of how I map these estimates from the literature to my parameters appear in Online Appendix C.2. I set $\gamma_{a,L}$ and $\gamma_{a,H}$, the elasticities of metro amenities with respect to the college non-college ratio, to 0 for my baseline estimates but experiment with non-zero values later in the paper.

The outcomes in the Los Angeles housing market that I target appear in Panels B and C of Table 3. I calculate each outcome for Los Angeles relative to the average change in non-superstar metros over the same time period. Because I am shocking the model fit to the 2019 equilibrium, I normalize these empirical differences by the level in Los Angeles in 2019. To match the outcomes in Panel B, I smooth the decile supply effects across all 100 percentiles of the quality distribution; the smoothed series appears in Figure A2 of the online appendix, with additional information about this smoothing procedure in the note to that figure. I then match the four targets in Panel C by choosing shocks to the exogenous productivity and amenity shifters for the noncollege and college groups: $\widetilde{A}_L$, $\widetilde{A}_H$, $\widetilde{a}_L$, and $\widetilde{a}_H$.[15] Intuitively, the productivity shocks pin down the income changes by education group. The sum of the amenity shocks determines average house price growth, while the difference between the amenity shocks pins down the relative growth of the college population.

To validate the match between the data and the model's predicted changes, in Panels C and D of Figure 3 I plot population growth by education and housing decile in both the data and the model. In the data, I calculate population growth as the difference between Los Angeles and non-superstar metros, normalized by the count of occupied housing units in the corresponding decile

[15] The calculated shocks to the logs of these variables are respectively −0.041, 0.073, 0.195, and 0.071.

in Los Angeles in 2019. The model's predicted population change matches the data well at most deciles. This exercise validates the model fit because I do not target the population growth of each education group in each decile. The model also matches the slight gentrification of Los Angeles relative to non-superstars over this time: the difference between total normalized college growth and non-college growth is 4.4 p.p. in the data and 4.2 p.p. in the model. Therefore, the model matches aspects of the 2000–2019 Los Angeles housing market beyond the targets in Table 3.

6 Effects of construction in the model

In this section, I report the effects of construction within the estimated model. I first focus on a counterfactual in which the housing stock in Los Angeles grows between 2000 and 2019 like the average non-superstar metro. As a second exercise, I investigate the heterogeneous effects of construction at difference quality levels.

6.1 Superstar effect

In the first counterfactual, I explore how Los Angeles's housing market would have evolved had the housing stock grown like the average non-superstar metro between 2000 and 2019. To conduct this exercise, I hold constant the exogenous shocks to productivity and amenities I calculated in Section 5.2 to fit the targets in Panel C of Table 3. At the same time, I change the growth in the housing stock at each percentile of the quality distribution to zero. Therefore, this counterfactual illustrates Los Angeles with housing growth resembling that of non-superstar metros but with shocks to productivity and amenities particular to Los Angeles.

The targeting of counterfactual construction to the data is a key difference relative to other papers that evaluate the equilibrium effects of new construction. For instance, Favilukis, Mabille and Van Nieuwerburgh (2023) explore a policy that increases the housing stock in high-priced areas but simultaneously decreases the housing stock in less expensive areas, while Couture et al. (2023) evaluate a uniform increase in housing supply through out a city. In contrast, I study a policy that increases the supply of high-quality housing while leaving the stock of low-quality housing essentially unchanged. This policy differs from the ones considered in the prior work and is grounded in the empirical difference between Los Angeles and non-superstar metros.

In Figure 4, I plot house price growth by quality percentile in Los Angeles relative to non-superstar metros, in both this counterfactual and the baseline. At every quality percentile, house price growth is smaller in the counterfactual. Differences relative to the baseline range from 13.9 percentage points for the lowest percentile to 4.1 for the highest percentile, corresponding to 25.4% and 50.4% of the baseline effect. The attenuation of house price growth even at low quality deciles is consistent with the reduced-form guess, based on Panel B of Figure 2, that a lack of high-quality construction in superstar metros raises house prices throughout the quality distribution. It may also explain the difference between the results here and those in Anenberg and Kung (2020), who predict that a similarly sized expansion of the housing stock in Los Angeles lowers rents only -0.71%. While Anenberg and Kung (2020) hold constant rents in the outer PUMAs of Los Angeles, I allow prices of low-quality standard housing units to change in response to the new construction. With this additional construction, average house price growth (calculated as in Table 3) in Los Angeles relative to non-superstars falls to 9.4% from 15.0%. Excess house price growth persists due to the calibrated shocks to productivity and amenities.

To illustrate the mechanism through which construction lowers house prices, I next quantify the marginal effect of construction on population flows into and out of each housing type. I measure these marginal effects as differences between population growth in the counterfactual and the baseline, and I decompose the population flows into across-metro and within-metro effects. The across-metro effect represents changes in the population of each household type choosing the metro in response to changes in prices, wages, and amenities as described in Eq. (10). The within-metro effect arises as existing households change their housing quality choices; this effect appears in the model as changes to the cutoffs assigning households to housing quality types. To make these flows analogous to those in Panels C and D of Figure 3, I normalize population flows by the initial stock of housing in each percentile.

The results of this exercise, which appear in Figure 5, illustrate the trickle-down mechanism. Households already in the city—not migrants from outside—occupy most of the new housing in the highest quality percentiles. This result is apparent from the excess of the "within metro" curve above the "across metro" curve for top percentiles in Panel B, and the equality of these curves for top percentiles in Panel A. As a result of the upward movement of existing households along the quality distribution, the demand from these households for low-quality housing declines, as is apparent from the negativity of the within-metro curves below the 40th percentile. For the housing market at low qualities to clear, the prices of such housing units must fall to induce in-migration. Because they occupy low-quality housing, these migrant households are likely to be poor and lack a college degree, as is apparent from the height of the across-metro curve in the non-college panel for low quality percentiles. Hence, construction of high-quality housing trickles down to allow poor non-college households to live in the metro via cheaper house prices.

The results in Figure 5 also reveal the effect of construction on gentrification: the difference between the "total" curves across the two panels gives the relative increase of the college versus the non-college population. I plot this difference in Figure A3 of the online appendix. In top quality percentiles, construction increases gentrification as college households occupy more of this new housing than non-college households. At the same time, construction decreases gentrification at lower quality percentiles, specifically, all those below the 80th. Non-college households who migrate into lower quality housing drive this effect. Hence, the construction of high-quality housing has ripple effects on gentrification throughout the metro, offsetting the direct positive effect on the submarkets where construction is located. The net effect on gentrification for the metro (difference between normalized college and non-college growth) is −3.7 percentage points, essentially offsetting the excess gentrification in Los Angeles relative to non-superstars over this time. This finding suggests that the positive effect of superstar status on gentrification is due largely to the lack of high quality construction in superstar metros. Consistent with this channel, construction in the model relieves gentrification most strongly in middle quality percentiles, which matches the particularly strong empirical effect of superstar status on gentrification in middle deciles.

How many existing households move out of the bottom half, or bottom quintile, of the quality distribution for each new housing unit? These statistics gauge the degree to which in-migration attenuates the trickle-down mechanism; without migration, 100 new units in the top half of the quality distribution lead 100 households to move out of both the bottom half and bottom quintile (Braid, 1981). As I report in Panel A of Table 4, 100 new units in the counterfactual lead 33 (resp. 20) households to move out of the bottom half (resp. quintile) of the quality distribution into nicer housing.[16] Hence, migration attenuates much of the potential effect of high-quality construction on the housing outcomes of lower-income households.

In Panel A of Table 4, I also report that construction decreases the counts of homeless persons in the metro. Construction lowers the price of the first housing quality percentile (see Figure 4), so it induces some homeless households to switch into this type of housing. Quantitatively, 100 new housing units in the counterfactual lead three to four households to move from homelessness to standard housing. According to this result, some of the growth in homelessness in superstar

[16]Mast (2023) estimates that 100 new market rate units free up 40 (resp. 15) units in neighborhoods in the bottom half (resp. quintile) of the income distribution in the same metro; see his Figure 3. These numbers are close to the corresponding estimates in Table 4.

metros, documented in Table 2, comes from the lack of construction in superstar metros, even though the missing housing units are primarily high-quality units.

Construction raises the average welfare of households living in the metro. As reported in Panel B, building like a non-superstar raises non-college welfare by 4% and college welfare by 5.6%. I calculate welfare as consumption-equivalent units by normalizing the change in indirect utility by the migration elasticities, β_L and β_H. These welfare benefits are large relative to the 11.8% expansion of the housing stock considered in the experiment. They are also larger than the welfare benefits of upzoning in Favilukis, Mabille and Van Nieuwerburgh (2023) by an order of magnitude. Welfare increases because construction lowers the price of housing by 5 to 6%, which increases expenditure on non-housing consumption and also lowers the price of the non-tradeable good. There is a small, offsetting effect on welfare for non-college households because their wages fall as relatively more non-college households migrate into the metro. College wages rise 1% due to this effect, increasing their welfare.

6.2 Construction at different quality levels

In the second counterfactual, I measure the extent to which households prefer construction at one quality level versus another. I focus on two groups of households. The first, which I call "rich college," are those with a college degree initially living in the top 25 housing percentiles; "poor non-college" are those without a college degree in the bottom 25 housing percentiles initially. I measure welfare effects as consumption-equivalent changes in the average utility of households in the metro before the construction experiment. To make this experiment comparable to that in Section 6.1, I increase the metro's housing stock by the same amount here as I do in that section. However, in this counterfactual, I concentrate that entire increase in a single quality percentile. The results from this exercise appear in Figure 6.

As shown in Panel A, poor non-college households benefit most from construction at the ninth percentile. Quantitatively, building one unit of this quality raises their welfare 87% more than building at the top percentile and 42% more than the growth of their welfare in the construction experiment in Section 6.1. Conversely, rich college households benefit most from construction at the top quality level, with a 140% gain relative to building at poor non-college households' preferred level at the sixth percentile. Hence, the two groups have households hold conflicting preferences about the quality of new construction.

These conflicting preferences arise due to the trickle-down effect of new construction onto house prices. I illustrate this mechanism in Figure A4 of the online appendix, which shows the effect on house prices throughout the metro of building at the 9th percentile versus the 75th percentile. Construction at the 9th percentile sharply lowers the prices of low-quality homes while having close to no effect on the prices of the highest-quality homes. Intuitively, low-quality house prices fall to clear the market by moving existing poor households toward the 9th percentile and attracting new poor households from outside the city; high-quality prices do not need to fall to clear the market. In contrast, construction at the 75th percentile lowers house prices in a roughly uniform way throughout the metro. In this case, the fall in prices clears the market by moving existing households up the quality distribution towards the 75th percentile and by drawing in new households across the income distribution. Because this construction draws migrants across the income distribution, the price effect on low-quality housing is less extreme than the case of construction at the 6th percentile.

Although the two groups of households disagree on the preferred quality of construction, each group benefits from each type of construction. However, under alternative parameters, rich college households become worse off from low-quality construction. I display this result in Panel B of Figure 6, which is analogous to Panel A except that amenities are now endogenous to the metro's college share, as in Diamond (2016).[17] With endogenous amenities, construction at any of the bottom 40 percentiles lowers the welfare of rich college households. Intuitively, such construction draws in relatively more non-college migrants, lowering the metro's college/non-college ratio and thereby decreasing the amenities enjoyed by rich college households. This amenity effect dominates the near-zero effect of low-quality constriction on high-quality house prices, leading to a decrease in welfare. Interestingly, poor non-college households gain more from construction when amenities are endogenous compared to when they are exogenous. This phenomenon stems from the parameter values, based on Diamond (2016), under which non-college households value amenities less than college households. As a result, the decline in amenities triggered by new construction lowers house prices in a way that, on net, makes non-college households better off.

According to the results in Figure 6, high-quality construction can make poor non-college households worse off if it comes through demolishing low-quality units. For instance, tearing down units at the 9th percentile to build an equal number of high-quality units lowers the average welfare of these households. This decline in welfare occurs purely through equilibrium effects and

[17]Specifically, I set $\gamma_{a,L} = 0.548$ and $\gamma_{a,H} = 2.024$. I discuss how I calculate these numbers using estimates from Diamond (2016) in Online Appendix C.2.

not due to any displacement costs borne by the people previously living in the demolished unit. However, their welfare increases if a sufficient quantity of new construction replaces the torn-down units. As an example, if developers tear down 100 units at the 9th percentile to build units at the 75th percentile, then poor non-college welfare increases as long as the developers build 142 new units or more. This transformation necessitates increasing the market value of housing at least by a factor of 3.9, given that units at the 9th percentile are only 37% as valuable as units at the 75th. In contrast, rich college households require fewer new units to achieve a welfare gain: 77 units under exogenous amenities, and zero units under endogenous amenities, as simply demolishing low-quality units raises their welfare in equilibrium.

These results may raise a consideration for policymakers giving out permits for urban renewal projects that involve tearing down low-quality units to build high-quality units. More generally, any policy that encourages construction of high-quality units at the expense of low-quality units could lower the welfare of the poor, depending on how many low-quality units are forgone and how many high-quality units are built. This observation may explain why Favilukis, Mabille and Van Nieuwerburgh (2023) find small welfare effects of upzoning the core of New York City, as this policy shifts construction resources away from the periphery and lowers the housing stock in this less expensive area.

7 Conclusion

In this paper, I highlight the trickle-down effects of high-quality construction on the housing market throughout a metro area. Reduced-form evidence from superstar metros suggests strong trickle-down effects on house prices, gentrification, and homelessness. I confirm these effects in an equilibrium model fit to Los Angeles, which permits estimating the effects of high-quality construction in a more causal framework. The predicted effects of construction are stronger than those in some recent papers, potentially because my counterfactual matches the quality distribution of the difference in construction between Los Angeles and non-superstar metros.

While some factors relevant for policy are missing from my framework, it suggests that policies encouraging market-rate construction could relieve housing affordability issues for the poor, even if most of this construction is at the top end of the market. However, policies that encourage the supply of high-quality units at the expense of low-quality units have more subtle effects and can potentially make the poor worse off.

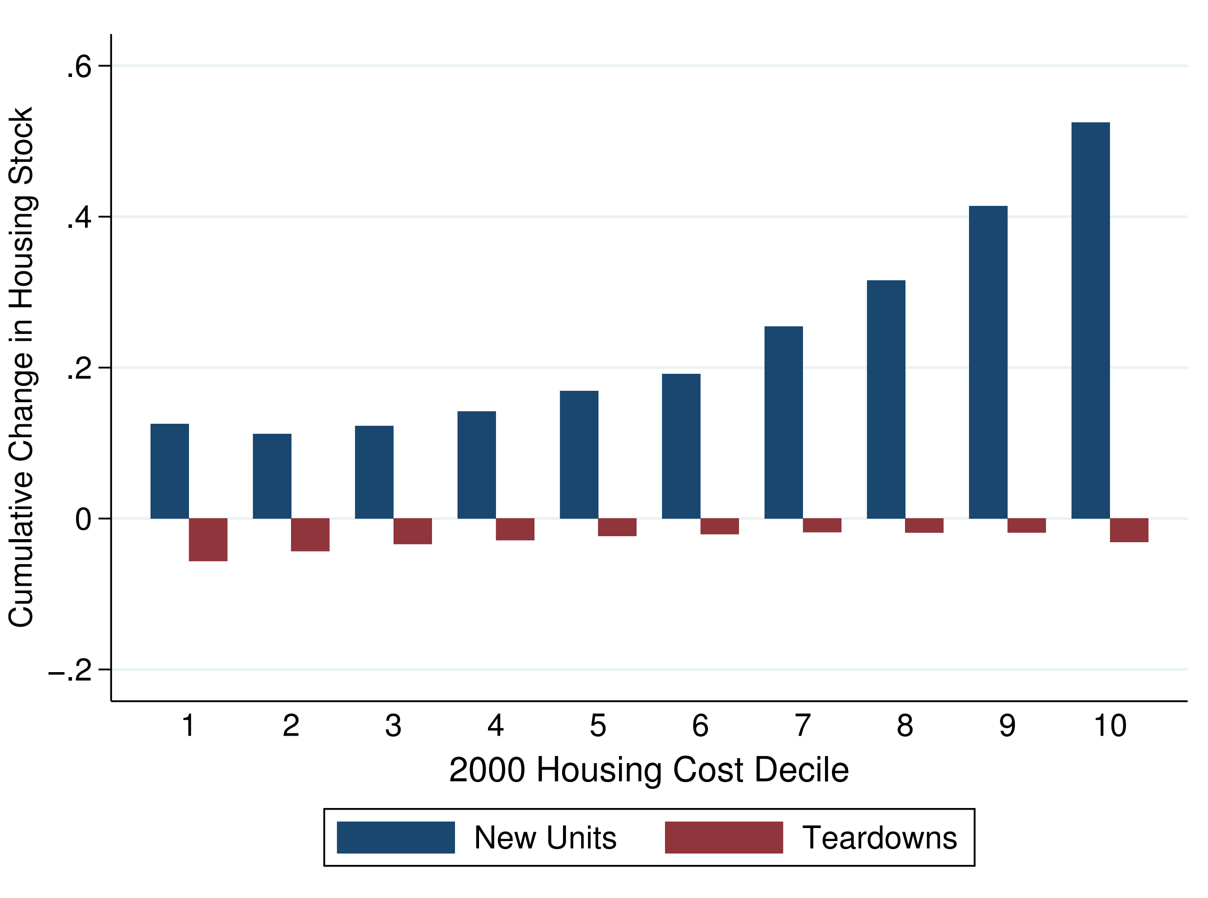

Figure 1. 2000–2019 Housing Stock Growth by 2000 Cost Decile

Notes: Data cover housing units within 2018 CBSA boundaries in the 2000 5% U.S. Census sample and the 2015–2019 ACS. I annualize housing costs by multiplying gross rents by 12 and home values by a user cost specific to each PUMA and survey year. For each CBSA, I form deciles of this cost using all units built before 2000 in each of the 2000 and 2015–2019 surveys. When assigning deciles in the 2015–2019 survey, I reweight housing units to match the distribution by year-built within the same Census tract in the 2000 survey. In the figure, growth from new units equals units built on or after 2000 in the 2015–2019 survey divided by units built before 2000 in the 2000 survey; growth from teardowns equals units built before 2000 in the 2015–2019 survey divided by units built before 2000 in the 2000 survey, minus one. I report the average of these growth components across CBSAs weighted by housing units in 2000.

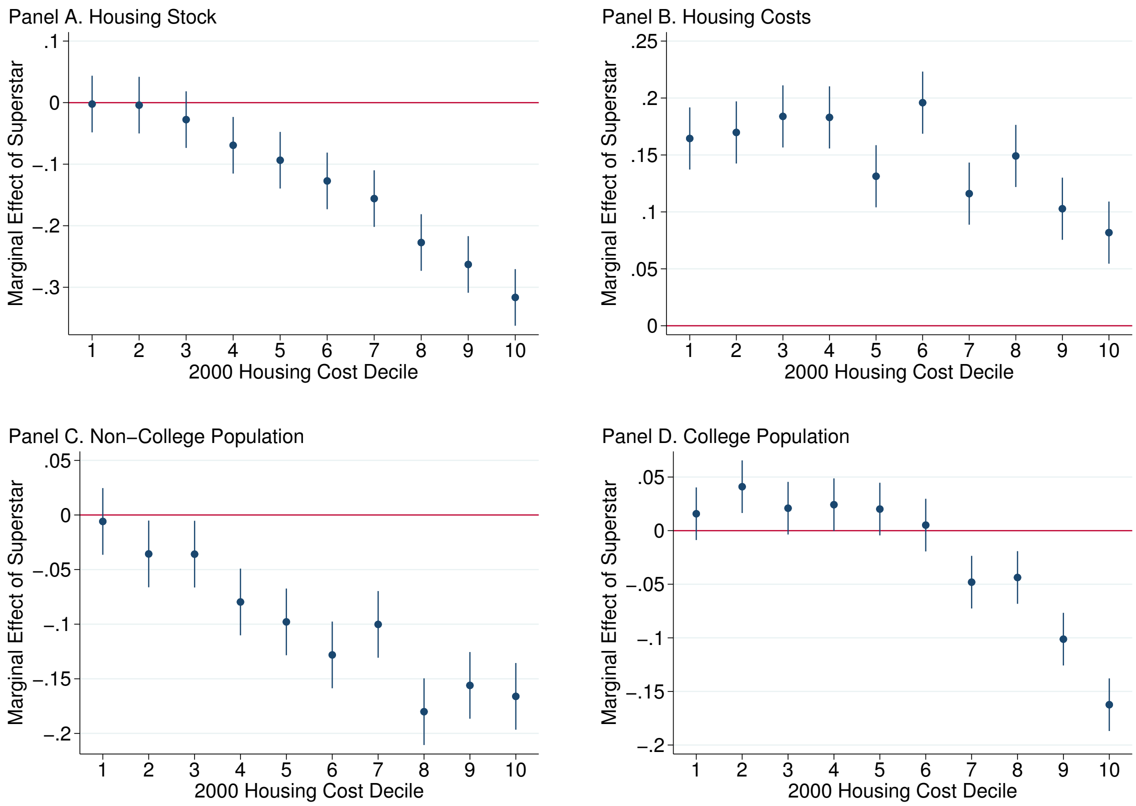

Figure 2. Marginal Effects of Superstar Indicator on 2000–2019 Growth

Notes: Data cover housing units within 2018 CBSA boundaries in the 2000 5% U.S. Census sample and the 2015–2019 ACS. I assign housing units to deciles as described in the note to Figure 1. I designate 15 CBSAs as superstars as described in the note to Table 1. In each panel, I regress an outcome at the CBSA-decile level on decile dummies and deciles dummies interacted with the superstar indicator, and I plot the coefficients on the interaction terms. In these regressions, I weight observations by each CBSA's housing stock in 2000. In Panel A, the outcome is the sum of housing stock growth from new units and from teardowns as defined in the note to Figure 1. In Panel B, the outcome is real median housing cost among occupied units, excluding rented single-family homes on 10 or more acres in 2000 and renter-occupied units without cash rent in both survey years. In Panel C (resp. D), the outcome is the difference in the number of non-college (resp. college) households between survey years, divided by the number of occupied units in 2000. I annualize housing costs and impute college status for households in NHGIS as described in the note to Table 1.

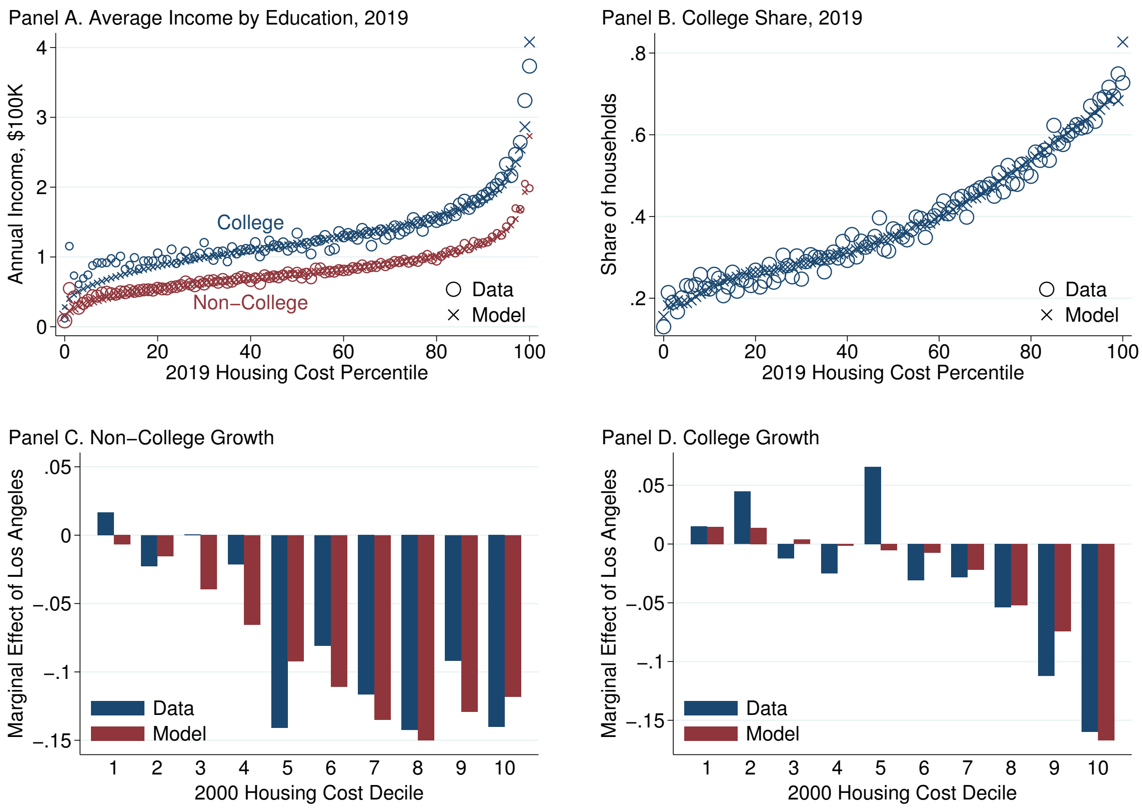

FIGURE 3. GOODNESS OF FIT

Notes: In Panels A and B, data come from PUMAs in the Los Angeles metro in the 2015-2019 ACS. I annualize housing costs as described in the note to Figure 1. Percentile 0 represents homeless persons, defined as in the note to Table 2. The size of each market in Panel A is proportional to the number of observations in that bin and education group in the data. Income is units of 2019 dollars. In Panels C and D, I measure population changes in the data as described in the note to Figure 2 and report the differences between Los Angeles and the average non-superstar metro, using housing units in 2000 as weights to take this average. In the model, I assign 2019 percentiles to 2000 deciles using the shares of 2019 housing units in Los Angeles in each quality decile in the data underlying Figure 1.

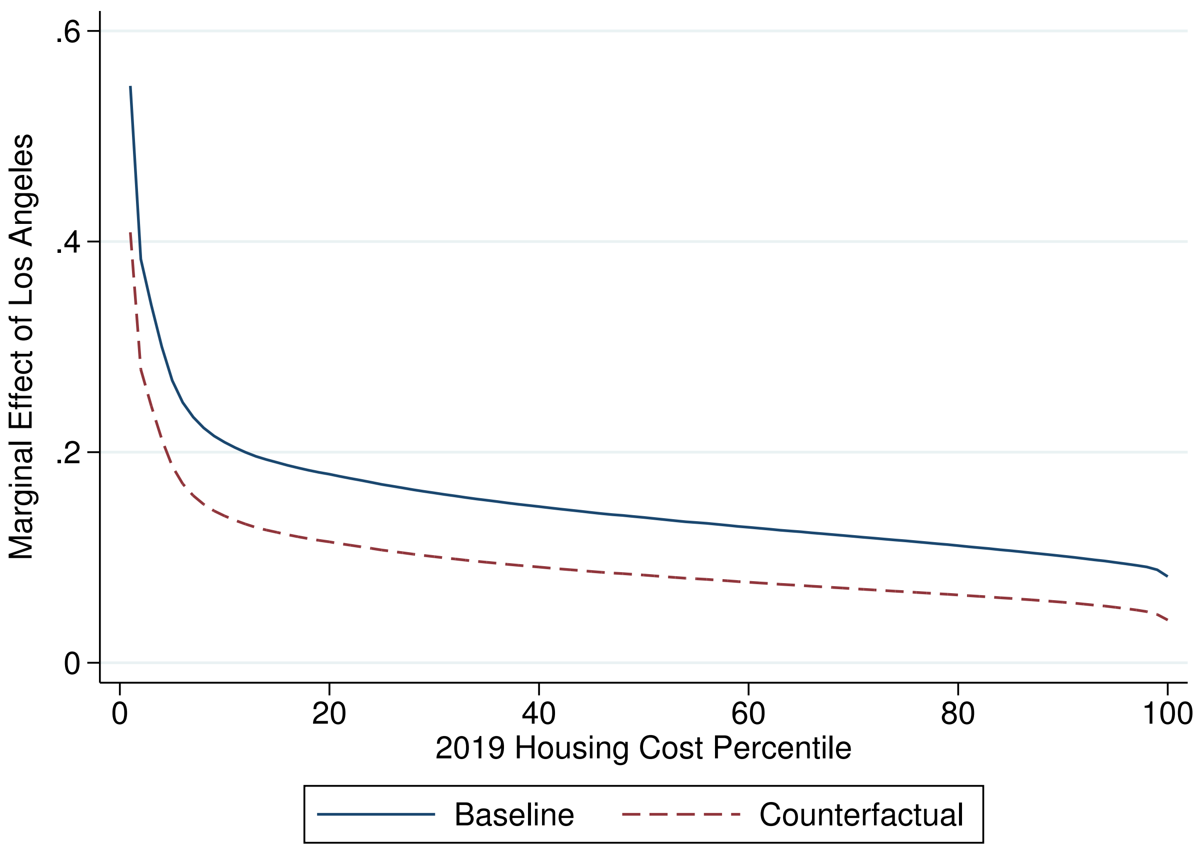

Figure 4. 2000–2019 Housing Cost Growth, Model Baseline and Counterfactual

Notes: The results are house price changes of different quality percentiles within a model fit to the Los Angeles metro housing market in 2019. In the baseline, I pick changes to the housing stock and metro productivity and amenities to match the targets in Panels B and C of Table 3. In the counterfactual, I hold constant the shocks to productivity and amenities while altering the housing stock changes to zero.

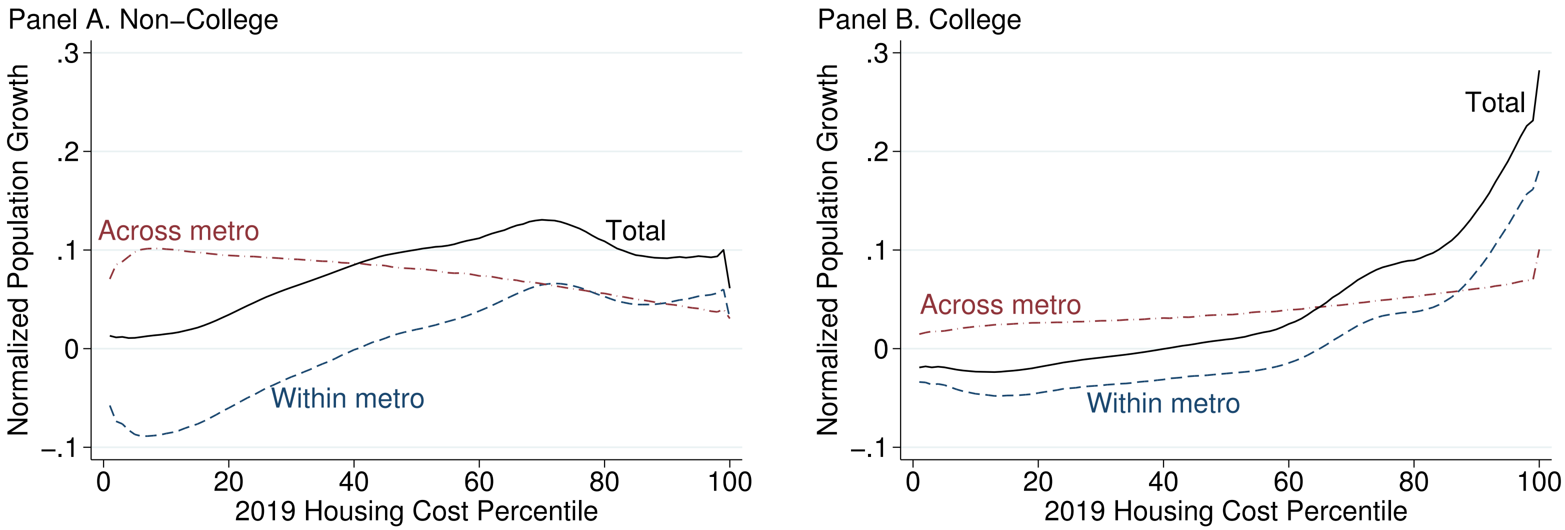

Figure 5. Marginal Effects of Building like Non-Superstar on Populations

Notes: The results are counterfactuals within a model fit to the Los Angeles metro housing market in 2019. I increase the housing stock at each quality percentile by the difference in the growth of such housing between the average non-superstar and Los Angeles between 2000 and 2019. The note to Figure A2 of the online appendix contains additional information on measuring these differences. Population growth is normalized by the initial count of housing units in each percentile. "Within metro" gives the change coming from households already in the metro, and "across metro" provides the change arising from households moving into or out of the metro.

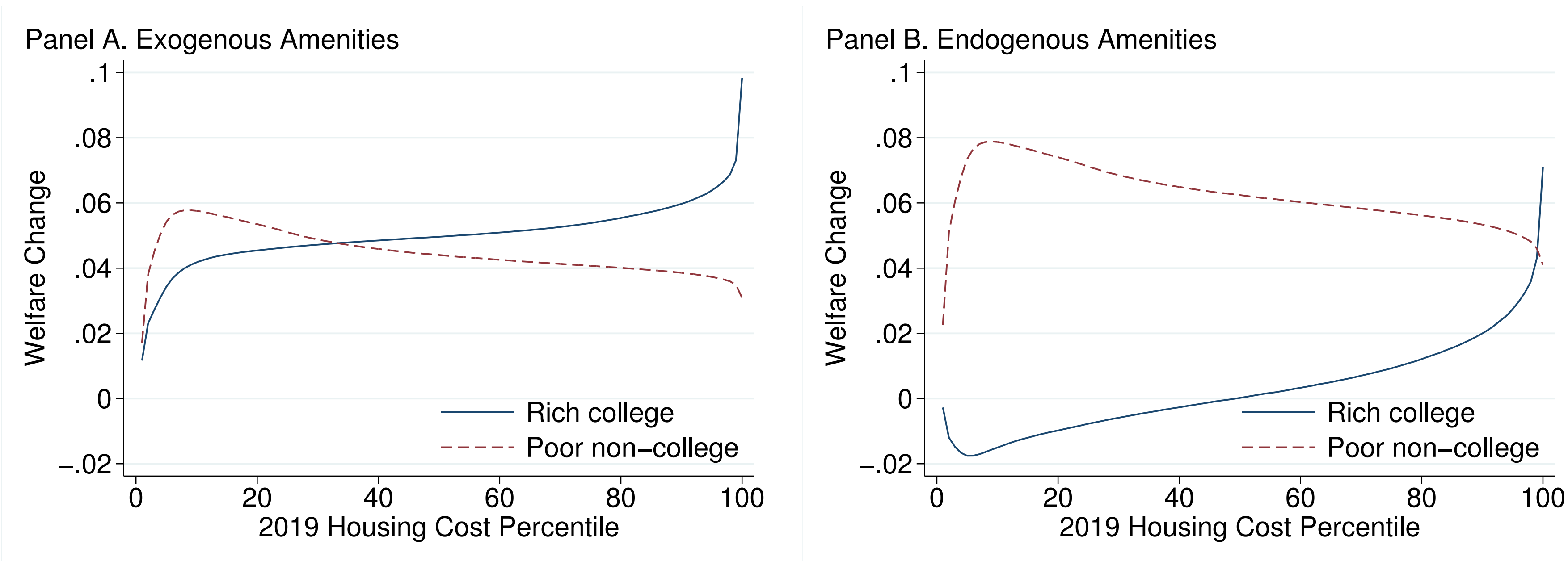

Figure 6. Effects of Construction at Different Percentiles on Welfare

Notes: The results are counterfactuals within a model fit to the Los Angeles metro housing market in 2019. For each percentile, I conduct a counterfactual in which I increase the metro's housing stock by 11.8% entirely with housing at that percentile. I report the average welfare change for rich college households, defined as those initially living in the top 25 housing percentiles, and poor non-college households, defined as those initially living in the bottom 25 housing percentiles. I normalize welfare as described in the note to Table 4. To produce the counterfactuals, I use the parameters in Panel A of Table 3, except that I set $\gamma_{a,L} = 0.548$ and $\gamma_{a,H} = 2.024$ to produce Panel B of this figure.

TABLE 1
Superstar vs. non-superstar metros, 2000–2019

	Superstar	Non-superstar
Panel A. Superstar counts		
Number of CBSAs	15	369
Share of 2000 housing stock in CBSAs	0.158	0.842
Panel B. Cumulative growth rates		
Housing units	0.097	0.221
Home values	0.563	0.236
Housing costs	0.278	0.128
Household income, non-college	−0.041	−0.084
Household income, college	0.011	−0.052
Panel C. Normalized growth rates		
Non-college households	−0.068	0.031
College households	0.131	0.154
Panel D. Largest CBSAs by 2000 housing units		
1	New York	Chicago
2	Los Angeles	Philadelphia
3	San Francisco	Miami
4	Baltimore	Dallas
5	Honolulu	Washington

Notes: Data cover housing units within 2018 CBSA boundaries in the 2000 5% U.S. Census sample and the 2015–2019 ACS. I assign superstar status as in Gyourko, Mayer and Sinai (2013) using the annualized growth rates of total housing units and real mean owner-occupied home values between the two surveys. I annualize growth rates by raising the cumulative growth rate to the 1/17 and subtracting 1. Superstars are CBSAs above the 50th percentile in the sum of these growth rates and above the 90th percentile in the ratio of the price to quantity growth rates. Outcomes in Panels B and C are averages weighted by 2000 housing stock within each group of CBSAs. Dollar amounts are adjusted for inflation. For home values, I use the mean among owner-occupied units. I annualize housing costs by multiplying gross rents by 12 and home values by a user cost specific to each PUMA and survey year; I use the median of this cost among all occupied housing units, excluding rented single-family homes on 10 or more acres in 2000 and renter-occupied units without cash rent in both survey years. I impute college status and income of occupied housing units in NHGIS using the means by survey year, PUMA, tenure, and housing cost bin in IPUMS. The normalized growth in non-college and college households is the difference in the count of each group between the survey years, divided by the count of occupied units in 2000. In addition to the CBSAs in Panel D, the other superstars are Santa Barbara, Longview, Santa Cruz, Ocean City, Chico, Abilene, Fort Drum, Odessa, Napa, and San Angelo.

TABLE 2

Marginal effect of superstar status, 2000–2019

	Log(persons per unit)	Log(occupancy rate)	Log(homeless rate)
	(1)	(2)	(3)
2019 dummy	0.011*	−0.031***	−0.327***
	(0.006)	(0.004)	(0.036)
Superstar dummy	0.065***	0.023***	0.515***
	(0.010)	(0.008)	(0.063)
2019 × Superstar	0.000	−0.001	0.250***
	(0.014)	(0.011)	(0.090)
Constant	0.946***	−0.083***	−5.028***
	(0.004)	(0.003)	(0.025)
R^2	0.106	0.093	0.275
Number of observations	768	768	768

Notes: Data cover housing units within 2018 CBSA boundaries in the 2000 5% U.S. Census sample and the 2015–2019 ACS. Observations are at the level of CBSA-year. I weight observations by the number of housing units in 2000. I assign superstar status to CBSAs as described in the note to Table 1. Persons per unit equals the count of persons in housing units divided by the number of housing units. Occupancy rate equals the ratio of occupied units to all units. Homeless rate equals the number of homeless persons divided by the count of housing units. In IPUMS, the persons I categorize as homeless are those in non-institutionalized group quarters who are at least 18, either unemployed or out of the labor force, and not in school. I aggregate homeless counts to the CBSA level using the average homeless status among group quarters persons in each survey year and PUMA in IPUMS, which I multiply by the count by CBSA and PUMA of group quarters persons in NHGIS. I denote significance levels 10%, 5%, and 1% by *, **, and ***, respectively. Standard errors appear in parentheses.

TABLE 3

INPUTS INTO MODEL CALIBRATION

Parameter or target	Value
Panel A. Parameters	
Migration elasticity, non-college (β_L)	3.051
Migration elasticity, college (β_H)	1.755
Preference for non-tradeables, non-college (ξ_L)	0.501
Preference for non-tradeables, college (ξ_H)	0.543
Substitutability of college, non-college labor (ρ)	0.300
Population spillover on productivity (γ_N)	0.055
College share spillover on productivity (γ_H)	0.097
Amenity elasticity, non-college ($\gamma_{a,L}$)	0.000
Amenity elasticity, college ($\gamma_{a,H}$)	0.000
Panel B. Targets, housing stock growth	
Decile 1	0.008
Decile 2	−0.001
Decile 3	−0.036
Decile 4	−0.067
Decile 5	−0.097
Decile 6	−0.118
Decile 7	−0.157
Decile 8	−0.202
Decile 9	−0.203
Decile 10	−0.285
Panel C. Targets, other growth	
Average housing cost	0.150
Household income, non-college	0.041
Household income, college	0.031
College households (normalized)	−0.032

Notes: Sources for the parameters in Panel A appear in Section 5.2, and details on calculating the parameters from these sources appear in Online Appendix C. In Panels B and C, I report differences between Los Angeles and the average non-superstar metro in growth between 2000 and 2019, weighting by the housing stock in 2000 when averaging over non-superstars. Because I calibrate the model to 2019 data, I divide these differences by 1 plus the growth in Los Angeles between 2000 and 2019 (in the case of normalized college households, by 1 plus the growth of occupied units). Average housing cost is the average of the growth of the median housing cost over the ten deciles. Further information on measuring the variables in Panels B and C appears in the notes to Figures 1 and 2 and Table 1.

TABLE 4

MARGINAL EFFECTS OF BUILDING LIKE NON-SUPERSTAR

Outcome	Non-college	College
Panel A. Population relative to new housing		
Migrants	0.638	0.327
Existing households <p50	−0.168	−0.158
Existing households <p20	−0.131	−0.072
Homeless	−0.027	−0.008
Panel B. Other outcomes		
Housing costs	−0.060	−0.053
Wages	−0.005	0.012
Amenities	0.000	0.000
Welfare	0.040	0.056

Notes: The results are counterfactuals within a model fit to the Los Angeles metro housing market in 2019. Detail on the counterfactual experiment appears in the note to Figure 5. In Panel A, I report the change in each outcome divided by the change in the housing stock. In Panel B, housing costs equal the average over all quality percentiles weighted by the initial population of non-college or college households. I normalize welfare by the migration elasticities to convert it into consumption-equivalent units.

References

Aliprantis, Dionissi, and Daniel Hartley. 2015. "Blowing it up and knocking it down: The local and city-wide effects of demolishing high concentration public housing on crime." *Journal of Urban Economics*, 88: 67–81.

Almagro, Milena, Eric Chyn, and Bryan A. Stuart. 2022. "Urban Renewal and Inequality: Evidence from Chicago's Public Housing Demolitions." Becker Friedman Institute Working Paper 2022-163.

Anagol, Santosh, Fernando Ferreira, and Jonah Rexer. 2023. "Estimating the Economic Value of Zoning Reform." University of Pennsylvania Working Paper.

Anenberg, Elliot, and Edward Kung. 2020. "Can More Housing Supply Solve the Affordability Crisis? Evidence from a Neighborhood Choice Model." *Regional Science and Urban Economics*, 80.

Asquith, Brian J., Evan Mast, and Davin Reed. 2023. "Local Effects of Large New Apartment Buildings in Low-Income Areas." *Review of Economics and Statistics*, 105(2): 359–375.

Baum-Snow, Nathaniel, and Justin Marion. 2009. "The Effects of Low Income Housing Tax Credit Developments on Neighborhoods." *Journal of Public Economics*, 93(5-6): 654–666.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy*, 115(4): 588–638.

Blanco, Hector. 2023. "Pecuniary Effects of Public Housing Demolitions: Evidence from Chicago." *Regional Science and Urban Economics*, 98.

Braid, Ralph M. 1981. "The Short-Run Comparative Statics of a Rental Housing Market." *Journal of Urban Economics*, 10(3): 286–310.

Card, David. 2009. "Immigration and Inequality." *American Economic Review*, 99(2): 1–21.

Chetty, Raj, and Adam Szeidl. 2007. "Consumption Commitments and Risk Preferences." *Quarterly Journal of Economics*, 122(2): 831–877.

Collinson, Robert, and Peter Ganong. 2018. "How Do Changes in Housing Voucher Design Affect Rent and Neighborhood Quality?" *American Economic Journal: Economic Policy*, 10(2): 62–89.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook of Regional and Urban Economics*. Vol. 5A, , ed. Gilles Duranton, J. Vernon Henderson and William C. Strange, Chapter 5, 247–348. Amsterdam:North-Holland.

Couture, Victor, Cecile Gaubert, Jessie Handbury, and Erik Hurst. 2023. "Income Growth and the Distributional Effects of Spatial Sorting." *Review of Economic Studies*.

Damiano, Anthony, and Chris Frenier. 2020. "Build Baby Build?: Housing Submarkets and the Effects of New Construction on Existing Rents." University of Minnesota Working Paper.

Davis, Donald R., and Jonathan I. Dingel. 2020. "The Comparative Advantage of Cities." *Journal of International Economics*, 123.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review*, 106(3): 479–524.

Diamond, Rebecca, and Tim McQuade. 2019. "Who Wants Affordable Housing in their Backyard? An Equilibrium Analysis of Low Income Property Development." *Journal of Political Economy*, 127(3): 1063–1117.

Diamond, Rebecca, Tim McQuade, and Franklin Qian. 2019. "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco." *American Economic Review*, 109(9): 3365–3394.

Epple, Dennis, Luis Quintero, and Holger Sieg. 2020. "A New Approach for Estimating Equilibrium Models for Metropolitan Housing Markets." *Journal of Political Economy*, 128(3): 948–983.

Eriksen, Michael D., and Stuart S. Rosenthal. 2010. "Crowd Out Effects of Place-Based Subsidized Rental Housing: New Evidence from the LIHTC Program." *Journal of Public Economics*, 94(11-12): 953–966.

Evans, William N., James X. Sullivan, and Melanie Wallskog. 2016. "The Impact of Homelessness Prevention Programs on Homelessness." *Science*, 353: 694–699.

Favilukis, Jack, Pierre Mabille, and Stijn Van Nieuwerburgh. 2023. "Affordable Housing and City Welfare." *Review of Economic Studies*, 90(1): 293–330.

Ganong, Peter, and Daniel Shoag. 2017. "Why Has Regional Income Convergence in the U.S. Declined?" *Journal of Urban Economics*, 102: 76–90.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2013. "Human Capital and Regional Development." *Quarterly Journal of Economics*, 128(1): 105–164.

Glaeser, Edward L., and Joseph Gyourko. 2003. "The Impact of Building Restrictions on Housing Affordability." *Economic Policy Review*, 9(2): 21–39.

Glaeser, Edward L., Joseph Gyourko, and Raven Saks. 2005. "Why Is Manhattan So Expensive? Regulation and the Rise in Housing Prices." *Journal of Law and Economics*, 48(2): 331–369.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race Between Education and Technology.* Cambridge, MA:Belknap Press.

Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. "Superstar Cities." *American Economic Journal: Economic Policy*, 5(4): 167–199.

Landvoigt, Tim, Monika Piazzesi, and Martin Schneider. 2015. "The Housing Market(s) of San Diego." *American Economic Review*, 104(4): 1371–1407.

Li, Xiaodi. 2022. "Do new housing units in your backyard raise your rents?" *Journal of Economic Geography*, 22(6): 1309–1352.

Locks, Gedeão, and Josselin Thuilliez. 2023. "The Impact of Minimum Income on Homelessness: Evidence from France." *Journal of Urban Economics*, 135.

Lucas, Jr., Robert E. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1): 3–42.

Määttänen, Niku, and Marko Terviö. 2014. "Income Distribution and Housing Prices: An Assignment Model Approach." *Journal of Economic Theory*, 151: 381–410.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2022. "IPUMS National Historical Geographic Information System: Version 17.0." Minneapolis, MN: IPUMS (accessed August 15 and 17, 2023).

Mast, Evan. 2023. "The Effect of New Market-Rate Housing Construction on the Low-Income Housing Market." *Journal of Urban Economics*, 133.

Mense, Andreas. 2023. "Secondary Housing Supply." IAB Nuremberg Working Paper.

Meyer, Bruce D., Angela Wyse, and Kevin Corinth. 2023. "The size and Census coverage of the U.S. homeless population." *Journal of Urban Economics*, 136.

Missouri Census Data Center. 2010. "Geocorr 2000: Geographic Correspondence Engine, Version 1.3.3." https://mcdc.missouri.edu/applications/geocorr2000.html (accessed June 15, 2022).

Molloy, Raven, Charles G. Nathanson, and Andrew D. Paciorek. 2022. "Housing supply and affordability: Evidence from rents, housing consumption and household location." *Journal of Urban Economics*, 129.

Moretti, Enrico. 2004*a*. "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics*, 121(1-2): 175–212.

Moretti, Enrico. 2004*b*. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review*, 94(3): 656–690.

Pennington, Kate. 2021. "Does Building New Housing Cause Displacement?: The Supply and Demand Effects of construction in San Francisco." University of California Working Paper.

Reed, William J. 2003. "The Pareto Law of Incomes—An Explanation and an Extension." *Physica A: Statistical Mechanics and its Applications*, 319: 469–486.

Reed, William J., and Murray Jorgensen. 2004. "The Double Pareto-Lognormal Distribution—A New Parametric Model for Size Distributions." *Communications in Statistics - Theory and Methods*, 33(8): 1733–1753.

Rosenthal, Stuart S. 2014. "Are Private Markets and Filtering a Viable Source of Low-Income Housing? Estimates from a "Repeat Income" Model." *American Economic Review*, 104(2): 687–706.

Rosenthal, Stuart S. 2020. "Owned Now Rented Later? The Effect of Housing Stock Transitions on Affordable Housing and Market Dynamics." Working Paper, Syracuse University.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler. 2023. "IPUMS USA: Version 13.0." Minneapolis, MN: IPUMS (accessed June 15, 2022 and May 11, August 3, and August 7, 2023).

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply." *Quarterly Journal of Economics*, 125(3): 1253–1296.

Schwartz, Amy Ellen, Ingrid Gould Ellen, Ioan Voicu, and Michael H. Schill. 2006. "The External Effects of Place-Based Subsidized Housing." *Regional Science and Urban Economics*, 36(6): 679–707.

Sweeney, James L. 1974. "Quality, Commodity Hierarchies, and Housing Markets." *Econometrica*, 42(1): 147–167.

U.S. Census Bureau. 2010. "2010 Census Tract to 2010 PUMA Relationship File." https://www2.census.gov/geo/docs/maps-data/data/rel/2010_Census_Tract_to_2010_PUMA.txt (accessed June 23, 2021).

U.S. Census Bureau. 2012. "2010 Census Summary File 1: Technical Documentation." Washington, DC.

U.S. Census Bureau. 2013. "2010 Census Geography Notes and Errata." Washington, DC.

U.S. Census Bureau. 2014. "American Community Survey Design and Methodology (January 2014)." Version 2.0 ed., Washington, DC.

U.S. Census Bureau. 2018. "List 1. Core Based Statistical Areas (CBSAs), Metropolitan Divisions, and Combined Statistical Areas (CSAs), September 2018." https://www2.census.gov/programs-surveys/metro-micro/geographies/reference-files/2018/delineation-files/list1_Sep_2018.xls (accessed August 15, 2023).

For Online Publication: Internet Appendix

A Data build

A.1 NHGIS

A.1.1 2000

For 2000, I download data at the level of a block-group part from NHGIS (Manson et al., 2022). A block group part is a refinement of a block-group, as it represents a block group intersected with other geographies such as places or county subdivisions. I use block-group parts because they map more cleanly into geographies in 2019 than do block groups.

I merge block-group parts to contemporaneous PUMA identifiers using a crosswalk from the Missouri Census Data Center's Geocorr program (Missouri Census Data Center, 2010). The block-group parts nest within PUMAs.

To map block-group parts to 2018 CBSA boundaries, I first I map block-group parts to 2010 US counties using a crosswalk from NHGIS. In cases where a block-group part matches to more than one 2010 US county, I select the county to which the NHGIS crosswalk assigns the greatest share of the housing units of that block-group. I verify that this county always exists (i.e., there are no ties among counties with the greatest allocated share) when the block-group part has positive housing units in 2000. I then map 2010 US counties to 2018 CBSAs using a delineation file from the US Census Bureau (U.S. Census Bureau, 2018). Each county is either a rural county that maps to no CBSA, or it maps to a unique CBSA.

I use data on the number of housing units by year built and by occupancy type. Year built is binned as follows: on or before 1939, 1940–1949, 1950–1959, 1960–1969, 1970–1979, 1980–1989, 1990–1994, 1995–1998, and 1999–2000 (specifically March 2000, as census day is April 1 2000). Occupancy type is owner occupied, renter occupied, and vacant. Counts of housing units by year built and occupancy for the entire US appear in Appendix Table A1. When I count housing units built before 2000, I multiply housing units built in the 1999–2000 window by 0.8, as 20% of the time window includes 2000.

I also use data on the number of units by housing cost. For owner-occupied units, housing cost is the occupant's estimate of the market value of the home, which is binned as shown in Table A3, which reports the counts of owner-occupied units for the US in each bin. For renter-occupied units, I use data on monthly gross rents, which includes utilities such as electricity. These data are binned as shown in Table A4, which reports counts of renter-occupied units in each bin. Data on rents is not available for units rented without cash rent and for non-specified units, which are rented single-family homes on 10 or more acres. The counts of units in each category also appear in Table A4.

Finally, I use data on the number of persons in households and the number of persons in group quarters.

A.1.2 2019

For 2019, I download data for the 2015-2019 ACS at the level of a block-group from NHGIS. I drop observations in Puerto Rico.

To match these data to PUMAs, I use a crosswalk from the Census Bureau going from tracts to PUMAs (U.S. Census Bureau, 2010). Each block group is fully contained in a single census tract, so this crosswalk maps block groups to PUMAs. However, the crosswalk uses tract definitions as of 2010, whereas the NHGIS data have tract definitions as of 2019, which correct for some minor errors that appear in the 2010 Census. These errors are documented in U.S. Census Bureau (2013). In particular, I correct the tract numbering for several tracts in Madison County, NY, Oneida County, NY, Pima County, AZ, and Los Angeles County, CA. I also split a block group in Los Angeles County between block groups in different tracts proportionally using the ratio of housing units in the 2010 Census. Finally, I correct the county code changes, which are delineated on the Census Bureau's website, for the following counties: Kusilvak Census Area, AK, Oglala Lakota County, SD, and Bedford City, VA.

To match block groups to CBSAs, I match them to counties using the definitions in the NHGIS data. I then match those counties to CBSAs using the county to CBSA crosswalk mentioned above. I similarly match tracts, which nest within counties, to CBSAs.

As with 2000, I use housing unit counts by year built and occupancy; summary statistics appear in Table A1. The year-built bins differ slightly from 2000: there is now a single bin for the 1990s, and there are three separate bins for housing built on or after 2000. I also use housing counts by home value for owner-occupied units and gross rent for renter-occupied units; summary statistics appear respectively in Tables A3 and A4. For home values, the bins are the same as 2000 except there are now three separate bins for values at or above \$1,000,000 instead of one. Similarly, for rents the bins are the same except there are now four separate bins for rents at or above \$2,000. Another difference is that rents are reported even for units that would qualify as non-specified in the 2000 Census. Finally, I again use data on the number of persons in households and the number of persons in group quarters.

To assign superstar metro status, I also use data at the CBSA level from NHGIS on aggregate housing values for owner-occupied units. I use these data to calculate the mean value of owner-occupied housing in 2019. In 2000, these data are available at the level of block-group parts, but the Census Bureau suppresses aggregate housing values at the block-group level for many block groups (and even at the county level for some counties). Therefore, I use the data directly at the CBSA level in NHGIS.

A.1.3 Crosswalk between survey years

To map 2000 block-group parts to 2019 tracts, I start with the 2000 block-group part to 2010 tract crosswalk from NHGIS. This crosswalk maps each block-group part to one or more tracts and gives the share of the housing units in the block-group part lying in each matching tract. I map both 2000 block-group parts and 2010 tracts to 2018 CBSA boundaries as described above. I then keep 2000 block-group part to 2010 tract matches only if both lie in the same CBSA and if the NHGIS crosswalk allocates a positive share of housing units from the block-group part to that tract. I verify that all block-group parts in 2000 with positive housing units have at least one such matching tract. If multiple matches remain, I proportionally allocate housing units from the 2000 block-group to the matching tracts weighting by the NHGIS allocation factor.

A.2 IPUMS

I use data from the 5% sample of the 2000 US Census and the 2015–2019 ACS from IPUMS (Ruggles et al., 2023). These data are at the person, household, or housing unit level and have PUMA identifiers.

For persons in group quarters (gq code of 3 or 4 in IPUMS), I impute a status of homelessness if the following conditions are met: the group quarters type is non-institutionalized (gq code of 4), the age is at least 18, the person is not in school (school code of 1), and the employment status is unemployed or not in the labor force (employment status code of 2 or 3). To measure homeless counts in each CBSA, I multiply the count of group quarters persons in NHGIS by the share in IPUMS of all group quarters persons in the same PUMA and survey year that have this imputed homelessness status. I then aggregate to the CBSA level.

For all housing units, I use data on year built and occupancy type. For 2000, year built is binned in IPUMS in the same way as in NHGIS. For 2019, year built is binned in the same way in IPUMS and NHGIS for units built before 2000. For units built on or after 2000, the bins in IPUMS are finer than the bins in NHGIS. For these units, I assign year built bins using the same classification as in NHGIS. Aggregate counts by occupancy, survey year, and year built appear in Table A2. By and large, these numbers are close to the corresponding counts in NHGIS shown in Table A1.

I also use data on home values for owner-occupied units and gross rents for renter-occupied units. To enable comparison with NHGIS, I assign bins using the same classification in the corresponding year in NHGIS. For renter-occupied units in 2000, I also tag non-specified units using the units-in-structure and acreage variables in IPUMS. The counts of housing units in IPUMS in each bin appear for owner-occupied units in Table A3 and for renter-occupied units in Table A4. The numbers are again close between IPUMS and NHGIS between the corresponding bins.

I use the un-binned home value and gross rent data when calculating user costs using the GMM specification in Section 3. In that specification, I do not use renter-occupied units without cash rent, and I do not use non-specified rental units in 2000. I also use the un-binned data to assign values and rents to each in in NHGIS. Specifically, for each survey year I take the average across the entire US of home values and rents by bin, again excluding non-specified units in 2000 and no-cash rent units in both survey years. I use each average as the value or rent for each corresponding bin when analyzing housing costs in NHGIS.

To analyze heterogeneity by college degree, I tag a household as college-educated if the educational attainment of the household head is at least 4 years of college (educ code of 10 or 11). To measure household income, I use the hhincome variable in IPUMS, except for a small number of households in 2000 with group quarters code equal to 5, for whom I calculate household income by summing total income (inctot) across household members who are at least 15 years old. When analyzing occupied housing units in NHGIS, I impute college status using the mean college status by survey year, PUMA, occupancy type, and housing cost bin in IPUMS. If there are no such units in IPUMS, I use the mean by survey year, PUMA, and occupancy type.

Finally, I use CPI-U values from IPUMS to inflation-adjust home values and rents when calculating real growth rates between survey years. The adjustment factor equals 1 for 2000 and 0.652 for the 2015-2019 survey.

B Proofs

B.1 Proposition 1

When the price of a type of housing is positive, $p_j > 0$, rentiers optimize by selling all h_j units of housing of quality q_j. When the price of that housing is 0, $p_j = 0$, rentiers optimize by selling any quantity of that quality of housing. As I now demonstrate, this observation allows me to prove the first two statements in Proposition 1.

Assumption 1 implies that some housing remains vacant in equilibrium, as the number of households, N, is less than the quantity of housing in the metro, $\sum_{j=0}^{J} h_j$. The equilibrium price of such housing must be 0, given the rentier optimization just described. Suppose, for a contradiction, that the equilibrium price of homeless shelters is positive: $p_0 > 0$. Then there must exist another index, $j > 0$, such that price of housing of quality q_j is 0: $p_j = 0$. In this case, all households strictly prefer choosing quality q_j over the homeless shelter of quality q_0, as the quality is higher and the price is lower. As a result, no households choose the homeless shelter, which contradicts market clearing, as rentiers sell all h_0 units of the housing of quality q_0 because the price of such housing, p_0, is positive. This contradiction proves that in equilibrium, the price of the homeless shelter must be 0: $p_0 = 0$.

Assumption 1 also implies that some households must choose the homeless shelter, as there is an insufficient quantity of standard housing units for all households: $\sum_{j=1}^{J} h_j < N$. As a result, the prices of all standard housing units must be positive: $p_j > 0$ for $j > 0$. To prove this statement, I assume the opposite to arrive at a contradiction, namely, that there is a $j > 0$ such that the price of the corresponding housing unit is 0: $p_j = 0$. Then all households strictly prefer to choose housing of quality q_j over the inferior quality homeless shelter of quality q_0, implying that no households choose the homeless shelter, a contradiction. Therefore, the price of all standard housing units is positive: $p_j > 0$ for $j > 0$. Given the optimizing behavior of rentiers, the market for standard housing units clears only when households occupy all of the standard housing units, which proves that the housing demanded by households equals h_j for $j > 0$.

The fact that households occupy all standard housing units in equilibrium allows me to prove another useful fact: that prices strictly increase in quality among standard housing units in equilibrium. For a contradiction, suppose this monotonicity condition fails, so that there are indices j and j', both greater than 0, such that $j < j'$ but $p_j \geq p_{j'}$. In this case, all households strictly prefer choosing housing of quality $q_{j'}$ over that of quality q_j, meaning that no households choose housing of quality q_j, contradicting the result that households occupy all standard housing units.

Finally, I prove the final statement of Proposition 1, that a household's housing quality weakly increases in its labor endowment, z, within each education group e. I prove this statement by contradiction as well. Consider two households with the same education e, labor endowments z and z', and equilibrium housing choices q_j and $q_{j'}$. Suppose that the former household has a smaller labor endowment but makes a larger quality choice: $z < z'$ and $j > j'$.

In this case, both households can afford both housing units: $w_e z$ and $w_e z'$ are both larger than p_j and $p_{j'}$. By assumption, both households an afford the houses they choose. The household of endowment z can afford the house of quality $q_{j'}$ because it has a lower quality, and hence price, than the house of q_j that he can afford. The household of endowment z' can afford the house of quality q_j because the household of endowment z, who has a smaller income, can afford it. Given that both households can afford both houses, optimization by the household of endowment z implies the inequality

$$\log(w_e z - p_j) + \omega_e q_j \geq \log(w_e z - p_{j'}) + \omega_e q_{j'},$$

and optimization by the household of endowment z' implies the inequality

$$\log(w_e z' - p_{j'}) + \omega_e q_{j'} \geq \log(w_e z' - p_j) + \omega_e q_j.$$

Rearranging these inequalities implies that

$$\log(w_e z - p_{j'}) - \log(w_e z - p_j) \leq \omega_e(q_j - q_{j'}) \leq \log(w_e z' - p_{j'}) - \log(w_e z' - p_j).$$

Exponentiating the terms in the outer inequality, cross-multiplying, and then collecting terms yields

$$0 \leq w_e(z' - z)(p_{j'} - p_j),$$

a contradiction because $z < z'$ and $p_{j'} < p_j$. Therefore, the equilibrium quality choices must weakly increase in labor endowment within each education group, as claimed.

B.2 Solution to Eqs. (8) and (9)

Eq. (8) gives $2J$ equations, one for each combination of education, e, and index of standard housing units, j. Eq. (9) gives J equations, one for each index of standard housing units, j. The unknowns are the J prices of the standard housing units, p_j for $j > 0$, and the $2J$ endowment cutoffs, $z_{L,j}$ and $z_{H,j}$ for $j > 0$.

I show that a unique solution exists to this system of equations as long as the population distributions, $n_L(z)$ and $n_H(z)$, are atomless distributions with full support on $(0, \infty)$. This condition holds in any equilibrium: given household optimization across metros, Eq. (7) holds. This equation implies that $n_e(z)$ is atomless and has full support on $(\min_j p_j, \infty)$, as the distribution $\widetilde{n}_e(z)$ is atomless and has full support on $(0, \infty)$ by assumption. As shown in Proposition 1, $p_0 = 0$ in equilibrium, so the support of $n_e(z)$ is $(0, \infty)$, as claimed.

Using (8), I solve for each endowment cutoff, $z_{e,j}$, as a function of prices:

$$z_{e,j} = w_e^{-1}\left(q_j^{\omega_e} p_j - q_{j-1}^{\omega_e} p_{j-1}\right).$$

Substituting this expression into Eq. (9) yields:

$$\sum_{j'=j}^{J} h_{j'} = \sum_{e \in \{L,H\}} \int_{w_e^{-1}\left(q_j^{\omega_e} p_j - q_{j-1}^{\omega_e} p_{j-1}\right)}^{\infty} n_e(z) dz.$$

Given p_{j-1}, the right side of this equation strictly decreases in p_j as this price runs from p_{j-1} to ∞, because $n_L(z)$ and $n_H(z)$ are atomless distributions with full support on $(0, \infty)$. Furthermore, as p_j runs over this interval, the right side of this equation decreases from N to an asymptote of 0. The left side of this equation is less than N due to Assumption 1. Therefore, given p_{j-1}, a unique solution to this equation for p_j exists, and the value of p_j that solves the equation is greater than p_j. Proposition 1 shows that $p_0 = 0$ in equilibrium. Therefore, by induction, a unique set of prices p_j solve this system of equations, and prices increase in this solution. These prices then uniquely pin down the cutoffs, $z_{e,j}$, proving that the solution to the original system of $3J$ equations is unique, as claimed.

B.3 Proposition 2

To prove the proposition, I provide equations such that model parameters and the objects in the proposition determine all the coefficients of these equations. The unknowns in the equations determine all model outcomes, and the number of equations equals the number of unknowns. The other terms in the equations are the exogenous shocks: ∂h_j, $\partial \log \widetilde{A}_e$, and $\partial \log \widetilde{a}_e$.

Before listing the equations, I provide a useful fact about transforming variables from the labor endowment, z, to income, y. The measure of households of education e with endowment less than z equals $N_e \int_0^{w_e z} f_e(y) dy$, and by definition, the derivative of this measure with respect to z equals $n_e(z)$. Therefore:

$$n_e(z) = w_e N_e f_e(w_e z). \tag{IA1}$$

Differentiating the log of Eq. (2) yields:

$$\partial \log w_e = \frac{(1-\rho)\left(A_L^\rho Z_L^\rho \partial \log Z_L + A_L^\rho Z_L^\rho \partial \log A_L + A_H^\rho Z_H^\rho \partial \log Z_H + A_H^\rho Z_H^\rho \partial \log A_H\right)}{(A_L Z_L)^\rho + (A_H Z_H)^\rho}$$
$$+ \rho \partial \log A_e + (\rho - 1)\partial \log Z_e.$$

I let Y_e denote the aggregate income for households of education e, so that $Y_e = w_e Z_e$. I also let $Y = Y_L + Y_H$ denote the aggregate income for all households. Eq. (2) implies that

$$\frac{Y_e}{Y} = \frac{(A_e Z_e)^\rho}{(A_L Z_H)^\rho + (A_H Z_H)^\rho}$$

for each e. Therefore, the expression for $\partial \log w_e$ yields two equations:

$$\partial \log w_e = \left(1 - (1-\rho)\frac{Y_{\sim e}}{Y}\right)\partial \log A_e + (1-\rho)\frac{Y_{\sim e}}{Y}\partial \log A_{\sim e} - (1-\rho)\frac{Y_{\sim e}}{Y}\partial \log Z_e + (1-\rho)\frac{Y_{\sim e}}{Y}\partial \log Z_{\sim e} \quad \text{(IA2)}$$

for each e, where $\sim e$ denotes the opposite eduction: $\sim L = H$ and $\sim H = L$.

Differentiating the log of Eq. (3) yields two equations:

$$\partial \log A_e = \partial \log \widetilde{A}_e + (\gamma_N - \gamma_H)\,\partial \log N + \gamma_H \partial \log N_H \quad \text{(IA3)}$$

for each e.

Differentiating the log of Eq. (4) yields two equations:

$$\partial \log a_e = \partial \log \widetilde{a}_e + \gamma_{a,e} \partial \log N_H - \gamma_{a,e} \partial \log N_L \quad \text{(IA4)}$$

for each e.

Differentiating the log of Eq. (8) yields:

$$\frac{w_e z_{e,j} \partial \log z_{e,j} + w_e z_{e,j} \partial \log w_e - \partial p_j}{w_e z_{e,j} - p_j} = \frac{w_e z_{e,j} \partial \log z_{e,j} + w_e z_{e,j} \partial \log w_e - \partial p_{j-1}}{w_e z_{e,j} - p_{j-1}}.$$

Substituting the expression for the income cutoffs, $y_{e,j} = w_e z_{e,j}$, yields J equations:

$$\partial p_j = \frac{y_{e,j} - p_j}{y_{e,j} - p_{j-1}}\partial p_{j-1} + \frac{y_{e,j}(p_j - p_{j-1})}{y_{e,j} - p_{j-1}}\partial \log z_{e,j} + \frac{y_{e,j}(p_j - p_{j-1})}{y_{e,j} - p_{j-1}}\partial \log w_e \quad \text{(IA5)}$$

for each $j > 0$.

Differentiating Eq. (9) yields:

$$\sum_{j'=j}^{J} \partial h_{j'} = \sum_e -n_e(z_{e,j}) z_{e,j} \partial \log z_{e,j} + \int_{z_{e,j}}^{\infty} \partial \log n_e(z) n_e(z) dz.$$

Substituting Eq. (10) yields:

$$\sum_{j'=j}^{J}\partial h_{j'} = -\sum_{e} n_e(z_{e,j})z_{e,j}\partial\log z_{e,j} - \left(\sum_{e}\int_{z_{e,j}}^{\infty}\xi_e\beta_e n_e(z)dz\right)\partial\log r + \sum_{e}\left(\int_{z_{e,j}}^{\infty} n_e(z)dz\right)\partial\log a_e$$
$$+\sum_{e}\left(\sum_{j'=j}^{J}\int_{z_{e,j'}}^{z_{e,j'+1}}\frac{\beta_e w_e z}{w_e z - p_{j'}}n_e(z)dz\right)\partial\log w_e - \sum_{j'=j}^{J}\left(\sum_{e}\int_{z_{e,j'}}^{z_{e,j'+1}}\frac{\beta_e}{w_e z - p_{j'}}n_e(z)dz\right)\partial p_{j'},$$

where $z_{e,J+1}$ is defined formally as ∞ for each e. Substituting Eq. (IA1) and changing variables in the integrals from z to y yields $2J$ equations:

$$\sum_{j'=j}^{J}\partial h_{j'} = -\sum_{e}\big(N_e y_{e,j} f_e(y_{e,j})\big)\partial\log z_{e,j} - \left(\sum_{e}\int_{y_{e,j}}^{\infty}\xi_e\beta_e N_e f_e(y)dy\right)\partial\log r$$
$$+\sum_{e}\left(\int_{y_{e,j}}^{\infty}N_e f_e(y)dy\right)\partial\log a_e + \sum_{e}\left(\sum_{j'=j}^{J}\int_{y_{e,j'}}^{y_{e,j'+1}}\frac{\beta_e N_e y}{y-p_{j'}}f_e(y)dy\right)\partial\log w_e \tag{IA6}$$
$$-\sum_{j'=j}^{J}\left(\sum_{e}\int_{y_{e,j'}}^{y_{e,j'+1}}\frac{\beta_e N_e}{y-p_{j'}}f_e(y)dy\right)\partial p_{j'}$$

for each e and $j > 0$.

To obtain expressions for the changes in aggregate populations, I differentiate the log of $N_e = \int_0^\infty n_e(z)dz$ and substitute Eq. (10) to obtain:

$$\partial\log N_e = -\xi_e\beta_e\partial\log r + \partial\log a_e + \left(\sum_{j=0}^{J}\int_{z_{e,j}}^{z_{e,j+1}}\frac{\beta_e w_e z n_e(z)dz}{N_e(w_e z - p_j)}\right)\partial\log w_e - \sum_{j=0}^{J}\left(\int_{z_{e,j}}^{z_{e,j+1}}\frac{\beta_e n_e(z)dz}{N_e(w_e z - p_j)}\right)\partial p_j.$$

Substituting Eq. (IA1) and changing variables from z to y gives two equations:

$$\partial\log N_e = -\xi_e\beta_e\partial\log r + \partial\log a_e + \left(\sum_{j=0}^{J}\int_{y_{e,j}}^{y_{e,j+1}}\frac{\beta_e y f(y)dy}{y-p_j}\right)\partial\log w_e - \sum_{j=0}^{J}\left(\int_{y_{e,j}}^{y_{e,j+1}}\frac{\beta_e f(y)dy}{y-p_j}\right)\partial p_j \tag{IA7}$$

for each e. Differentiating the log of $N = N_L + N_H$ yields one equation:

$$\partial\log N = \left(\frac{N_L}{N}\right)\partial\log N_L + \left(\frac{N_H}{N}\right)\partial\log N_H. \tag{IA8}$$

To obtain expressions for the changes in the aggregate labor endowments, I differentiate the log of $Z_e = \int_0^\infty z n_e(z)dz$ and substitute Eq. (10) to obtain:

$$\partial\log Z_e = -\xi_e\beta_e\partial\log r + \partial\log a_e + \left(\sum_{j=0}^{J}\int_{z_{e,j}}^{z_{e,j+1}}\frac{\beta_e w_e z^2 n_e(z)dz}{Z_e(w_e z - p_j)}\right)\partial\log w_e - \sum_{j=0}^{J}\left(\int_{z_{e,j}}^{z_{e,j+1}}\frac{\beta_e z n_e(z)dz}{Z_e(w_e z - p_j)}\right)\partial p_j.$$

Substituting Eq. (IA1) and changing variables from z to y gives two equations:

$$\begin{aligned}\partial \log Z_e = -\xi_e \beta_e \partial \log r + \partial \log a_e + \left(\sum_{j=0}^{J} \int_{y_{e,j}}^{y_{e,j+1}} \frac{\beta_e N_e y^2 f_e(y) dy}{Y_e (y - p_j)} \right) \partial \log w_e \\ - \sum_{j=0}^{J} \left(\int_{y_{e,j}}^{y_{e,j+1}} \frac{\beta_e N_e y f_e(y) dy}{Y_e (y - p_j)} \right) \partial p_j \end{aligned} \tag{IA9}$$

for each e.

Differentiating the log of Eq. (6) yields one equation:

$$\partial \log r = \sum_{j=1}^{J} \left(\frac{1}{J} \right) \partial \log p_j. \tag{IA10}$$

Eqs. (IA2)–(IA10) provide $2 + 2 + 2 + J + 2J + 2 + 1 + 2 + 1 = 3J + 12$ equations in the following $3J + 12$ unknowns: $\partial \log z_{e,j}$, ∂p_j, $\partial \log w_e$, $\partial \log A_e$, $\partial \log a_e$, $\partial \log N_e$, $\partial \log Z_e$, $\partial \log N$, and $\partial \log r$, where $e \in \{L, H\}$ and $j \in \{1, ..., J\}$. The objects p_j, N_e, $f_e(y)$, $y_{e,j}$, ρ, γ_N, γ_H, $\gamma_{a,e}$, β_e, ξ_e, and J determine the coefficients, as claimed.

C Calculating parameter values

C.1 Parameter space restriction for estimation

I describe a non-empty set Θ such that if $\theta \in \Theta$, then θ uniquely determines the income cutoffs, $y_{e,j}$, given estimates of house prices, p_j, and housing stocks, h_j, and that these income cutoffs increase in the quality index , j, for each education, e. To emphasize that the income distributions $f_e(y)$ depend on θ, I denote them here as $f_e(y; \theta_e)$.

I define Θ using two conditions. The first condition is that the following inequality holds for each $j > 0$:

$$\sum_{j'=j}^{J} \frac{h_{j'}}{N} \leq \sum_{e} \frac{N_e}{N} \int_{p_j}^{\infty} f_e(y; \theta_e) dy. \tag{IA11}$$

I verify numerically that there exist values for θ_L and θ_H such that this inequality holds.

I first show that when this inequality holds, unique solutions for the income cutoffs, $y_{e,j}$, exist. To do so, I first express the cutoffs for college households as functions as the cutoffs for non-college households. I solve for q_{j-1}/q_j in Eq. (8) and then change variables from z to y to obtain:

$$\left(\frac{y_{H,j} - p_j}{y_{H,j} - p_{j-1}} \right)^{\omega_H} = \left(\frac{y_{L,j} - p_j}{y_{L,j} - p_{j-1}} \right)^{\omega_L}.$$

I rewrite these expressions as:

$$\left(1 - \frac{p_j - p_{j-1}}{y_{H,j} - p_{j-1}} \right)^{\omega_H} = \left(1 - \frac{p_j - p_{j-1}}{y_{L,j} - p_{j-1}} \right)^{\omega_L}.$$

Solving for $y_{H,j}$ yields:

$$y_{H,j} = p_{j-1} + \frac{p_j - p_{j-1}}{1 - \left(1 - \frac{p_j - p_{j-1}}{y_{L,j} - p_{j-1}}\right)^{\zeta}}.$$

I denote this solution as $y_{H,j}(y_{L,j};\zeta)$.

Using this solution, I now show that if the inequality in (IA11) holds, then θ uniquely determines the income cutoffs. I divide Eq. (9) by the total number of households, N, and then change variables from z to y to obtain:

$$\sum_{j'=j}^{J} \frac{h_j}{N} = \frac{N_L}{N} \int_{y_{L,j}}^{\infty} f_L(y;\theta)dy + \frac{N_H}{N} \int_{y_{H,j}(y_{L,j};\zeta)}^{\infty} f_H(y;\theta)dy \tag{IA12}$$

for each $j > 0$. This equation strictly increases in $y_{L,j}$ over the interval (p_j, ∞) because the function $y_{H,j}(y_{L,j};\zeta)$ has this property, as is clear from inspection. Furthermore, $y_{H,j}(p_j;\zeta) = p_j$. Therefore, as long as the inequality in (IA11) holds, a unique solution for $y_{L,j}$ exists by the intermediate value theorem. This solution then determines a unique corresponding solution for $y_{H,j}$. I denote these solutions $y_{L,j}(\theta)$ and $y_{H,j}(\theta)$.

The second condition defining Θ is that these solutions, $y_{L,j}(\theta)$ and $y_{H,j}(\theta)$, increase in the quality index, j, for each e. The set of such Θ is non-empty, because this monotonicity condition holds whenever $\zeta = 1$: in this special case, $y_{H,j}(y_{L,j}, 1) = y_{L,j}$, and then it is clear that the solutions to Eq. (IA12) increase in j because the left side decreases in j.

C.2 Elasticities from literature

Diamond (2016) estimates that the elasticity of city population with respect to wages equals 4.026 for non-college workers and 2.116 for college workers. These estimates come from the third specification in Table 5 of her paper. They represent preferences for workers without differential effects for race and immigration status, and they hold for a small city. I denote these estimates by $\beta_{d,e}$. Using Eq. (10), the equation for city population in my model, I equate my model's implied population elasticity with respect to wages to the estimates from Diamond (2016) to obtain:

$$\beta_{d,e} = \beta_e E_e\left(\frac{y}{y - p_j}\right),$$

where the $E_e(\cdot)$ operator denotes the average over households of education e in my estimated model. This average equals 1.2338 for non-college households and 1.1547 for college households, implying values of $\beta_L = 3.263$ and $\beta_H = 1.833$.

Diamond (2016) assumes a value of 0.62 for the share of income spent on both housing and non-tradeable consumption, which she terms local goods. In my model, the share of income going to local goods equals $(p_j + \xi_e(y - p_j))/y$. Equating the average of this share to Diamond's value gives:

$$0.62 = \xi_e + (1 - \xi_e)E_e\left(\frac{p_j}{y}\right).$$

Given the model estimates, the average ratio of house prices to income equals 0.1869 for non-college households and 0.1327 for college households, implying values of $\xi_L = 0.533$ and $\xi_H = 0.562$.

Several papers in labor economics estimate the inverse elasticity of substitution between college and non-college labor to be about 0.7; see the discussion in Card (2009). This inverse elasticity corresponds to $1-\rho$, so $\rho = 0.3$.

Combes and Gobillon (2015) find that the typical estimate in the literature of the elasticity of productivity with respect to population density lies between 0.04 and 0.07. I set $\gamma_N = 0.055$, the midpoint of this range. Moretti (2004*b*) estimates that log output in an industry within a metropolitan area rises about 0.0055 when the college share in other industries in the same area rises by one percentage point. Interpreting this estimate as 100 times the derivative of log productivity with respect to the college share, I obtain the equation $0.55 = \gamma_H \text{CollegeShare}$. The college shares in the two years in the sample in Moretti (2004*b*) are 0.161 and 0.191. Setting CollegeShare equal to the average gives $\gamma_H = 0.097$.

In Figure 6, I explore an alternative calibration in which amenities are endogenous to the metro's college share. In this calibration, I raise $\gamma_{a,L}$ and $\gamma_{a,H}$ above 0 using estimates from Diamond (2016). She calculates elasticities of city population with respect to amenities of 0.274 for non-college workers and 1.012 for college workers. As with the wage elasticities I use above, these estimates come from the third specification of Table 5 of her paper, and they represent preferences for workers without differential effects for race and immigration status and elasticities that hold for a small city. In that specification, she calculates that the elasticity of amenities with respect to the college/non-college ratio is 2.60, with a standard error of 1.13. In my setting, I normalize amenities by assuming a unit elasticity of metro population with respect to amenities, as is apparent from Eq. (10). Therefore, to map her estimates to the values of $\gamma_{a,L}$ and $\gamma_{a,H}$ in my model, I multiply her population elasticities with respect to amenities by the elasticity of amenities with respect to the college/non-college ratio. Doing so gives me values of $\gamma_{a,L} = 0.712$ and $\gamma_{a,H} = 2.631$. However, my model becomes unstable under these parameters, in that the baseline construction experiment, which expands the metro's housing stock by 11.8%, *raises* house prices. To show results for a stable equilibrium, I therefore reduce her estimated value of the elasticity of amenities with repsect to the college/non-college ratio to 2, which is less than one standard error below her point estimate. Doing so produces values of $\gamma_{a,L} = 0.548$ and $\gamma_{a,H} = 2.024$. Therefore, the results in Panel B of Figure 6 illustrate that low-quality construction can reduce rich college welfare under parameters that are plausible because they are close to the point estimates in Diamond (2016).

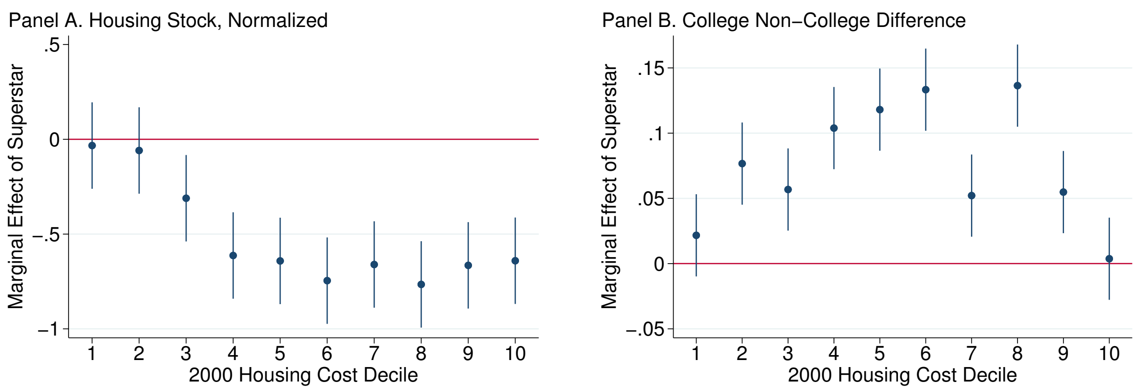

Figure A1. Marginal Effects of Superstar Indicator on 2000–2019 Growth

Notes: Data cover housing units within 2018 CBSA boundaries in the 2000 5% U.S. Census sample and the 2015–2019 ACS. I calculate the results in Panel A of this figure and Panel A of Figure 2 the same way, except that here I normalize the change in the housing stock for each CBSA-decile by the average for that decile across CBSAs, given by the sum of new units and teardowns in Figure 1. I calculate the results in Panel B of this figure the same was as the results in Panels C and D of Figure 2, except that here I use normalized college growth minus normalized non-college growth as the outcome.

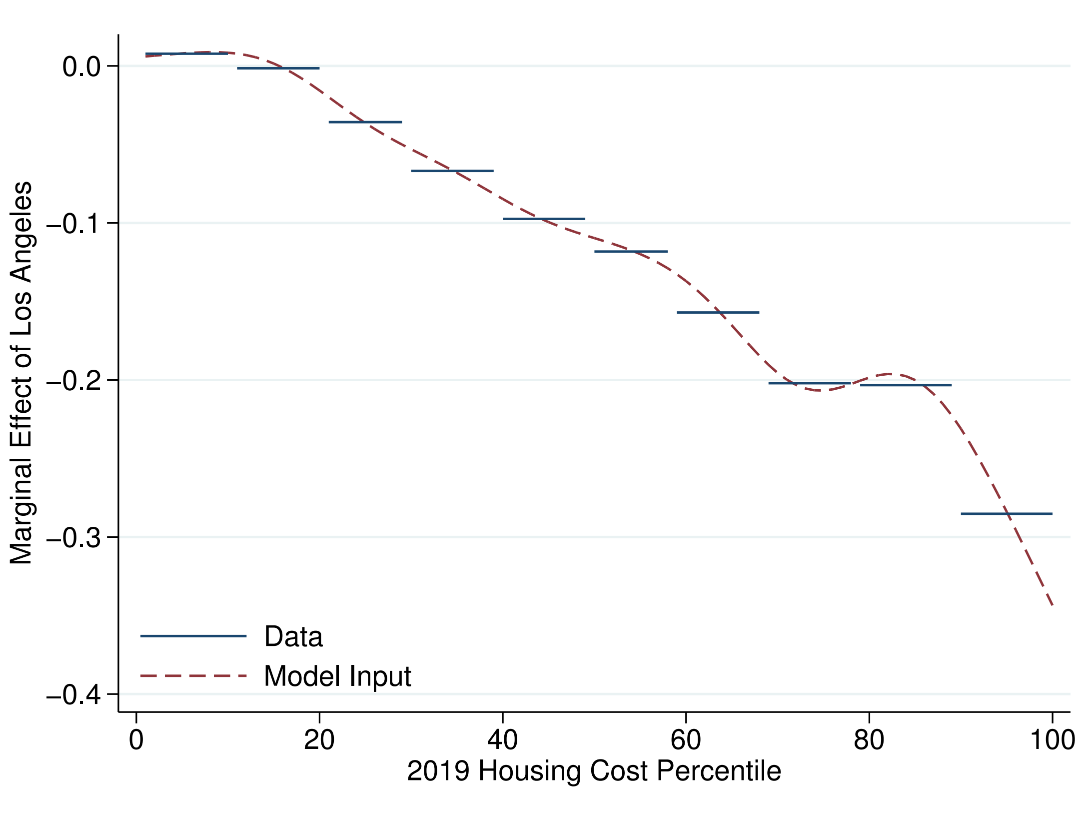

Figure A2. Marginal Effects of Los Angeles on 2000–2019 Housing Stock Growth

Notes: To produce the Data series, I assign 2019 percentiles to 2000 deciles using the shares of 2019 housing units in Los Angeles in each quality decile in the data underlying Figure 1. The effect for each decile matches the targets in Panel B of Table 3. To produce the Model input series, I calculate values for each percentile so that the mean in each decile matches the data while the sum of the squares of the second difference of the series is minimized.

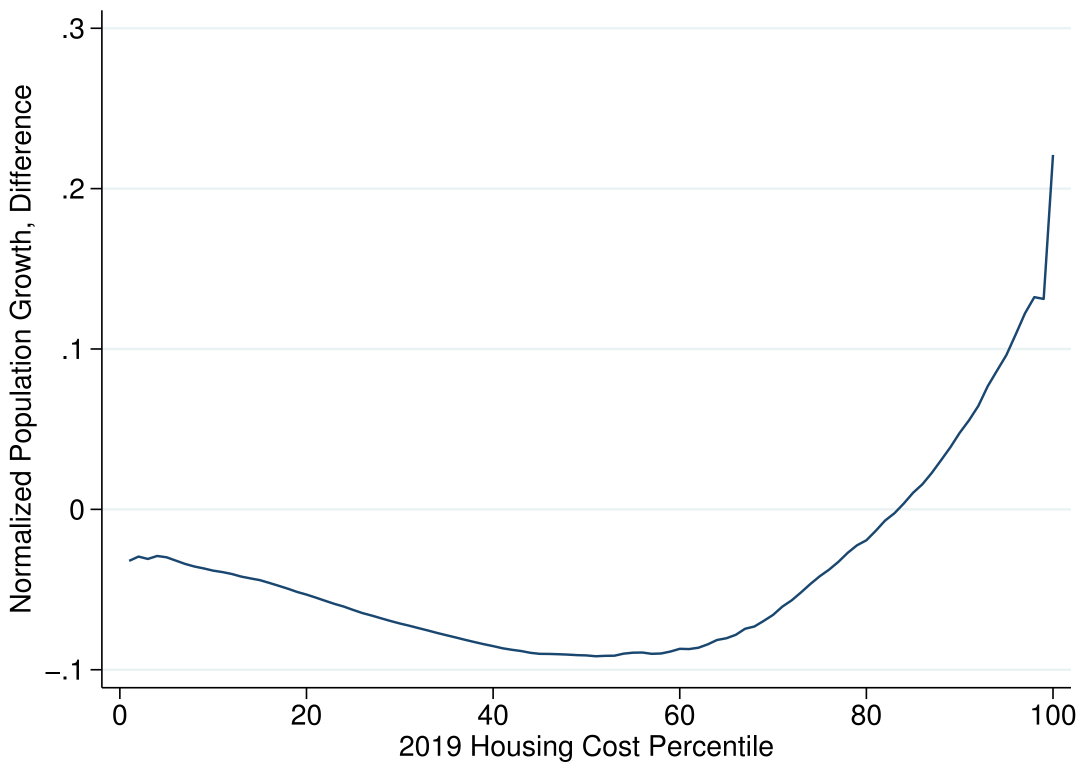

Figure A3. Marginal Effects of Building like Non-Superstar on Gentrification

Notes: The results are counterfactuals within a model fit to the Los Angeles metro housing market in 2019. Detail on the counterfactual experiment appears in the note to Figure 5. I plot the Total series in Panel B of Figure 5 minus the Total series in Panel A of that figure.

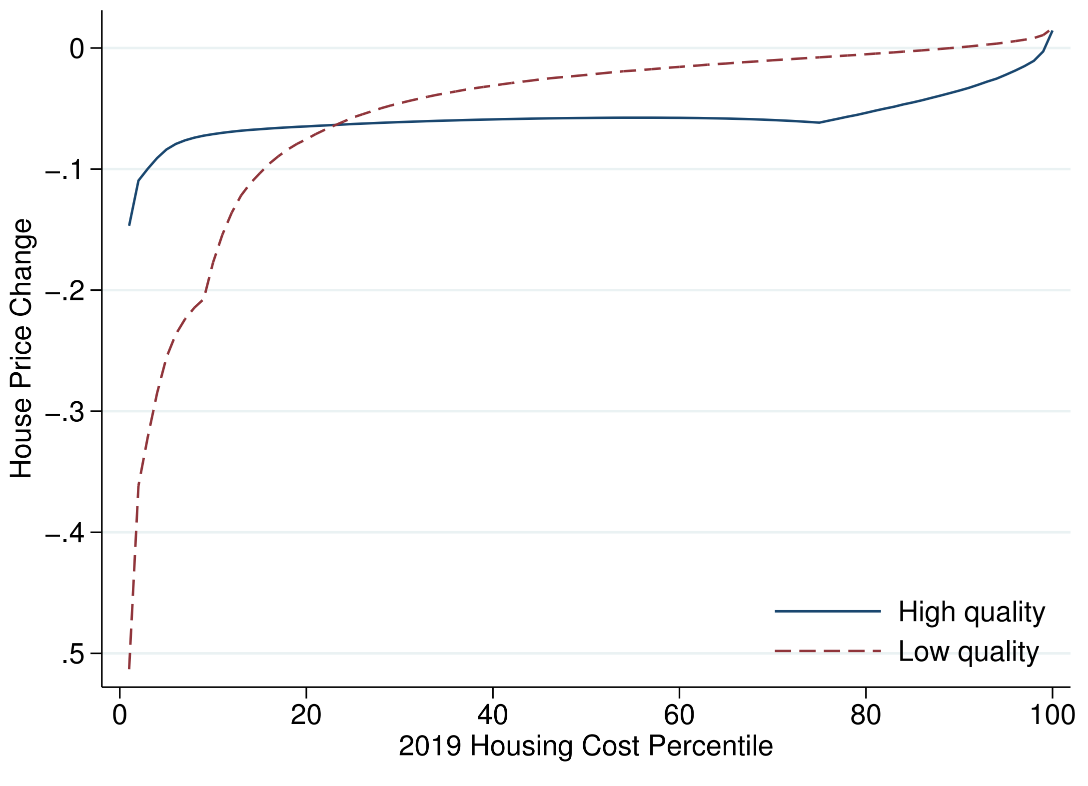

Figure A4. Effect of Construction on Housing Costs

Notes: The results are counterfactuals within a model fit to the Los Angeles metro housing market in 2019. In the High quality counterfactual, I increase the metro's housing stock by 11.8% by supplying housing at the 75th percentile. In the Low quality counterfactual, I increase the metro's housing stock by the same amount by supplying housing at the 9th percentile. For each percentile, I plot the resulting change in the price of housing.

TABLE A1

Housing units by year built, NHGIS

	Owner occupied		Renter occupied		Vacant		Total	
Year built	2000	2019	2000	2019	2000	2019	2000	2019
2014–2019	-	1,842,570	-	1,106,082	-	451,190	-	3,399,842
2010–2013	-	2,017,682	-	1,335,437	-	336,919	-	3,690,038
2000–2009	-	12,189,768	-	4,823,321	-	2,173,843	-	19,186,932
1990–1999	-	11,394,937	-	5,621,132	-	2,056,538	-	19,072,607
1999–2000	1,740,646	-	445,111	-	569,318	-	2,755,075	-
1995–1998	5,914,129	-	1,846,103	-	718,743	-	8,478,975	-
1990–1994	5,756,320	-	2,052,015	-	658,673	-	8,467,008	-
1980–1989	11,001,379	10,056,592	5,672,973	6,067,082	1,652,495	2,331,633	18,326,847	18,455,307
1970–1979	12,260,326	11,022,459	7,308,622	7,299,004	1,869,915	2,556,092	21,438,863	20,877,555
1960–1969	9,146,099	7,935,694	5,356,867	4,884,545	1,408,937	1,706,590	15,911,903	14,526,829
1950–1959	9,334,476	8,517,512	4,264,793	4,101,333	1,110,880	1,523,302	14,710,149	14,142,147
1940–1949	4,727,489	3,582,779	2,916,774	2,266,377	791,505	899,443	8,435,768	6,748,599
≤ 1939	9,935,649	8,714,388	5,800,330	5,977,354	1,644,074	2,637,388	17,380,053	17,329,130
Total	69,816,513	77,274,381	35,663,588	43,481,667	10,424,540	16,672,938	115,904,641	137,428,986

Notes: Data cover housing units in the United States in the 2000 5% U.S. Census sample and the 2015–2019 ACS, as appearing in Census tables accessed through NHGIS (Manson et al., 2022). The 1999–2000 category includes properties built in 1999 and the first quarter of 2000.

TABLE A2

Housing units by year built, IPUMS

	Owner occupied		Renter occupied		Vacant		Total	
Year built	2000	2019	2000	2019	2000	2019	2000	2019
2014–2019	-	1,860,553	-	1,126,938	-	451,954	-	3,439,445
2010–2013	-	1,996,609	-	1,329,856	-	338,292	-	3,664,757
2000–2009	-	12,072,109	-	4,848,958	-	2,167,029	-	19,088,096
1990–1999	-	11,333,190	-	5,654,442	-	2,052,567	-	19,040,199
1999–2000	1,737,516	-	444,559	-	572,190	-	2,754,265	-
1995–1998	5,902,770	-	1,855,981	-	714,685	-	8,473,436	-
1990–1994	5,760,375	-	2,046,397	-	656,875	-	8,463,647	-
1980–1989	10,999,041	9,999,805	5,673,367	6,143,993	1,657,945	2,328,024	18,330,353	18,471,822
1970–1979	12,252,861	10,967,138	7,323,337	7,364,863	1,868,230	2,547,103	21,444,428	20,879,104
1960–1969	9,148,770	7,907,131	5,351,435	4,947,042	1,403,596	1,714,618	15,903,801	14,568,791
1950–1959	9,338,964	8,464,703	4,261,205	4,155,390	1,110,025	1,534,984	14,710,194	14,155,077
1940–1949	4,724,511	3,559,953	2,911,419	2,280,407	800,439	896,649	8,436,369	6,737,009
$\leq$ 1939	9,953,737	8,708,716	5,793,856	6,034,219	1,640,555	2,641,751	17,388,148	17,384,686
Total	69,818,545	76,869,907	35,661,556	43,886,108	10,424,540	16,672,971	115,904,641	137,428,986

Notes: Data cover housing units in the United States in the 2000 5% U.S. Census sample and the 2015–2019 ACS, as appearing in public-use Census micro data accessed through IPUMS (Ruggles et al., 2023). The 1999–2000 category includes properties built in 1999 and the first quarter of 2000.

TABLE A3
VALUE DISTRIBUTION, OWNER-OCCUPIED UNITS

	2000		2019	
Value, $	NHGIS	IPUMS	NHGIS	IPUMS
< 10,000	1,213,739	1,211,737	1,028,776	1,008,690
10,000–14,999	875,743	876,325	555,266	553,945
15,000–19,999	881,701	882,000	493,678	487,069
20,000–24,999	983,663	993,165	541,349	535,955
25,000–29,999	1,074,059	1,075,206	496,197	492,304
30,000–34,999	1,241,653	1,240,773	605,943	601,202
35,000–39,999	1,358,507	1,356,474	448,984	442,593
40,000–49,999	2,802,293	2,803,821	1,165,696	1,160,405
50,000–59,999	3,340,679	3,331,501	1,451,180	1,434,765
60,000–69,999	3,883,453	3,892,580	1,685,990	1,667,708
70,000–79,999	4,224,104	4,236,508	1,937,078	1,928,461
80,000–89,999	4,786,358	4,784,804	2,288,047	2,270,978
90,000–99,999	4,453,978	4,440,716	1,946,546	1,935,377
100,000–124,999	8,003,853	8,012,791	5,290,207	5,248,186
125,000–149,999	7,190,118	7,200,090	4,952,877	4,896,808
150,000–174,999	5,466,573	5,466,121	6,272,151	6,234,748
175,000–199,999	3,900,943	3,885,425	4,582,862	4,541,507
200,000–249,999	4,727,813	4,715,150	8,274,180	8,234,499
250,000–299,999	3,073,248	3,083,685	6,877,112	6,851,044
300,000–399,999	2,910,231	2,904,423	9,450,290	9,409,140
400,000–499,999	1,364,279	1,360,434	5,498,064	5,496,368
500,000–749,999	1,192,681	1,198,470	6,314,185	6,321,223
750,000–999,999	420,835	416,348	2,506,929	2,468,529
≥ 1,000,000	446,009	449,998	-	-
1,000,000–1,499,999	-	-	1,400,103	1,267,374
1,500,000–1,999,999	-	-	508,010	593,130
≥ 2,000,000	-	-	702,681	787,899
Total	69,816,513	69,818,545	77,274,381	76,869,907

Notes: Data cover owner-occupied housing units in the United States in the 2000 5% U.S. Census sample and the 2015–2019 ACS, as appearing in Census tables accessed through NHGIS (Manson et al., 2022) or in public-use Census micro data accessed through IPUMS (Ruggles et al., 2023).

TABLE A4
RENT DISTRIBUTION, RENTER-OCCUPIED UNITS

	2000		2019	
Rent, $	NHGIS	IPUMS	NHGIS	IPUMS
< 100	338,433	323,888	123,528	120,899
100–149	619,084	560,842	110,554	109,275
150–199	886,664	923,380	206,021	199,936
200–249	842,638	847,583	593,412	589,659
250–299	976,126	971,346	566,246	573,353
300–349	1,345,152	1,331,834	506,897	517,724
350–399	1,756,259	1,733,476	494,868	512,581
400–449	2,198,741	2,184,903	583,129	576,848
450–499	2,439,363	2,417,070	680,382	655,264
500–549	2,628,948	2,636,056	883,086	857,897
550–599	2,554,636	2,558,578	1,046,955	1,071,198
600–649	2,464,722	2,478,694	1,282,428	1,279,470
650–699	2,225,202	2,231,752	1,416,121	1,418,519
700–749	1,986,790	1,993,980	1,578,633	1,585,896
750–799	1,683,956	1,688,955	1,695,293	1,731,082
800–899	2,612,631	2,637,550	3,521,999	3,551,544
900–999	1,748,586	1,769,613	3,531,570	3,494,969
1,000–1,249	2,151,418	2,139,832	7,362,992	7,375,905
1,250–1,499	902,681	896,789	5,040,510	5,054,900
1,500–1,999	685,164	555,441	5,796,670	5,955,638
≥ 2,000	339,132	501,836	-	-
2,000–2,499	-	-	2,324,841	2,374,892
2,500–2,999	-	-	987,766	1,065,283
3,000–3,499	-	-	531,931	598,798
≥ 3,500	-	-	446,040	429,070
Non-specified	464,086	283,215	-	-
No cash rent	1,813,176	1,994,943	2,169,795	2,185,508
Total	35,663,588	35,661,556	43,481,667	43,886,108

Notes: Data cover renter-occupied housing units in the United States in the 2000 5% U.S. Census sample and the 2015–2019 ACS, as appearing in Census tables accessed through NHGIS (Manson et al., 2022) or in public-use Census micro data accessed through IPUMS (Ruggles et al., 2023). Rent is measured as gross rent, which includes utilities. Non-specified units are renter-occupied single-family homes on 10 or more acres. The Census Bureau does not provide rent data for such units in 2000 but does in the 2015–2019 survey.

TABLE A5
ESTIMATION SAMPLE FOR LOS ANGELES-LONG BEACH-ANAHEIM, CA

	Homeless	Renters	Owners
Panel A. Averages			
Income	$8,938	$70,002	$138,861
Education	0.130	0.320	0.470
Rent	-	$1,659	-
Home value	-	-	$766,741
Panel B. Totals			
Sum of weights	54,755	2,204,557	2,093,953
Observations	5,650	101,963	125,174

Notes: Data include group quarters persons and households in the Los Angeles-Long Beach-Anaheim, CA metro area in the 2015–2019 ACS. I categorize a subset of group quarters persons as homeless as described in the note to Table 2. I drop renters who do not pay cash rent. To take averages in Panel A, I weight observations by household weights (hhwt in IPUMS). I report the sum of these weights in Panel B.