NBER WORKING PAPER SERIES

URBAN REVIVAL IN AMERICA, 2000 TO 2010

Victor Couture
Jessie Handbury

Working Paper 24084
http://www.nber.org/papers/w24084

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2017

We thank Prottoy Aman Akbar, Hero Ashman, Yue Cao, Yizhen Gu, Jeffrey Jacobs, Hae Nim Lee, Ellen Lin, Daniel Means, and Jungsoo Yoo for outstanding research assistance. We thank David Albouy, Nate Baum-Snow, Don Davis, Jorge De la Roca, Rebecca Diamond, Jonathan Dingel, Gilles Duranton, Ben Faber, Fernando Ferreira, Ed Glaeser, Joe Gyourko, Andy Haughwout, Jeffrey Lin, Hal Martin, Christopher Palmer, Jordan Rappaport, Jesse Shapiro, Matt Turner, and Owen Zidar, as well as participants at the 2014 Urban Economics Association Meeting, the 2015 Duke-ERID Conference, the 2016 Stanford-SITE Conference, the 2016 NBER SI, the 2017 Chicago BFI Conference and numerous seminars for useful comments. Jessie Handbury thanks the Research Sponsors' Program of the Wharton Zell-Lurie Real Estate Center and the Wharton Dean's Research Fund, and Victor Couture thanks the Fisher Center for Real Estate and Urban Economics for generous financial support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Urban Revival in America, 2000 to 2010
Victor Couture and Jessie Handbury
NBER Working Paper No. 24084
November 2017
JEL No. R23

ABSTRACT

This paper documents and explains the striking rise in the proclivity of college-educated individuals to reside near city centers. We show that this recent urban revival is driven entirely by younger cohorts in larger cities. With a residential choice model, we quantify the role of jobs, amenities, and house prices in explaining this trend. We find that changing preferences of young college graduates for non-tradable service amenities like restaurants, bars, gyms, and personal services account for more than 50 percent of their growth near city centers. Complementary datasets confirm that the young and college-educated are indeed spending more on and taking more trips to non-tradable service establishments. Our investigation into the causes of rising preferences for non-tradable services highlights their expanding role in generating socializing opportunities with other young college graduates, but also indicates roles played by delayed family formation, rising incomes, and improvements in the quality and diversity of non-tradable services.

Victor Couture
University of California at Berkeley
Berkeley, CA 94720-1900
couture@berkeley.edu

Jessie Handbury
The Wharton School
University of Pennsylvania
1463 Steinberg-Dietrich Hall
Philadelphia, PA 19104
and NBER
handbury@wharton.upenn.edu

This paper documents and seeks to explain the striking reversal in the fortunes of urban America since 2000. We show that the college-educated population started urbanizing in most large U.S. cities between 2000 and 2010, in spite of continued suburbanization of the aggregate population.[1] This reversal was entirely driven by rapid growth of the young and college-educated near city centers, as older college-educated cohorts also continued to suburbanize.

To quantify the importance of various competing hypotheses explaining these stylized facts, we estimate a residential choice model with a rich database at a fine spatial scale. In our model, individuals choose a CBSA and tract to live in based on access to jobs and amenities, house prices, and nested-logit idiosyncratic taste parameters. Our estimates reveal a decisive role for high initial levels of non-tradable service consumption amenities like restaurants, bars, gyms and personal services downtown in explaining the urbanization of young college graduates. We find that recent changes in well-studied characteristics like job density and public amenities (school, crime and transit) struggle to explain the distinctive urbanization of young college graduates, even though these characteristics are often important determinants of locational choices across all tracts and in the broader population. We also measure changes in access to various types of service and retail consumption amenities, and find that such changes did not favor downtown tracts over the last decade.

The intuition here is simple: an important explanatory factor for urban revival must 1) be highly prevalent near city centers relative to elsewhere, and 2) strongly attract young college graduates relative to other age-education groups. Our data shows that downtowns are characterized by a high density of non-tradable service amenities. Our residential choice model estimates suggest that high initial levels of non-tradable services like restaurants have attracted young college graduates more than their non-college educated and older college-educated counterparts. Our results suggest that initial levels of non-tradable service amenities account for between 50 to 80 percent of young college-educated urbanization from 2000 to 2010. Other factors, like the slight urbanization of high-wage jobs, appear to be less important.

To identify time-varying, age-education group-specific preference parameters, we include both the change in a variable, whose coefficient captures the level of the preference for a characteristic, and the initial level of a variable, whose coefficient captures the change in those preferences over the previous period, in our estimating equation. In the context of our model, a positive impact of initial levels of non-tradable service amenities on location choice reflects rising preferences for these amenities. We do not, however, interpret our model's group-specific

[1] This suburbanization has been extensively studied, such as in Glaeser et al. (2004), Baum-Snow (2007), and Boustan (2010). The reversal of this trend was already apparent in the 1990s and before in a handful of gateway cities like New York, Chicago, Boston and San Francisco. Carlino and Saiz (2008) also show that, while central cities do not experience a revival in the 1990s, some recreational districts were already seeing college-educated growth by the 1990s. Our finding is that urban revival emerges as a distinct widespread phenomenon in the 2000s, and is restricted to areas smaller than the central city.

and time-varying preferences as reflecting shifts in deep underlying preference parameters.[2] Instead, we view these preference changes as capturing, in a reduced-form way, changes in collective preferences across cohorts or time driven by, for example, changes in the underlying demographic composition of each age-education group or changes in access to technology that we explore in a complementary inquiry.

We begin this inquiry by assessing the external validity of our model's preference parameter estimates using expenditure data from the Consumer Expenditure Survey (CEX) and travel data from the National Household Travel Survey (NHTS). The patterns in these expenditure and travel data are consistent with our model's result that the young and college-educated have a stronger preference for non-tradable service amenities like restaurants and bars than other age-education groups, and that they have experienced the most positive change in this preference over the last decade. These patterns do not hold for other types of consumption amenities like tradable retail.

We then present evidence supporting three hypotheses for the changing preferences of the young and college-educated for non-tradable service amenities. First, we find that young college graduates' propensity to move to locations with high initial levels of non-tradable services is much higher in areas with a high initial share of their own age-education group. This link between homophily and non-tradable services like restaurants and bars suggests that changing preferences for these establishments derive in part from opportunities that they provide to socialize, network, and date within one's own group. Second, what we interpret as a changing preference for high amenity density may be due to changing amenity composition in high density areas toward a more diverse set of establishments that cater to the special tastes of young college graduates. We find that both rising diversity and improving quality of restaurants attract young college graduates downtown. Moreover, only the largest cities have experienced a relative improvement in restaurant diversity and quality near their city centers, consistent with the urbanization of young college graduates being primarily a large city phenomenon. However, the size of our estimated parameter, which indicates the changing preference for non-tradable service density, does not decline with the inclusion of quality and diversity controls. Third, we show that richer and solo-living young college graduates devote a larger fraction of their expenditures and travel to non-tradable services. Therefore, delayed family formation in the 2000s can explain the changing preference of the young and college-educated for non-tradable services, because it mechanically shifts their travel and expenditure shares toward this type of establishments. Rising disposable income would have a similar impact, and we have evidence that urban revival starts during a period of rising income before the Great Recession. However,

[2] Becker (1965), for example, describes the pitfalls of explaining new trends with a changing utility function.

the Great Recession overlaps with most of our post-period data and obscures these trends.[3] [4]

Our analysis contrasts with existing work on residential choice in the U.S. in three important ways. First, our empirical approach incorporates a broad set of competing explanatory factors to quantify their relative importance. This comprehensive approach distinguishes our work from a concurrent set of papers on central city gentrification that documents and explains the rising socio-economic status of downtown inhabitants over the last decade. Baum-Snow and Hartley (2016) identify rising amenity values after 2000 as the most important driver of downtown gentrification, consistent with our results. In their model, amenities are an unobserved residual that compensate for differences in employment opportunities and house prices à la Rosen (1979) and Roback (1982). Our model is rich enough to reveal the special importance of non-tradable services relative to other types of residential amenities. Edlund et al. (2015) instead focus on the longer hours worked by college-educated workers after 1970 as an explanation for their centralization. Our model estimates suggest that this loss of leisure time may have had a larger impact on valuation for proximity to non-tradable service amenities useful in outsourcing home production, as in Murphy (2017), than on valuation for proximity to jobs. Ellen et al. (2017) find that the 1990s crime drop predicts central city gentrification in the 2000s. Our model confirms that prior crime decline also contributes to the urbanization of young college graduates in the 2000s. It is not surprising that more than one factor contributes to explain a phenomenon as pervasive as urban revival. However, our empirical results suggest that the contribution of non-tradable services is more important than that of crime.

Second, we document the important role of consumption amenities in residential choice within CBSAs. Existing work in the U.S. has either considered only public amenities like schools and crime (see, e.g., Epple and Sieg, 1999 and Bayer et al., 2007) or the role of consumption amenities within larger geographical units like CBSAs (e.g., Diamond, 2016).[5] We build tract-level density indexes capturing proximity to consumption amenities in various types of non-tradable services, tradable retail, and activities. The localized nature of these density indexes matters, as one may move to the Bay Area primarily for job opportunities, but choose to live in the center of San Francisco for the consumption amenities. We additionally construct a novel Bartik-type instrument for changes in these consumption amenity density indexes, drawing from a recent IO literature on the determinants of entry and exit for various types of retail establishments (e.g., Igami and Yang, 2016).

Finally, our empirical framework is methodologically distinct from existing work study-

[3] One variant of this explanation for changing tastes is recent income growth among the college-educated, which tends to make them more likely to pay for locations with a high perceived quality of life, as hypothesized by Rappaport (2009) and Gyourko et al. (2013).

[4] Recent innovation in mobile technology may also complement urbanized amenities, which benefits digitally savvy young college graduates. We do not find any evidence to support this hypothesis, so it remains speculative.

[5] Teulings et al. (2014) consider restaurant density in a residential choice model at the zip code level in the Netherlands.

ing within-CBSA location choices. For instance, our model differs from Bayer et al. (2007)'s important application of McFadden (1973) and McFadden (1978)'s random utility model to neighborhood choice in the Bay Area in 1990. Unlike Bayer et al. (2007), we derive our indirect utility from a primitive Cobb-Douglas consumer optimization problem, and we add a time dimension, a CBSA dimension, and many additional neighborhood characteristics. We do not use the technique developed in Berry et al. (1995) to estimate unobserved neighborhood heterogeneity. Instead, we derive simpler first-difference linear regressions that control for any time-invariant unobservables. In robustness checks, we extend our residential choice model to a residential-workplace choice model, which allows us to propose an alternative identification strategy based on workplace fixed-effects.[6] With tract-level data on the joint residential and workplace choice, a workplace tract fixed-effect removes the bias due to endogenous firm location from our coefficients on residential characteristics. The intuition for this identification strategy relates to the argument in Glaeser et al. (2001) that a rising share of people who live in the central city but work in the suburbs signals the rising importance of central city amenities. Our findings demonstrate that little bias results from using only residential data instead of residence-workplace data.[7]

More generally, our findings can inform theories on location choice within cities (see Duranton and Puga (2015) for a review) and contribute to an empirical literature on the determinants of college-educated location choices. The seminal monocentric city model (Alonso 1964; Mills 1969; Muth 1969), for instance, has no consumption amenities and only access to jobs determines the high residential density near the city center. Influential empirical studies by Glaeser et al. (2004), Moretti (2012), and Diamond (2016) demonstrate that the share of college-educated individuals is a key determinant of economic success *across* cities since 1980. The new within-city trends that we study may have similarly far-reaching implications.

Indeed, the demographic shift that we document has become increasingly visible in the streets of urban America. In October 2017, CNN's Ron Brownstein portrayed "center cities pulling away the most upwardly mobile people" as a "giant sucking sound [...] changing the economy and disrupting politics." This paper focuses on documenting and explaining the recent

[6]Baum-Snow and Hartley (2016) also estimate a tract-level residential choice model across years and CBSAs, but not a residential-workplace choice model. Other studies that we are aware of have coarser geography or lack a time dimension. Albouy and Lue (2015) estimate a residential-workplace model using one year of data at the larger PUMA geography. Their finding that variation in quality of life is as important within metropolitan areas as across them motivates the within-city analysis in our paper. Important contributions to residential-workplace modeling include Waddell et al. (2007), who use 1999 data in the Puget Sound Region in WA, and Monte et al. (2015), who use commuting-zone or county-level data from the ACS 2010.

[7]We relegate this residential-workplace model to our robustness section because the LODES residence-workplace data is not available by age and education groups, but only by either broad income or broad age categories. Moreover, adding a workplace fixed-effects only allows us to identify preferences for residential characteristics at the cost of removing our ability to separately estimate the importance of proximity to jobs in residential location choice. Therefore a model with workplace fixed-effect cannot distinguish the contribution of jobs to urban revival from that of residential amenities, as our residential-only model does.

urbanization of the college-educated, but what we call "urban revival" may have adverse welfare consequences for other groups. For instance, poorer individuals may incur welfare losses if they are being priced out of urban areas that catered to their specific needs (e.g., transit access). We are investigating the welfare impact of such gentrification in complementary work. Amid growing public concerns for gentrification and geographic disparities in economic opportunities, this paper is a step toward understanding the socio-economic forces behind the recent urban revival in America.

The rest of the paper is divided as follows. We describe the data in section 1. Section 2 presents the stylized facts on urban revival. Sections 3 and 4 present the residential choice model and our empirical application of this model to identifying the key drivers behind the urbanization of the young and college-educated. Section 5 presents various robustness checks on our results and section 6 provides external validity for the changing preferences for non-tradable service amenities that we find to drive urban revival. Section 7 explores various hypotheses for these changing preferences and section 8 concludes.

1 Data

The main geographical unit in our analysis is a census tract within a Core-Based Statistical Area (CBSA). We construct constant 2010-boundary CBSAs using constant 2010-boundary tracts from the Longitudinal Tract Data Base (LTDB). We define the city center of each CBSA using the definitions provided by Holian and Kahn (2012), obtained by entering the name of each CBSA's principal city into Google Earth and recording the returned coordinates.

To establish the stylized facts on recent urban growth that motivate our empirical analysis, we assemble a database describing the residential locations of U.S. individuals at a decennial frequency. Tract-level population counts by education level are from the decennial censuses of 1980 to 2000 and the American Community Survey (ACS) 2008-2012 aggregates, downloaded from the National Historical Geographic Information System (NHGIS). Population counts by age and education level, which we use to estimate our residential choice model, are only available in the 2000 Census and in the 2008-2012 ACS.

Next, we build datasets describing access to jobs, consumption amenities, and house prices in the vicinity of each census tract in 2000 and 2010. To measure job density by wage groups, we use the LEHD Origin-Destination Employment Statistics (LODES) datasets for 2002 and 2011. The LODES data provide counts of people who live and work in a given census block pair by three different nominal wage groups: high-wage workers earning more than \$3,333 per month, middle-wage (\$1,000-\$3,333) and low-wage (<\$1,000).[8] To measure consumption

[8] The share of workers in the upper bracket is 27 percent in 2002 and 37 percent in 2011. To address confidentiality issues, the LODES data are partially synthetic. We describe the generation of synthetic data in Appendix A,

amenity density, we use two main datasets: 1) a geo-coded census of establishments in 2000 and 2010 from the National Establishment Time-Series (NETS), and 2) travel times between these establishments and census tract centroids by foot from Google Maps.[9] We define up to nine types of consumption amenity in non-tradable services (restaurants, bars, gyms, and personal services), tradable retail (food, apparel, and general merchandise stores) or activities (amusement parks/golf and museums/libraries/galleries). We measure consumption amenity diversity as an inverse-Herfindahl index using the most refined industry classification available in the NETS (at the SIC8 level, e.g., Korean restaurants). For a few consumption amenity types, we can also measure quality using Esri's Market Potential Index (MPI), which captures the propensity of different socio-economic groups to shop in a given chain store. Our primary house price index for 2000 and 2010 is the Zillow House Value Index for all homes, which measures median house prices at the zip code level. We match these to 2010 tract geography using a zip code to tract crosswalk from the U.S. Department of Housing and Urban Development (HUD).[10] In robustness checks, we use two alternative house price indexes: Zillow's two bedroom index, and a hedonic price index calculated using DataQuick data and the model from Ferreira and Gyourko (2011). We further include rental prices using HUD's Fair Market Rent Series for one bedroom, two bedroom and three bedroom homes (available at the county level), and finally we include the median age of the housing stock from the 2000 census and the 2008-2012 ACS, to measure one aspect of housing quality.

We complement these three main datasets with information on public amenities (transit times, violent crime per capita, school district rankings) and natural amenities. Our measure of transit performance at the tract-level comes from Google Maps in 2014, and is the average travel time of a five-mile trip from a tract centroid to a random set of 100 NETS establishments nearby. We measure violent crime (murder, rape, robbery, and aggravated assault) at the police district-level using the Uniform Crime Reporting Program (UCR) in 2000 and 2010. We measure school quality using within-state rankings of school districts in 2004 and 2010 from SchoolDigger.com.[11] There are typically multiple tracts within a particular police and school

and show how aggregation of census block data at the tract level ensures that 90 percent of the LODES data are unaffected by synthesis.

[9]We computed density indexes by car and by transit, but only use indexes by foot because other modes delivered weak instruments. The popularity of the Walk Score, which rates neighborhoods by how walkable they are, hints at the importance of such indexes in location decisions. We also failed to obtain strong instruments for two types that we exclude in the paper: "Theater" (theater, operas, symphonies, etc) and "Movie" (movie theater and bowling).

[10]We expand the dataset beyond tracts that Zillow covers by about 20 percent. To do this, we calculate the average house price index within a tract-group for tracts covered by Zillow, and apply this average to tracts that are not covered by Zillow. Tract-groups are a set of three to four neighboring tracts, as defined in Ferreira and Gyourko (2011), who similarly estimate hedonic price indexes at the tract-group level.

[11]While we believe that SchoolDigger.com is the most comprehensive database available, we have school ranking data for less than half of our CBSAs' sample of tracts. SchoolDigger.com compiles test scores and provides

district. We match these areas to 2010 tract boundaries using Census shapefiles.[12] Data on natural amenities, like the precipitation, hilliness, and coastal proximity of each census tract, are from Lee and Lin (forthcoming).

To investigate recent trends in household formation, income growth, expenditures, and travel that can explain the changing preferences of young college graduates, we use counts of individuals by household type and income within each age-education group. These counts come from the 5% Integrated Public Use Micro-data Series (IPUMS) sample of the 2000 census and the 5% IPUMS sample from 2007-2011 ACS surveys, as well as micro-data from the 1998-2002 and 2008-2012 Consumer Expenditure Survey (CEX) and the 2001 and 2009 National Household Transportation Survey (NHTS). Appendix A provides detailed information on all data sources.[13]

2 Stylized Facts

Claims of urban revival are not new. The 1960s and 1970s were times of rapid decline for urban America, with many central cities losing population. Various forms of urban comeback have been documented since the early 1990s (e.g., Frey, 1993). In recent years, tales of urban revival in America have become commonplace and widely relayed by the popular press. Census tables, however, tell an unequivocal story of continued suburbanization (Kotkin and Cox, 2011). In this section, we establish a number of stylized facts about changes in the within-city location choice of Americans from 2000 to 2010, most of which are new. These facts motivate our empirical analysis. In section 4.1, we further document the changing spatial distribution of key residential location characteristics, such as house prices, proximity to jobs, and amenities. In section 5.2, we provide additional facts about changes in commute patterns since 2000.

To establish facts on urban revival, we construct kernel density plots of tract population growth at various distances from the city center, shown in Figure 1. Each kernel plot displays population growth gradients for four groups of CBSAs defined by population size.[14] We

a ranking of each school district within each U.S. state. The ranking averages over test scores in different fields for schools from grades 1 through 12. We use the inverse of that ranking in percentile for 2004 - the earliest year available - and for 2010 in the school district that a tract falls into as our measure of school quality in 2000 and 2010.

[12]This mapping projects 11,044 police districts to 57,095 census tracts, and 12,956 school districts to 24,283 census tracts. Police districts are mostly cities and, while CBSAs consist of many cities, the central city in most CBSAs is larger than the downtown experiencing urban revival. In some cases like Houston and Atlanta, police districts are at the county level, so the parts of the respective central city in different counties report different numbers. Our results are robust to using a sub-sample containing only those CBSAs where the largest police district contains less than 30 percent of the CBSA population.

[13]We use the 2007-2011 ACS samples of the IPUMS data, rather than the more recent 2008-2012, because a change in PUMA boundaries in the 2008-2012 ACS sample prevents us from defining urban areas consistent with our urban definitions for this later period.

[14]Note that the CBSA size group definition (e.g., 10 largest CBSAs) can change at the beginning of each decade.

measure distance from the city center as the cumulative share of the pre-period CBSA group population. As we move further from the city center, the share of the included population increases until the area measured encompasses the entire CBSA group population. The first row of Figure 1 shows the continuing suburbanization of the general population from 1980 to 2010. All three decades since 1980 feature a positive growth gradient from the city center in CBSAs of all sizes.

Slower urban population growth does not preclude urban revival. In fact, many authors argue that the college-educated population is instead the better indicator of spatial success in recent decades (Glaeser et al., 2004; Moretti, 2012).[15] Accordingly, the second row of Figure 1 replicates the first row with the college-educated population only. These plots uncover a new, previously undocumented trend: between 2000 and 2010, in both the 10 largest and the 11-50$^{\text{th}}$ largest CBSAs, college-educated growth is fastest near the city center. This trend is specific to the 50 largest cities and, outside of three or four gateway cities like New York, Chicago and San Francisco, only emerged in the most recent decade. This explains why early claims of urban revival were not backed by systematic evidence. Defining downtowns as the set of tracts closest to the city center accounting for 5 percent of a CBSA's population, we find that in 28 of the 50 largest CBSAs, the college-educated population grew faster downtown than elsewhere in the 2000s. This compares to only 10 out of 50 in the 1990s and 9 out of 50 in the 1980s. This trend is robust to a number of downtown definitions, but too localized to show in a simple comparison of central cities to the suburbs.[16]

The third row of Figure 1 further refines our investigation and breaks down college-educated growth in the 2000s by age group. This breakdown is relevant because we expect the residential preferences of the younger generation to differ from that of older Americans.[17] Our results show a negative growth gradient from the city center in the 50 largest cities for both the "young" 25-34 year-olds and "middle-aged" 35-44 year-old college-educated groups. The youngest group experiences the sharpest gradient with nearly 40 percent growth near city centers relative to around 15 percent growth past the twentieth distance percentile. This trend is much less pronounced outside of the 50 largest cities. Contrary to claims by the popular press that retiring

We weight the kernel regression by initial tract population to ensure that local growth estimates are independent of tract size.

[15] Note also that in downtowns that are already built up and subject to heavy housing regulations limiting supply (Glaeser et al., 2006), increased desirability is more likely to trigger rising house prices and gentrification rather than aggregate population growth.

[16] An online Appendix E available on the authors' website proposes different ways of tabulating the data shown in Figure 1. It compares downtown growth - using various downtown definitions - to that in the rest of the CBSA to document the scope of urban revival across cities.

[17] A recent industry report by CEO for Cities (Cortright, 2014) also uses 2000 census data and 2008-2012 ACS data. Considering only the 51 largest MSAs and the 25-34 college-educated population, the report shows that the 25-34 college-educated population is growing faster downtown, defined as within a 3-mile radius from the city center.

baby boomers are urbanizing, the older 45-64 and 65+ year-old (not shown) college-educated groups are still suburbanizing.[18]

These local trends are strong enough to have an aggregate impact. About 150 million Americans live in the 50 largest CBSAs. In these large cities, downtowns accounting for five percent of the population experienced 24 percent of the total increase in the young college-educated population and 11.5 percent of the middle-aged college-educated population between 2000 and 2010. Strikingly, young college graduates grew faster downtown than elsewhere in the CBSA in 23 of the 25 largest CBSAs. The exceptions are Riverside, CA, whose downtown is small, and Detroit.

Figure 1 also highlights the compositional shift driving these trends. In the 2000s, population growth near the city center of the largest cities is near zero. Urban revival is therefore entirely driven by the rising urban share of the young and college-educated, with no contribution from general population growth.[19]

Our stylized facts are robust to using a different city center definition (i.e., defined as Central Business Districts from the 1982 census of retail trade), income groups instead of age-education groups, and alternative datasets, such as the LODES data of commute by wage groups that we use to estimate our residential-workplace choice model, as well as the earliest ACS data available as a post-period (2005-2009, showing that urban revival starts before the Great Recession). Other current work on central city gentrification also supports our findings, with Baum-Snow and Hartley (2016) showing that downtowns are becoming richer, more educated and more white.

The objective of the rest of this paper is to find the factors driving the 2000-2010 population growth gradients by age-education groups documented in Figure 1, with a sharp focus on explaining the remarkable growth of the young and college-educated near city centers.

[18]The popular press also emphasizes the urbanization of "millennials," those born from 1980 to the late 1990s, but this generation is too young to drive urban revival, which shows even in 2005-2009 ACS data. The oldest millennials, born in 1980, are only 30 in 2010. Rappaport (2015) suggests that aging baby boomers will support strong demand for multi-family units, but that these downsizing households will remain close to their original suburban locations. This is consistent with our finding that baby boomers do not contribute to urban revival.

[19]Rust-belt cities like Cleveland and Detroit provide interesting case studies. Cleveland experienced "urban revival" despite a declining downtown population (a 12 percent drop from 2000 to 2010), thanks to changes in downtown composition (78 percent growth in young-college graduates from 2000 to 2010). Detroit also has a downtown population that declines as it shifts towards the young and college-educated. However, Detroit's downtown had the sharpest population drop and the smallest young college-educated growth of any large city. Detroit's downtown still shows promise of future revival: its youngest college-educated group - 18-24 year-olds, a very small group - urbanized quickly from 2000 to 2010.

3 Residential Choice Model

To explain the changing residential location choices of different age-education groups, we specify a workhorse discrete choice model. The model delivers an estimating equation capturing the effects of changes in the environment (jobs, amenities, and house prices) from 2000 to 2010, as well as initial 2000 levels in these variables, on changes in the share of an age-education group living in a given tract. In Appendix D we augment the model to study the joint workplace-residential location decision.

Each individual i in group d selects a tract j in CBSA c in which to reside in year t and chooses how to allocate their expenditure between units of housing H, consumption amenities A, and a freely-traded outside good Z in order to maximize the following Cobb-Douglas utility function:

$$U^i_{jct} = \alpha^i_{jct} H^{\beta^{d(i)}_{Ht}} A^{\beta^{d(i)}_{At}} Z^{\beta^{d(i)}_{Zt}}$$

subject to a budget constraint:

$$w^{d(i)}_{jct} = p_{Hjct} H + p_{Ajct} A + Z,$$

where w^d_{jct} is the wage net of commute costs, which we assume to be common to all individuals in group d residing in tract j, p_{Hjct} is the price of housing, and p_{Ajct} is a price index for consumption amenities that varies with transport costs to these amenities. α^i_{jct} reflects the utility that an individual receives for residing in tract j in CBSA c at time t, regardless of their expenditure in that location. This taste shifter captures utility from public amenities, a_{jct}, such as school quality and crime, as well as unobserved group- and individual-specific tastes:

$$\alpha^i_{jct} = \exp\left(\beta^{d(i)}_{at} \ln a_{jct} + \mu^{d(i)}_{jc} + \xi^{d(i)}_{jct} + \varepsilon^i_{jct}\right).$$

The group-specific tastes for each tract are represented by the sum of two group-specific terms: a time-invariant component μ^d_{jc}, and a time-varying component, ξ^d_{jct}. The individual-specific tastes, ε^i_{jct}, take a nested-logit structure with tracts nested by CBSA with a within-group correlation parameter σ^d.[20]

After solving the Cobb-Douglas utility maximization problem, each individual i chooses its

[20] This implies that individual-specific taste shocks, ε^i_{jct}, are themselves the weighted sum of two shocks, $\varepsilon^i_{jct} = \psi^i_{ct}(\sigma^{d(i)}) + (1 - \sigma^{d(i)})\nu^i_{jct}$. Tract-specific taste shocks, ν^i_{jct}, are independent draws from the extreme value distribution, while CBSA taste shocks, ψ^i_{ct}, are independent draws from the unique distribution such that $\psi^i_{ct}(\sigma^{d(i)}) + (1 - \sigma^{d(i)})\nu^i_{jct}$ is also an extreme value random variable. The parameter $0 \leq \sigma^d < 1$ governs the within-group correlation in the error term $\psi^i_{ct}(\sigma^{d(i)}) + (1 - \sigma^{d(i)})\nu^i_{jct}$. As σ^d approaches zero, the model collapses to a standard logit model.

residential tract j to maximize its indirect utility:

$$(1)\qquad V^{i}_{jct} = \beta^{d(i)}_{wt} \ln w^{d(i)}_{jct} - \beta^{d(i)}_{At} \ln p_{Ajct} - \beta^{d(i)}_{Ht} \ln p_{Hjct} + \beta^{d(i)}_{at} \ln a_{jct} + \mu^{d(i)}_{jc} + \xi^{d(i)}_{jct} + \varepsilon^{i}_{jct},$$

where $\beta_w \equiv \beta_H + \beta_Z + \beta_A$.

This utility maximization problem, outlined in Berry (1994), yields a linear equation for the share $\tilde{s}^{d}_{jct}$ of individuals in group d who choose tract j relative to a base tract $\bar{j}$:[21]

$$(2)\qquad \ln \tilde{s}^{d}_{jct} = \beta^{d}_{wt} \ln \tilde{\mathbf{w}}_{\mathbf{jct}} + \beta^{d}_{At} \ln \tilde{\mathbf{A}}_{\mathbf{jct}} - \beta^{d}_{Ht} \ln \tilde{p}_{Hjct} + \mu^{d}_{jc} + \tilde{\xi}^{d}_{jct} + \tilde{\xi}^{d}_{w,jct} + \sigma^{d} \ln \tilde{s}^{d}_{j|c},$$

where $\tilde{X}_j = X_j - X_{\bar{j}}$, we normalize $\mu_{\bar{j}c}$ to equal zero, and the final term is a "nested-logit" term, where $\ln \tilde{s}^{d}_{j|c}$ is equal to the share of group d choosing tract j within CBSA c. To simplify the presentation, we use the vector $\tilde{\mathbf{A}}_{\mathbf{jct}}$ to denote the sum of the public and consumption amenity terms, $\beta^{d}_{At} \ln (1/p_{Ajct}) + \beta^{d}_{at} \ln a_{jct}$. $\mathbf{w}_{\mathbf{jct}}$ denotes a vector of time-varying accessibility to jobs in three different wage brackets, which we use to proxy for w^{d}_{jct}, the group's wage net of commute costs. $\xi^{d}_{w,jct}$ reflects the residual variation in the wages earned by group d individuals residing in location j.

Differencing this equation between 2000 and 2010, the two years in our data, we obtain our estimating equation:

$$\begin{aligned}(3)\ \ \Delta \ln \tilde{s}^{d}_{jc} \ = \ & \beta^{d}_{w,2010} \Delta \ln \tilde{\mathbf{w}}_{\mathbf{jc}} + \Delta\beta^{d}_{w} \ln \tilde{\mathbf{w}}_{\mathbf{jc,2000}} + \beta^{d}_{A,2010} \Delta \ln \tilde{\mathbf{A}}_{\mathbf{jc}} + \Delta\beta^{d}_{A} \ln \tilde{\mathbf{A}}_{\mathbf{jc,2000}} \\ & + \beta^{d}_{p_H,2010} \Delta \ln \tilde{p}_{Hjc} + \Delta\beta^{d}_{p_H} \ln \tilde{p}_{Hjc,2000} + \sigma^{d} \Delta \ln \tilde{s}^{d}_{j|c} + \Delta\tilde{\xi}^{d}_{jc} + \Delta\tilde{\xi}^{d}_{w,jc} + \epsilon^{d}_{jc},\end{aligned}$$

where $\Delta X = X_{2010} - X_{2000}$ for both variables and coefficients.[22] Note that unobserved time-invariant tract characteristics like nice weather or historical architecture cancel out in first-difference. The error term is the sum of any unobserved changes in the perceived residential quality of tract j for group d (i.e., labor supply shocks $\Delta\tilde{\xi}^{d}_{jc}$), unobserved changes in the wages earned by group d individuals residing in tract j (i.e., labor demand shocks $\Delta\tilde{\xi}^{d}_{w,jc}$) , and an additional term ϵ^{d}_{jc} capturing any remaining measurement error.

We derived equation 3 from Cobb-Douglas preferences, so it delivers an intuitive structural interpretation of regression coefficients that we will use to interpret our results. In this interpretation, coefficients on changes in characteristics from 2000 to 2010 (e.g., $\Delta\tilde{\mathbf{A}}_{\mathbf{j}}$) capture the preference levels of demographic group d in 2010 (i.e., $\beta^{d}_{A,2010}$), while coefficients on initial lev-

[21] The steps of this derivation are standard and we present them in online Appendix G.

[22] Note that $\beta^{d}_{A,2010} X_{2010} - \beta^{d}_{A,2000} X_{2000} = \beta^{d}_{A,2010} (X_{2010} - X_{2000}) + \left(\beta^{d}_{A,2010} - \beta^{d}_{A,2000}\right) X_{2000} = \beta^{d}_{A,2010} \Delta X + \Delta\beta^{d}_{A} X_{2000}$ and that $\beta_{p_H} \equiv -\beta_H$ because house price enters our regressions as a positive number.

els of characteristics (e.g., $\tilde{\mathbf{A}}_{\mathbf{j},\mathbf{2000}}$) capture changes in the preferences of demographic group d from 2000 to 2010 (i.e., $\Delta\beta^{d}_{A,2010}$).

4 Empirical Strategy

In our model, changes in residential location decisions are driven by either changes in location characteristics (including prices), $\Delta\tilde{X}_{jc}$, or changes in the preferences of the relevant demographic group for these characteristics, $\Delta\beta^{d}_{X}$. The young and college-educated might be moving downtown either because characteristics of downtown tracts changed in ways correlated with their preferences (i.e., $Corr(\Delta\tilde{X}_{jc}, \beta^{d}_{X,2010}) > 0$) or because their preferences tilted towards characteristics in which downtown tracts were already advantaged (i.e., $Corr(\Delta\beta^{d}_{X}, \tilde{X}_{jc,2000}) > 0$). Our analysis therefore relies on two key ingredients: 1) data on the initial levels and changes in the characteristics of tracts at different distance from the city center, and 2) estimates of the parameters reflecting both the levels and change in the preferences of the young and college-educated for these characteristics. We now present data summarizing the initial levels and changes in tract characteristics. We then outline our estimation procedure, identification strategy, and baseline parameter estimates. Finally, we bring these two ingredients together to quantify the contribution of each factor in explaining the urbanization of young college graduates.

4.1 Recent Spatial Trends in Jobs, House Prices, and Amenities

Figures 2 and 3 show how key tract characteristics vary with distance to the city center. Panel A of each figure shows the kernel density plot of the 2000 logged level of a variable and Panel B shows a kernel density plot of the log change from 2000 to 2010, with kernel weights based on the 2000 tract share of young and college-educated individuals. The data presented include all tracts in our estimation sample of 355 CBSAs for which a variable is available. We provide details on the construction of all variables in Appendix B.

Figure 2 presents gradients from the city center for job density, house prices, and public amenities (school quality and crime). Job density is an inverse distance-weighted average of the number of jobs in tracts surrounding each residential tract in 2002 and 2011, computed using the LODES data by wage groups. Column 1 shows gradients from the city center for the initial level and change in both low-wage (dashed blue) and high-wage (solid red) job density. The density of both low and high-wage jobs is highest near the city center, but only high-wage jobs have grown faster near the city center over the last decade, albeit only slightly.

Column 2 shows similar gradients for house prices, plotting both our main Zillow all home index (dashed blue) as well as the two bedroom index (solid red). Houses are more expensive away from the city center in 2000, but less so when focusing on two-bedroom homes, which are a better control for size and quality. House price growth from 2000 to 2010 displays a strongly negative gradient from the city center, consistent with the urban revival that we document in this paper.[23]

Column 3 shows that public amenity levels are lower near the city center, with lower-ranked schools (dashed blue, not logged) and more violent crime per capita (solid red). Schools near the city center have dropped even further in state district rankings from 2004 to 2010. Violent crime rates are decreasing everywhere as expected, but surprisingly the percentage decline in crime has not been larger near city centers from 2000 to 2010.

Figure 3 presents similar gradients for two representative consumption amenities: restaurants (dashed blue) and food stores (solid red). Column 1 shows amenity density, column 2 shows quality, and column 3 shows diversity. These indexes are based on the CES price index methodology in Couture (2013). The density of restaurants and food stores is highest near the city center, but has grown faster in the suburbs from 2000 to 2010. Rising consumption amenity density is therefore unlikely to explain urban revival. Unlike density, quality and diversity have both increased faster near city centers, especially for restaurants.

4.2 Estimation

Our base specification of the estimating equation (3):

$$\begin{aligned}\Delta \ln \tilde{s}^d_{jc} \quad = \quad & \beta^d_{w,2010}\Delta \ln \mathbf{\tilde{w}_{jc}} + \Delta\beta^d_w \ln \mathbf{\tilde{w}_{jc,2000}} + \beta^d_{A,2010}\Delta \ln \mathbf{\tilde{A}_{jc}} + \Delta\beta^d_A \ln \mathbf{\tilde{A}_{jc,2000}} \\ & + \beta^d_{p_H,2010}\Delta \ln \tilde{p}_{Hjc} + \Delta\beta^d_{p_H} \ln \tilde{p}_{Hjc,2000} + \sigma^d \Delta \ln \tilde{s}^d_{j|c} + \Delta\tilde{\xi}^d_{jc} + \Delta\tilde{\xi}^d_{w,jc} + \epsilon^d_{jc}\end{aligned}$$

includes consumption amenity densities, job densities, and the “all home” house price index to reflect $\mathbf{w_{jc}}$, $\mathbf{A_{jc}}$, and p_{Hjc}, respectively. The dependent variable is the 2000 to 2010 log change in the share of age-education group d that lives in tract j of CBSA c relative to a base tract. We also include the 2000 levels of own-group tract shares and population density to control for changes in the taste for living among one’s own group type (i.e., homophily) and other unobserved endogenous amenities.[24] In robustness checks below, we add explanatory variables for public amenities, such as school quality, crime, and transit times, and other dimensions of the spatial distribution of consumption amenities, such as quality and diversity of establishments. These variables are excluded from our base specification because we either have no instrument

[23] Generally, gradients from the city center are similar in 2000 and in 2010, and the trends in a sample restricted to large CBSAs are visible in the current sample of all CBSAs. House prices are a notable exception: in the 10 largest CBSAs, the house price gradient from the city center *reverses* from 2000 to 2010.

[24] Controlling for additional 2000 demographic shares – such as share of college-educated individuals or share of a given age-education group who were 10 years younger in 2000 (“stayer”) - does not affect any of this paper’s main results.

for them or only limited spatial coverage.[25]

4.2.1 Identification

Identifying the effect of neighborhood characteristics on residential choice is challenging. The first-difference regression controls for time-invariant tract characteristics that could be correlated with our regressors. However, our regressors could still be correlated with unobserved changes in tract quality ($\Delta\tilde{\xi}^d_{jc}$) or local wage premia ($\Delta\tilde{\xi}^d_{w,jc}$). In robustness checks, we add an array of controls, including changes in own-group shares in nearby tracts, to alleviate omitted variable bias concerns. Neither first-differencing nor adding controls, however, resolves reverse causality, which affects variables in changes. For instance, an influx of young college graduates in response to unobserved shocks to tract quality or nearby wages may attract amenities and jobs and raise house prices. We describe instruments for our variables in changes below, devoting more time to the most novel of these, the instrument for the change in consumption amenity density. In section 5.2 we propose, as an alternative to instrumental variables, an identification strategy based on the addition of a workplace fixed-effect. This strategy requires joint workplace-residential choice data, and convincingly frees coefficients on residential characteristics from biases due to changes in workplace location.

Instruments for Consumption Amenity Density To design an instrument for the change in consumption amenity density, we seek factors that explain changes in amenity location from 2000 to 2010 but are exogenous to changes in neighborhood demographics. We exploit within-amenity type national variation in the business expansion strategy of chains or finely-defined industries, in conjunction with local variation in the preexisting commercial environment. Our instrument draws both from the Bartik (1991) instrument methodology familiar in labor and urban economics, and from findings in the industrial organization literature on the importance of cannibalization concerns in determining retailer entry (Toivanen and Waterson, 2005; Igami and Yang, 2016).

The instrument's computation proceeds in two steps. First, we model business expansion strategies by regressing SIC8-level establishment entry from 2000 to 2010 in each tract on variables capturing the preexisting commercial environment in 2000 at different distances from the tract centroid.[26] Second, we predict net establishment entry in each tract by summing the

[25] We have no instrument for our public amenity variables and only have school and crime data for a subset of CBSAs. Establishment quality is only available for a subset of establishments and we have no instrument for diversity.

[26] We expect entry to decrease with the concentration of establishments offering similar services in very close proximity, due to competition and cannibalization concerns. Past a very small radius from an entry point, the reverse may be true and entry may increase with broader density of establishments in the same SIC8 or chain, since these indicate proximity to the chain's upstream suppliers or distribution centers and some preexisting mar-

fitted values of these regressions over all SIC8 codes in each amenity type (e.g., all SIC8 codes in the "restaurant" type). We then use this predicted net entry to compute the predicted change in each of our nine consumption amenity density index. This predicted change is our instrument.

The first-step regression predicts exit and entry at the SIC8 level. We define n_{jt}^{sic8} as the number of establishments within a given SIC8 code in tract j in period t, so our dependent variable is $n_{j10}^{sic8} - n_{j00}^{sic8}$. We model entry and exit as a function of the commercial environment, more precisely of $n_{j00,dist}^{sic8}$, $n_{j00,dist}^{sic6|8}$ and $n_{j00,dist}^{sic4|6}$, which represent the number of establishments in the same SIC8, in the same SIC6 but not the same SIC8, and in the same SIC4 but not the same SIC6, within distance interval $dist$ from the centroid of tract j. The four distance intervals are 0-1, 1-2, 2-4, and 4-8 miles. For each SIC8 code, we estimate the following regression, in which each observation is a tract j:

(4)
$$n_{j10}^{sic8} - n_{j00}^{sic8} = \alpha^{sic8} + \sum_{dist \in \{[0,1],[1,2],[2,4],[4,8]\}} \left(\beta_{dist}^{sic8} n_{j00,dist}^{sic8} + \beta_{dist}^{sic6|8} n_{j00,dist}^{sic6|8} + \beta_{dist}^{sic4|6} n_{j00,dist}^{sic4|6} \right) + \varepsilon_j^{sic8}.$$

The estimation results indicate that competition and cannibalization concerns are strong predictors of establishment entry and exit.[27] In 93 percent of SIC8 codes, the presence of establishments in the same SIC8 within 0-1 miles significantly reduces entry in a tract. Agglomeration forces dominate for establishments in related but less similar product spaces: the 0-1 mile coefficient for same SIC6 and for same SIC4 are positive and significant in about 50 percent of cases and negative and significant in less than 10 percent of cases.

Our consumption amenity quality indexes can only be computed for establishments that are part of a rated chain. To instrument these indexes, we predict entry at the chain level (e.g., Pizza Hut) instead of the SIC8 level (pizza restaurants). The results of this entry regression highlight the strength of within-chain spatial economies of scale.[28] That is, the effect of proximity to an establishment in the same chain switches from a negative to a positive sign as distance increases beyond 2 miles. This implies that while chains avoid locating right next to an existing outlet, they tend to enter markets that they have already penetrated.

The second step of our methodology sums up the fitted value of these entry and exit regressions to compute predicted changes in each consumption amenity index. We start from the vector of all establishments in 2000, and use the fitted value from the entry regression to add "predicted" establishments to the centroid of each tract. Using this vector of "predicted"

ket knowledge. In addition to within-SIC8 and within-chain scale economies, we also account for sector-level co-agglomeration externalities, in the form of positive spillovers from local activity from non-competing or differentiated firms within the same industry. In addition to these direct effects, we expect to capture location-specific barriers to entry, such as natural or regulatory supply constraints.

[27] See Table A.2 in Appendix B.3 for aggregate results on the predictors of entry and exit across all 559 SIC8 codes used to define our consumption amenity density indexes.

[28] See Table A.3 of Appendix B.3.

2010 establishments, we compute a "predicted" amenity density index for 2010. The difference between the 2010 predicted index and the 2000 actual index is our instrument for the actual change in the index.

First stage statistics indicate that these instruments are relevant.[29] To understand why, first recall that the instrument depends on the mix of establishment types in a tract in 2000. The variation in the instrument across tracts depends on two key factors. First, the tract's proximity to establishments in SIC8 categories or chains that have strong cannibalization or competition concerns (as estimated by the large, negative coefficients on the count of establishments in the same SIC8 in our entry regression, β_{dist}^{sic8} in equation 4). Second, the tract's proximity to establishments upstream from SIC8 categories or chains that have high agglomeration potential (i.e., those with large positive coefficients on the counts of establishments in the same SIC6 but different SIC8 or same SIC4 but different SIC6 category in our entry regression, $\beta_{dist}^{sic6|8}$ and $\beta_{dist}^{sic4|6}$ in equation 4). Our estimates from equation 4 indicate, for example, that oyster bars are much less likely to enter in close proximity to other oyster bars, but are more likely to enter near other seafood restaurants. We therefore predict faster growth for the oyster bar establishment count between 2000 and 2010 in those tracts that are near few oyster bars, but close to other seafood restaurants.

A valid instrument must also be exogenous i.e., uncorrelated with the error terms in equation 3 conditional on other regressors. If only supply-side factors drive business entry decisions (e.g., cannibalization and competition concerns or technological factors like supplier networks) then this exclusion restriction holds. The exclusion restriction could be violated if the instrument correlates with supply shocks that affect the unobserved group-specific wage premia $\Delta\tilde{\xi}_{w,jc}^{d}$, but it is hard to find a story such that this would be true, especially for the college-educated groups who are unlikely to work in restaurants and food stores. The instrument is also robust to changes in local demand because the cross-tract variation in the instrument is determined by tract-invariant, national coefficient estimates interacted with the local, but predetermined, business mix. A violation of the exclusion restriction would require that demand factors not controlled for in equation 3 drive our estimated business expansion strategies. Specifically, this would require that locations characterized by a specific amenity mix in 2000, controlling for initial amenity density, experienced correlated unobserved tract quality shocks between 2000 and 2010 ($\Delta\tilde{\xi}_{jc}^{d}$). The granularity of the data implies that such correlated shocks require more than the systematic presence or absence of a handful of product categories from locations with correlated shocks. For the restaurant amenity index, for instance, our instrument is calculated from the business expansion strategy of 70 SIC8 categories. A violation would require an unlikely similarity in the presence or absence of establishments across these types in locations with correlated shocks.

[29] See Table A.4 of Appendix C.

Instruments for Housing Prices To overcome the endogeneity of house price changes, we exploit the correlation between housing prices and exogenous natural amenities identified by Lee and Lin (forthcoming). We expect geographic features like oceans and mountains to act like anchors imposing supply constraints on land, thereby driving up relative house price levels, as described in Gyourko et al. (2013). These supply constraints may also amplify the reaction of house prices to demand shocks, so we also use these natural amenities as instruments for changes in house prices. Our vector of geographic features includes the log Euclidean distances (in km) of the centroid of tract j from the coast of an ocean or Great Lake, from a lake, and from a river, the log elevation of the census tract centroid, the census tract's average slope, an indicator for whether the tract is at high risk of flooding, the log of the annual precipitation, and the log July maximum and January minimum temperatures in the tract averaged over 1971 and 2000.[30] As in Bayer et al. (2007), our instrument for tract j uses geographic features of tracts one to three miles away, controlling for the average geographic features of tracts within one mile. The key exclusion restriction is that geographic features further than one mile away from a tract do not impact demand for living in that tract, conditional on the geographic features within one mile. As an additional instrument for changes in housing prices (and for the levels of local demographic shares), we include historical tract-level 1970 population shares, by age and by education group.

In a robustness check, we exploit the Cobb-Douglas preference structure to simply difference out the CEX housing expenditure share of each age-education group from the utility function. Endogeneity of housing is then no longer an issue because housing variables are used to adjust the left-hand side variable and are excluded from the right-hand side regressors. This approach, taken in Baum-Snow and Hartley (2016), replaces a reliance on assumptions related to instruments with a reliance on assumed demand structure.

Instruments for Employment Density We use standard Bartik instruments for changes in the job density indexes. The LODES data include jobs in our three wage groups by 20 North American Industry Classification System (NAICS) sectors. This industry breakdown allows us to obtain Bartik predictions of wage group-specific employment growth that depend on the industrial composition of each tract, and on national industry growth.

Instruments for Nested-Logit Within-CBSA Share Instrumenting the change in the nested-logit share of type d individuals within CBSA c who live in tract j, $\Delta s^d_{j|c}$, requires exogenous factors affecting the attractiveness of tract j relative to all other tracts in its CBSA

[30] Such instruments have been criticized by Davidoff (2016) in the context of cross-CBSA regressions. Davidoff (2016) shows that geographical supply constraints are correlated with demand factors and that constrained cities like New York and San Francisco also have more productive workers. Our within-CBSA instrument is less vulnerable to this criticism.

c. For each instrument described above, we compute $instr(\Delta s^d_{j|c})$ as the average difference between the instrument in tract j and that in all other tracts k in CBSA c:

$$instr(\Delta s^d_{j|c}) = \frac{\sum_{k \in c_j \, and \, k \neq j} (instr_j - instr_k)}{N_{c_j}},$$

where N_{c_j} is the number of tracts in the same CBSA c as tract j.

4.2.2 Regression Results

Table 1 presents ordinary-least-squares (OLS) regression results for the nested-logit model in equation 3.[31] For the sake of parsimony, this base specification includes two representative consumption amenity density indexes, one for restaurants and one for food stores. Restaurants and food stores are the non-tradable service and the tradable retail amenity type with by far the largest CEX expenditure and NHTS trip share. Table 2 follows with the same specification in instrumental variables (IV).[32] In both tables, Panel A show the coefficients for the three college-educated age groups: 25-34 year-olds (young), 35-44 year-olds (middle-aged) and 45-64 year-olds (old). Panel B shows the same coefficients for the non-college educated age groups. For each age group, column 1 shows coefficients for variables in changes (i.e., the 2000 to 2010 first-difference in that variable) and column 2 shows coefficients for variables in initial levels (i.e., the 2000 value of the corresponding variables). Most coefficients are significant at the 1 percent level. The OLS coefficients are generally of the same sign as their IV counterparts but often smaller in magnitude, likely as a result of attenuation bias. These differences in magnitude do not impact our main results on the relative importance of various factors in explaining urban revival, which are invariant to scaling all coefficients up or down and, in particular, robust to whether we use the OLS or IV coefficient estimates.[33]

In the discussion that follows, we adopt the structural interpretation of the coefficients derived from the model in section 3. The coefficient on a variable in change has an interpretation as a preference parameter in 2010, $\beta^d_{X,2010}$. A positive sign denotes attraction to this tract characteristic. The coefficient on a variable in initial level has an interpretation as a change in

[31] The within-CBSA share is instrumented even in the OLS specification to ensure that it does not capture too much of the variation in the data. Our main results hold without taking this precaution.

[32] See Table A.4 in Appendix C for detailed first-stage statistics for all instrumented variables. The reduced-form first-stage statistics all easily reject that the instruments are irrelevant. The Sanderson and Windmeijer (2016) conditional F-statistics are lower but the lowest is at 9.17 for restaurants, which is at the margin of being weak using standard rule of thumbs and still implies an IV estimator that is considerably less biased than the OLS estimator. The other non-tradable service amenities that we include next have stronger instruments.

[33] All of the presented coefficients are standardized. For example, the positive IV coefficient of 0.212 on the change in high-income jobs for the young college-educated group means that moving up one standard deviation in the tract-level distribution of this change translates into a 0.212 standard deviation increase in the share of young college-educated individuals living in a tract.

preference from 2000 to 2010, $\Delta\beta_X^d$. We provide external validity for this interpretation in section 6, using data on expenditure and trip shares.

The IV coefficients in Table 2 generally have the expected sign. Considering coefficients on the variables in changes, all three college-educated groups have a preference for high-wage job density and all six age-education groups have a preference for restaurant density, which is strongest for the young and college-educated. The coefficient on the change in house prices has the expected negative sign in four out of six age-education groups (three significantly so, relative to only one in OLS).[34]

Considering coefficients on variables in initial levels, we find that the young and college-educated have the most positive change in preferences for restaurant density of any age-education group, which is also true in OLS, and highly significant. For food stores, however, we estimate a negative change in preferences for young college graduates. These preference patterns of young college graduates for non-tradable services and tradable retail generalize to consumption amenities other than restaurants and food stores. Column 1 of Table 3 adds bars and apparel stores to the set of amenities in our base specification, and column 2 includes all nine consumption amenities, in both OLS and IV. The coefficients confirm the attraction, getting stronger through time, of the young and college-educated to non-tradable services in general (restaurants, bars, gyms and personal services) and their reluctance to locate near tradable retail and activities like museums and amusement parks.[35,36] These stark preference patterns do not hold for other age-education groups.

One may worry that given the urbanization of young college graduates, their location choice must necessarily correlate with initial levels of variables that are urbanized, like non-tradable service amenities. It is worth noting that other consumption amenity types (e.g., tradable retail) are also urbanized but do not have initial levels with positive coefficients for young college graduates.[37] Further, our coefficients are robust to including a direct control for tract distance to the city center in column 3 of Table 3.

[34]We take this success rate as reasonable support for the specification, but note that positive coefficients are inconsistent with a structural interpretation as (the negative of) demand for housing. As explained later, our main results hold in regressions on housing-adjusted population shares that exclude housing, as in Baum-Snow and Hartley (2016).

[35]The four non-tradable service amenity density indexes are correlated and the restaurant coefficients decrease in size when we add other amenities. This is consistent with endogenous consumption amenities and dynamics that amplify any exogenous increase in one amenity. We explore issues of endogenous amenities and homophily further in subsection 5.4.

[36]The young and college-educated also have the most positive change in preferences for bars in the OLS specifications with 4 and 9 amenities, and the second most positive change in the IV specifications with 4 and 9 amenities. They have the second most positive change in preferences for gyms in both the OLS and IV specifications with 9 amenities. Personal services like hair, nail and beauty salons stand out as our estimates suggest that non-college educated groups have a more positive change in preferences for this amenity type.

[37]This result is not surprising. Built amenities that one rarely visits are probably dis-amenities, and indeed most jurisdictions have zoning regulations preventing commercial use near residential areas.

4.3 Which Variables Explain Urban Revival?

We now combine the preference estimates of the last section with the spatial distribution of each variable in urban relative to suburban areas from section 4.1 to identify the main factors explaining urban revival. Intuitively, a variable contributes to urban revival if: 1) young college graduates like it, and 2) it is highly prevalent downtown. In terms of our empirical framework, such a variable has: 1) a positive regression coefficient for young college graduates in Table 2, and 2) a negative gradient from the city center in Figures 2 and 3. To visualize these contributions, we again use kernel plots, which illustrate the contribution of each variable to log changes in the share of young college graduates living at a given distance from the city center.

We start with our regression equation 3 to extract the contribution of each variable to the log change in the share of a given demographic group living in a given tract. If $\tilde{X}_{jc,k}$ is the value of a regressor k and $\widehat{\beta_k^d}$ is the coefficient on that regressor for group d, then the fitted change in tract jc's share of demographic group d's national population, relative to base tract, is:

$$\widehat{\Delta \ln \tilde{s}_{jc}^d} = \sum_k \widehat{\beta_k^d} \tilde{X}_{jc,k}.$$

The contribution of each regressor k is $\widehat{\beta_k^d} \tilde{X}_{jc,k}$. We compute this contribution for each tract and plot it against the population-weighted distance of this tract from the city center.

Figure 4 shows the urbanizing contribution kernel plots for young college graduates for all explanatory variables, using coefficients from our base IV specification in Table 2.[38] The left-hand plots shows the contribution of variables in changes, and the right-hand plot shows that of variables in initial levels. As an example of how to interpret these plots, consider the urbanizing contribution of change in high-wage job density. Recall from Figure 2 that high-wage jobs grew only slightly faster downtown, i.e., they have a mild negative gradient from the city center. The change in high-wage job density has a large positive standardized coefficient, so it is an important determinant of location choice for young college graduates. Overall though, changes in high-wage job density contribute little to urbanizing young college graduates, because of their relatively flat gradient. That is, young college graduates value proximity to high-wage jobs, but these have not been growing much faster in urban relative to suburban areas. To make comparisons of contribution across variables easier, we normalize the contribution of each variable at the outskirts of a CBSA to zero. As a result, the intercept of each plot with the city center provides a ranking of each variable according to the importance of its contribution to urbanizing a given group.

Figure 4 shows the key result of the paper: the initial level of restaurant density, a non-

[38] The contribution plots for all other age-education groups are in online Appendix F and additional derivations of these plots are in online Appendix G.

tradable service, is the most important contributor to the urbanization of the young and college-educated. The young and college-educated also have the largest coefficient on the initial level of restaurants relative to other age-education groups, so its urbanizing contribution is largest for the young and college-educated. Restaurants are representative of non-tradable services more generally, and we replicate this exercise including all nine consumption amenities and find that the four non-tradable service amenity levels (restaurants, personal services, bars and gyms) are the four most important contributors to urban revival. Using our structural interpretation of the regression coefficients, the model suggests that the main contributing factor to the rising share of young college graduates near city centers is an increasing preference for urbanized non-tradable service amenities.

Table 4 quantifies these results. The table shows the contribution of non-tradable services initial levels in urbanizing young college-graduates as a share of the contribution of all types of variables in the model (levels and changes in jobs, house prices, tradable retail, activities, and non-tradable services, levels of population density and demographic share). The table also shows how the contribution of non-tradable services initial levels ranks relative to that of other types of variables. Each row of the table shows share and rank for a different specification. For the base IV specification with all nine consumption amenities, the four non-tradable service initial levels rank first and contribute 83 percent of the urbanization of young college graduates. Other variables making a positive contribution account for the remaining 17 percent.[39] In the base OLS specification this share is 50 percent, and in a non-nested CBSA fixed-effect IV specification this share is 75 percent. Other rows of Table 4 show the robustness of these conclusions for different specifications discussed later in the paper. We note that the variation across specifications in these tables often reflects the impact of re-estimating the model on a smaller selected sample rather than the impact of adding control variables.[40] One could also worry that the small contribution of variables in changes is due to the inclusion of variables in levels that are collinear. To investigate this, Table A.6 in Appendix C compares coefficients from our base specification with coefficients from the same specification but with only variables in changes. Coefficients for variables in changes remain of the same sign, significance, and order of magnitude after removing all variables in levels.

One way to characterize this analysis is as an attempt at distinguishing the role of changes

[39] In absolute magnitude, the contribution of non-tradable service initial levels to young college graduate growth near city centers is 1.8 times *larger* than the actual growth documented in Figure 1. However, other factors are also pushing against the urbanization of young college graduates. Aggregating over the contribution of all variables, the model correctly fits the urbanization of the young and college-educated and to a lesser extent of the middle-age and college-educated, and the suburbanization of every other group. Overall, fitted values from the model generate less urbanization than what actually happened.

[40] In Table A.5 of Appendix C we provide the rank of the contribution of restaurants out of every variable in the model for every specification shown in the paper with two consumption amenities, in both IV and OLS. This table also shows how the size of the young college graduates' restaurant coefficients ranks compared to that of other age-education groups.

in characteristics from the role of changes in the willingness to pay for those characteristics in explaining the difference in the spatial distribution of the young and college-educated between 2000 and 2010. This can be thought of as an application of the Oaxaca (1973) decomposition commonly employed in the labor literature attempting to understand wage differentials between two worker types (see, for example, Card and Krueger, 1992). Fortin et al. (2011) highlights that this decomposition is sequential, in the sense that the order of the decomposition matters for the conclusion. In our case, this implies that the use of the 2000 level variables, rather than 2010 level variables, could matter. For the purposes of estimation, we use the 2000 level variables since they are less subject to reverse causality biases than the 2010 level variables, which are mechanically correlated with the 2000-2010 change variables. Using the parameter estimates obtained from this base specification, our contribution plots are almost invariant to whether we decompose the log share using 2000 shares as in $\widehat{\Delta \ln \tilde{s}_{jc}^{d}} = \sum_{k \in K_1} \left(\widehat{\beta_{k,2010}^{d}} - \widehat{\Delta \beta_{k}^{d}} \right) \Delta \tilde{X}_{jc,k} + \sum_{k \in K_2} \widehat{\Delta \beta_{k}^{d}} \tilde{X}_{jc,k,2000}$ or using 2010 shares as in $\widehat{\Delta \ln \tilde{s}_{jc}^{d}} = \sum_{k \in K_1} \left(\widehat{\beta_{k,2010}^{d}} \right) \Delta \tilde{X}_{jc,k} + \sum_{k \in K_2} \widehat{\Delta \beta_{k}^{d}} \tilde{X}_{jc,k,2010}$.

4.3.1 Why Did Urban Revival Happen Primarily in Larger Cities?

In the urbanizing contribution plots above, the spatial distribution of each variable comes from our estimation sample of all tracts in all CBSAs. However, our stylized facts document that the urbanization of the young and college-educated is primarily a large city phenomenon. Figure 6 shows that non-tradable service levels can explain this as well. The plot on the left shows the contribution of initial level of restaurant density to urbanizing young college graduates for four groups of CBSAs ranked by population: top 10, top 11-50, top 50-100, and all other CBSAs. We find that the initial level of restaurant density - as well as that of every other non-tradable services - indeed provides a stronger urbanizing push in larger CBSAs, because larger CBSAs have higher density of non-tradable services near their city centers relative to their edges.

5 Robustness

We now present various robustness exercises where we explore the role of other factors for which our data is more limited, and therefore choose not to include in our main analysis.

5.1 Crime, School, and Transit

Public amenities like school quality, crime rates and transit availability are presumably important determinants of residential location choices. The well-documented decline in central city violent crime since 1990 (e.g., Levitt, 2004) is a potential explanation for urban revival. Anecdotal evidence suggests that school quality drives the suburban location choice of families with children. Transit availability is, on the other hand, a prominent characteristic of city centers. Table 5 reports coefficients for our base IV regression specification, adding controls for the changes and initial levels in local school district rankings and per capita violent crime, and for 2014 level of the transit time of a five-mile trip. Adding these public amenities variables reduces our sample size by two thirds. Table 4 documents that the initial level of non-tradable services is still the most important determinant of urbanization for the young and college-educated after including public amenities. The same is true in the corresponding OLS regression. Our regression results therefore do not support a key role for public amenities in explaining urban revival.

The sign of the coefficient on change in violent crime per capita is negative and significant for all groups *except* for the young and college-educated, who have a coefficient near 0 indicating little aversion to crime. The direction of the endogeneity bias is in theory ambiguous, but Autor et al. (2017) show that gentrification reduces crime, which suggests that our coefficient on crime change is biased in the direction of finding a larger negative impact on the location choice of college-educated groups. We also find no evidence that young college graduates move to areas with low crime in 2000. In fact, we find the reverse, signaling that a reduced aversion to crime makes "gritty" areas ripe for gentrification. Moreover, others (e.g., Kneebone and Raphael, 2011) have documented that the decline in urban crime was faster in the 1990s, a period over which the widespread urban revival that we document is not yet happening. To test the hypothesis in Ellen et al. (2017) that college-educated individuals move to central cities that experienced a *prior* decline in crime, we estimate the specification in Table 5 but use the crime levels from 1990 and changes from 1990 to 2000 (not shown). We also find a negative and significant coefficient on prior crime change that contributes to urbanizing the young and college-educated. However, this contribution is negligible compared to that of non-tradable service levels.[41] Finally, as noted in Edlund et al. (2015), there is anecdotal evidence that central locations in large European cities are also experiencing rising demand from the young and college-educated, despite not having had the high rates and subsequent decline in crime that U.S. central cities experienced. Combined, these pieces of evidence do not preclude a significant role of crime decline in generating favorable conditions for urban revival, but they suggest that the root of the widespread, recent urbanization of young college graduates lies elsewhere.

We also find that improvements in school quality are unlikely to be a factor in urban revival. Figure 2 showed that the relative ranking of schools near city centers worsened from 2004 to 2010. The young and college-educated show no preference for highly ranked school districts in

[41] We note two differences between our approach and that in Ellen et al. (2017). First, we focus on the younger college-educated group in particular and on areas smaller than central cities, motivated by the stylized facts in section 2. Second, our empirical approach infers a general aversion to violent crime regardless of the area, rather than assigning specific aversion to "central city" crime to different groups.

Table 5, unlike the middle-aged and older college-educated.

Finally, transit efficiency in 2014 does not correlate with a positive influx of the young and college-educated. Moreover, transit users are disproportionately low income (LeRoy and Sonstelie, 1983; Glaeser et al., 2008), which is inconsistent with urban transit explaining the urbanization of young college graduates.

5.2 Commuting

The simultaneous determination of work and residential locations is a key identification concern in residential choice model estimation. This problem is straightforward: young and college-educated workers can reduce their commute costs by moving to areas experiencing an influx of firms hiring them. At the same time, firms may move closer to a young, educated talent pool, which is often the given justification for new downtown offices by employers like Amazon, Twitter or Google (Johnson and Wingfield, 2013). To resolve this simultaneity problem, we use the LODES commute data in 2002 and 2011 to deliver *within*-work tract preference coefficients. This sharper identification strategy plausibly removes the bias due to job reallocation from coefficients on residential characteristics.

Including a workplace fixed-effect costs us the ability to estimate the importance of access to jobs in residential location choice. However, we can use a workplace fixed-effect model to estimate coefficients on other residential characteristics like consumption amenity density, and compare these fixed-effect coefficients with those from the pure residential choice model that we estimate everywhere else in the paper. Crucially, this comparison reveals little bias in estimating a pure residential choice model. These results are in Appendix D.[42]

We now provide a visual representation of recent changes in commute patterns, which supports our main findings by highlighting the role of residential amenities in driving the urbanization of high-wage workers. We aggregate the LODES data into commute matrices (Figure 7) showing the 2002-2011 percentage change in the number of workers living and working at various distances from the city center. Residential distance from the city center is fixed within each row of the matrix and workplace distance from the city center is fixed within each column. The seven row/column distance bins are: between 0-1 mile from the city center, 1-2, 2-4, 4-8, 8-16, 16-32, and 32+ mile. The color of a cell varies from dark blue for the most negative change to dark red for the most positive change. The matrix in Panel A displays data for all workers living and working in all CBSAs. Cells turn from blue to red as one looks down each

[42]Following Glaeser et al. (2001) and Moretti (2012), academics have debated the relative importance of consumption versus production in explaining college-educated location choices. A key contribution is Diamond (2016), who uses Bartik instruments for local labor demand interacted with housing supply elasticities to show that local labor demand shocks matter more than local amenity in the *cross-city* college-educated location choice. We show here how commute data provides a sharper identification strategy to distinguish consumption from production in a within-city context.

column, which is consistent with the national residential decentralization trends documented in Figure 1. Looking left to right along each row shows a similar and simultaneous pattern of job decentralization.

The stylized facts in Section 2, however, indicate that certain locations and groups have been bucking this national suburbanization trend in the last decade. Panel B of Figure 7 focuses exclusively on high-wage workers, again in all CBSAs. Unlike the general working population, high-wage workers are not systematically decentralizing their workplaces and residences. We instead observe increases in the number of high-wage workers either commuting from the suburbs to jobs downtown or reverse-commuting from downtown to jobs in the suburbs. Overall, average commute length increased slightly for high-wage workers from 2002 to 2011, with the largest increase for those living near the city center, due to the rising share of reverse commuters.

Finally, focusing on high-wage workers in the 10 largest CBSAs, Panel C displays commute patterns consistent with our stylized facts for college-educated individuals from section 2. High-income workers in large CBSAs are living and working closer to the city center in 2011 than in 2002. This correlation alone does not tell us whether high-wage workers are following jobs or whether jobs are following high-wage workers. To distinguish these explanations, one needs to consider commute patterns within each column (i.e., holding workplace location fixed). This mirrors our workplace fixed-effect identification strategy. The columns in Panel C demonstrate that, holding workplace distance from the city center fixed, high-wage workers in large cities live relatively closer to the city center in 2011 than they did in 2002. Reverse commuting also increased, illustrated by redder cells above the diagonal. These patterns imply that job location alone cannot drive the urbanization of high-wage workers in large cities. The longer commutes that high-wage workers are willing to incur to live near city centers are consistent with our main finding in this paper on the increasing attractiveness of downtown residential amenities.

Though based on the location patterns of a different set of people than our main census results (LODES high-wage workers are a much larger group than the young and college-educated), these commuting results support our conclusion that residential amenities contribute to urban revival. The importance of residential amenities in location choice should not come as a surprise. For instance, in the 2009 NHTS, high income individuals in the urban areas of large cities on average take more trips to non-tradable service amenities than commute trips to work.

5.3 Alternative Housing Cost Data

The house price index in our base specification is the Zillow House Value Index for all homes, which measures the median house price level in a tract. This index does not capture rental units that are prevalent in urban areas, and it depends on the average size and quality of housing units, as well as on market supply and demand conditions. We test the robustness of our results to our treatment of house prices in four different ways. First, we replace the all homes index with Zillow's two bedroom index, as well as with Ferreira and Gyourko (2011)'s hedonic price index that controls for more housing characteristics. Second, we include HUD's Fair Market Rent Series, in addition to the all homes house price index. Third, we use the Cobb-Douglas preference structure to remove endogenous housing prices by differencing out CEX group-specific housing expenditure shares from utility and running regressions on these housing-adjusted shares. Fourth we include median housing age in a tract as a proxy for quality. Panel C in Table 4 shows that initial levels of non-tradable services still make the largest contribution to urban revival in all cases.[43]

In an online Appendix H we investigate the possibility that urban revival is explained by limited mortgage credit availability following the housing crisis and recession of 2007-2009, which pushed individuals into urbanized rental housing. We find no support for this hypothesis in the ACS and IPUMS data, which instead suggests that urban revival starts before the recession, during a period of rising homeownership rates.

5.4 Homophily

Our regressions include the 2000 level of the own-group tract share to assess whether increases in the preference for living near one's own type, i.e., increasing homophily, could attract the young and college-educated downtown. However, we cannot include the change in own-group tract share, because it would mechanically co-vary with our dependent variable. Our coefficients therefore capture both the direct impact of a shock to tract characteristics or preferences over these characteristics, as well as any amplification of this direct effect through the resulting changes in demographics that make tracts more attractive.

To evaluate the share of our coefficients that reflect a direct utility from a variable as opposed to amplification related to homophily, we add a control for the change in the average own-group share near a given tract j, measured as an inverse-distance weighted average of the own-group share in all tracts excluding the given tract. We also control for the change in population density, computed in the same way. Table 6 shows the results of this specification for the young college-educated group side-by-side with the base specification results for this group. Adding these controls tempers the magnitude of almost all coefficients, as expected. However, our key result holds: Panel B of Table 4 shows that the initial levels of non-tradable services are still the most important factor drawing the young and college-educated downtown.

[43]Table 4 shows that the contribution of non-tradable services still ranks first but drops to 30 percent after adding housing age. This large drop is not a robust result. In the exact same IV specification with two instead of nine amenities, for instance, the contribution of non-tradable services is still 74 percent. Neither the initial level nor the change in housing age makes a positive contribution to urbanizing the young and college-educated.

6 External Validity: NHTS Trip Shares and CEX Expenditure Shares

We now use NHTS trips and CEX expenditures to corroborate the rising preference of young college graduates for non-tradable services. We focus on the four consumption amenity types that have a reasonable counterpart in both the CEX and the NHTS: two types of non-tradable services (restaurants and bars) and two types of tradable retail (food stores and apparel stores). For each amenity type in each year, we compute the average expenditure share from the CEX and average trip share from the NHTS across all individuals within each of our six age-education groups.[44]

The results for restaurants and food stores are in Panels A and B of Figure 8 (the results for bars and apparel stores are in Appendix C). Each panel shows the level shares in 2010 in the top row and the change in shares from 2000 to 2010 in the bottom row, with CEX expenditure in the left column and NHTS travel in the right column. Panel A shows that the young and college-educated have the largest expenditure and travel shares to restaurants of all six age-education groups. They have either the most positive change or the second most positive change (not significantly smaller from the most positive change) in expenditure and travel shares to restaurants. The exact same patterns hold for bars i.e., for trips to "go out" and for expenditures on alcohol away from home. These results help validate our model's conclusion that young college graduates have a stronger preference for restaurants and bars than other age-education groups, and have experienced the most positive change in such preferences from 2000 to 2010. The absolute magnitude of the change is hard to interpret because non-tradable services are luxuries and our post-period NHTS and CEX data overlaps with the Great Recession. Changes in restaurant expenditures are therefore negative for all groups; though, remarkably, changes in trip shares to restaurants (+4 percent) are positive for the young and college-educated. This positive change is even more pronounced for bars, with a 33 percent change in expenditure shares and an 8 percent change in travel shares.

The patterns above do not hold for a tradable retail like food stores, shown in Panel B. The young and college-educated have the lowest expenditure share on food stores; and the second lowest trip share to buy goods (groceries/apparel/hardware).[45]

[44] CEX expenditures on "restaurants" includes all food away from home, except alcohol which we classify as "bars." The NHTS identifies trips to restaurants, and we match the trip codes for "go out" (bar, entertainment, theater, sports event) with "bars" and the category "buy goods" (groceries/clothing/hardware store) with both food and apparel stores. To maximize sample size, we aggregate quarterly CEX data over 5 years (i.e., 1998-2002 and 2008-2012). We use the 2001 and 2009 NHTS. The CEX reports expenditures at the household ("consumption unit") level, so we attribute the expenditure shares of the household to its individual members. The NHTS records all trips on a single survey day separately for all members of participating households. All details appear in Appendix A.

[45] The CEX and NHTS each include categories that roughly correspond to expenditure on and trips to personal

Of course, travel and expenditure shares may not capture preferences if travel costs decline with proximity to consumption amenities, and if the young and college-educated live closer to these amenities in 2010. Using confidential geo-coded NHTS data, we verify that the travel patterns above hold when controlling for amenity density in a traveler's residential tract. We do not have geo-coded data for the CEX.

This increase in young college graduates' expenditures on and trips to non-tradable services relative to other groups supports our structural interpretation of the model's coefficients on non-tradable services initial levels. That is, CEX and NHTS data lend credence to our regression finding that young college graduates experienced a positive change in their preferences for non-tradable services like restaurants and bars that is larger than that for other age-education groups.

7 Explaining Changing Tastes

Our empirical analysis so far suggests that changing preferences for non-tradable service consumption amenities play an important role in the urbanization of the young and college-educated. In this section, we investigate four potential drivers of these changing preferences.

7.1 Homophily and Non-Tradable Services

We first explore the role of homophily in the rising preferences of young college graduates for non-tradable services (distinct from its more general role in the changing preference for downtown neighborhoods, as discussed in Section 5.4 above). We hypothesize that restaurants, bars, gyms and personal services establishments are increasingly attracting young college graduates because they provide opportunities to meet, network, and date other young college graduates. In this sense, urban revival is a dynamic process through which urban non-tradable services become even more desirable to the young and college-educated as others in the same group move in to patronize these establishments. This dynamic does not in itself provide a complete explanation for recent changes in preferences, but it offers an amplification mechanism and a hint at why young college graduates' experienced a positive change in preferences for proximity to non-tradable service amenities.

To test this hypothesis, we interact the level of and change in amenity densities with the 2000 share of the young and college-educated in a tract. Table 7 contains our base specification with these interactions in IV in columns 1 and 2, and the same specification in OLS in columns 3 and 4. Both specifications deliver similar results. The 2000 share of young and college-educated

services. Confidence intervals are too large for any valid inference, but the young and college-educated again have the largest (trip) or second largest (expenditure) positive change in these personal service categories. The young and college-educated also have the most positive change in expenditure to a much broader CEX "club" expenditure category that includes gyms, but the reverse is true for change in a broad categories of trips to play all sports.

individuals has a negative coefficient when entering alone, so stronger homophily does not in itself explain urban revival. However, the coefficient on 2000 restaurant density interacted with the 2000 young-college share is positive and significant in columns 2 and 4. This indicates a more positive change in preferences for restaurant density in locations with a large initial share of young college graduates. Notably, introducing these interactions reduces by more than half the coefficient on the uninteracted initial level of restaurant density.

The result that young college graduates' have a rising preference for non-tradable services especially in locations with large initial shares of their own type could reflect two specific phenomena that we explore further below. First, a strong relationship between homophily and non-tradable services may arise because these establishments are places for singles to mingle and meet potential partners. We explore the link between marital status and non-tradable services consumption later in this section. We also suspect that, in neighborhoods with lots of young college graduates, non-tradable service establishments have an unobserved quality that appeals specifically to young college graduates. The next section uses marketing studies to include such a quality measure in our regressions, which captures the desirability of specific establishments to young college graduates.

7.2 Changing Amenity Quality and Diversity

Food and restaurant quality and diversity increased fastest near city centers over the last decade (Figure 3). Therefore, quality and diversity improvements in dense areas could drive what we interpret as a change in taste for amenity density. To test this hypothesis empirically, we include quality and diversity indexes in our base specification. We describe these indexes briefly below and provide details in Appendix B.3.

Our quality index uses data from the Esri ArcGIS business analyst, which divides each neighborhood in the U.S. into market segments and assigns a Market Potential Index (MPI) to each chain in each segment. The MPI measures the propensity of each segment's inhabitants to visit a given chain relative to the average American. We use the MPIs in segments containing the largest share of young college graduates to identify the restaurant and food chains that they prefer.[46] We have MPIs for 24 food store chains and 61 restaurant chains, generally the largest family and fast-food restaurant chains. The three restaurants with the highest MPI for young-professionals are Starbucks (2.17), The Cheesecake Factory (2.11), and Chipotle (1.86), while the three lowest MPIs are Logan's Roadhouse (0.22), Church's Fried Chicken (0.33), and Bob Evans Farms (0.37). Our restaurant quality index is a weighted average of these MPI ratings near a tract, using the same transport costs weighting as for the amenity density indexes. To

[46]We average MPIs across all segments within which individuals are both more than 50 percent college-educated and more than 50 percent aged 18 to 44. Esri's MPIs come from the Survey of American Consumers, a proprietary dataset from GfK MRI.

alleviate concerns that changes in quality are driven by an influx of young college graduates who report visiting chains near where they live, we instrument quality changes by predicting entry and exit of chains as described in section 4.2.1.

Our diversity index is the inverse of a Herfindahl index, which decreases with the concentration of establishments within SIC8 codes near a tract (again using transport costs weights). For instance, diversity is lowest if every restaurant near a tract is a Korean restaurant, and highest if every restaurant belongs to a different SIC8 code. We have no instrument for diversity.

Regression results for the young and college-educated are in Table 8. We show, side by side, results with and without the additional quality and diversity variables. The coefficients on changes in restaurant quality and diversity are large and significant, while the coefficients on the quality and diversity of food stores are near zero. This confirms the primacy of non-tradable services like restaurants in attracting the young and college-educated. However, introducing these variables does not reduce the coefficient on initial restaurant level, which remains the main contributor to urban revival (shown in Panel D of Table 4).

Finally, Figure 6 shows that only the largest CBSAs have experienced a relative increase in quality and diversity near their city center, so these variables only contribute to urbanizing young college graduates in large cities. The ability of non-tradable service quality and diversity to explain why urban revival is primarily a large city phenomenon further establishes the role of non-tradable services in urban revival. To summarize, we find that improvements in non-tradable services attracts the young and college-educated into large cities' downtowns, but we do not find that these improvements explain the increasing taste of this group for non-tradable services.

7.3 Changing Family Structure and Income Distribution of the Young and College-Educated

We now assess the potential for recent trends in family formation and income distribution to explain urban revival. If richer and solo young college graduates spend more on and travel more to amenities like bars and restaurants, then delayed family formation and income growth may explain their rising preferences for non-tradable service amenities. In this section, we document significant differences in urban-vs.-suburban location choices across household and income types, as well as differences in the mix of consumption amenities that they consume and visit. We then investigate the potential for changes in household and income type composition to explain 1) the urban growth of young college graduates in general, and 2) the change in their preferences for non-tradable services highlighted in our model.

Our analysis combines IPUMS data on the distribution of five household types and four income brackets across space and over time (Figure 9) with NHTS and CEX data on the types most likely to spend on or travel to non-tradable service amenities (Figure 10). These three sources of microdata allow us to decompose the population into age-education-household types and age-education-income types not available in census tables, at the cost of smaller samples and, in the case of IPUMS, coarser geography.[47] The five household types are: 1. Solo, 2. Married couples with no children, 3. Households with oldest child younger than 5 years old, 4. Households with oldest child older than 5 years old, and 5. Others.[48] Household income is adjusted to reflect a "per capita" equivalent using the modified OECD equivalence scale.

All figures show data for the young and college-educated. Panel A of Figure 9 shows the spatial distribution of different household and income types in 2000, with the share of each type within urban areas in blue and the share of each type within suburban areas in green. Panel B shows time trends in the prevalence of each household and income types, with the 2000 share of each type in light orange and the 2007-2011 share of each type in dark orange. Figure 10 reports expenditure and travel shares to non-tradable services - restaurants in Panel A and bars in Panel B - by household and income types in 2000.[49]

Solo households stand out in both Figure 9 and 10. In 2000, a plurality of the young and college-educated live alone - more than for any other age-education group. Solos are by far the most urbanized household type, accounting for 60 percent of urban but only 30 percent of suburban young college graduates. Solos also have larger expenditure (CEX) and travel (NHTS) shares on bars and restaurants than any other household type. For instance, the solo's expenditure share on restaurants is twice as large as that of families with young children, and that on bars is four times as large. Considering the variation in location decisions of the young and college-educated by income instead of by household type, we observe more elaborate spatial sorting patterns, with both the poorest and the richest over-represented in urban areas. Non-tradable service amenities, however, are luxury goods, with richer young and college-educated households spending more on and traveling more to bars and restaurants. We note that poor and rich households possibly value downtowns for different reasons; for instance, non-tradable services draw in the rich while access to transit draws in the poor (Glaeser et al. 2008).

We now investigate whether changes in household or income composition from 2000 to 2007-2011 can explain the urbanization of young college graduates. Panel B of Figure 9 shows

[47] We construct urban areas in IPUMS as groups of Public Use Micro Areas (PUMAs) in each CBSA by sequentially adding the PUMAs closest to the city center until the total urban population reaches 10 percent of total CBSA population. PUMAs contain at least 100,000 individuals and they are the smallest geographical unit at which census and ACS microdata are available. We are able to create such downtowns for the 50 largest CBSAs.

[48] We define solos in IPUMS as individuals who do not live with anyone related to them by either blood, marriage or adoption. This category includes people who live alone, couples who live together and are unmarried, and people living with multiple non-related people, such as college dormitories. In our IPUMS sample, just over half of solo households live in a house with more than one person. In the CEX and NHTS, however, unrelated individuals living as roommates are probably not always reported as solos, and instead classified as "others" in this data. Indeed, Figure 10 shows that expenditures and trip shares for "others" are similar to that for "solo."

[49] Appendix C provides a similar figure for food stores, apparel stores and trips to buy goods.

that young college graduates are shifting from suburbanized household types, such as families with young children, towards urbanized solo households. A simple shift share analysis, therefore, suggests that recent changes in the distribution of household types are pushing the young and college-educated into urban areas. These changes predict 17 percent of the actual urban-suburban growth differential for young college graduates. For other age-education groups, however, these changes provide counterfactual predictions. A similar shift share analysis by income type also pushes young college graduates downtown, but this result is hard to interpret because of the Great Recession.

We now perform a similar analysis to explain changes in preferences for non-tradable service amenities. Figure 9 shows that, for young college graduates, household types with the highest propensity to spend on and travel to restaurants and bars ("solos" and "others") are growing, while types with the lowest propensity (families with children) are shrinking.[50] A shift share analysis therefore suggests that changes in the family structure of young college graduates increased their expenditures on and travel to both restaurants and bars. This increase in non-tradable service expenditure and trade shares is less than 5 percent of their initial 2000 levels.[51] A similar shift share analysis for income does not generate an increase in non-tradable service expenditures, because bars and restaurants are luxury goods and the recession reduces income measured in the post-period.[52]

Finally, one may worry that solo and richer households patronize non-tradable services only because they live close to such amenities. In this case, proximity to non-tradable services is a byproduct of these types' urban location choice. Using confidential geo-coded NHTS data, we find that the higher propensity of rich and solo households to travel to non-tradable services relative to other types of households persists almost entirely after controlling for amenity density near a traveler's residence.

To summarize, we find that the delayed family formation of young college graduates implies that they increasingly live in household types with high propensity for co-locating with, spending on, and traveling to non-tradable services. Simple decompositions show that although differences across household types in their propensity to travel to and spend on restaurants and bars are large, changes in family structure over the last decade are relatively small, which somewhat

[50] We cannot directly use the NHTS and CEX to measure changes in household and income types across surveys, because these samples are much too small in size and not stratified for this purpose. In these datasets, changes in household and income type have essentially zero impact on changes in NHTS and CEX trip and expenditure shares, meaning that the increase that we document in section 7.3 is not due to changes in household and income types.

[51] We compute this percentage change as $\sum(s_{n,10} - s_{n,00})x_{n,00})/x_{00}$, where $s_{n,10}$ is the share of households of type n in 2010 and $x_{n,00}$ is the expenditure (or travel) share for type n in 2000.

[52] A similar shift share analysis for race does not predict urban revival. Young college-educated Blacks, for instance, are overrepresented in urban areas in 2000 but have grown faster in suburbs, unlike Whites, Asians, Hispanics and others who grew faster in urban areas. Note that contrary to the perception that center cities are becoming more white, the share of urban young college graduates that are White has declined from 2000 to 2010, consistent with general demographic trends in the United States.

mutes the overall impact of these trends. Income growth for the young and college-educated could have a similar impact because richer young college graduates are over-represented in urban areas and devote a larger share of their travel and income to non-tradable services. These trends are harder to interpret given non-monotonic spatial sorting patterns by income and the impact of the Great Recession. We pursue this investigation in future work.

7.4 Changing Mobile Technology and Review Platforms

Recent innovations in mobile technology like mapping applications and establishment-rating aggregators may complement urban amenities and disproportionately benefit digitally savvy young college graduates. This hypothesis is hard to test directly. We look at the local share of NETS establishments that are independent, because they stand to benefit more than chains from maps and review portals.[53] However, we find no evidence that the share of independent restaurants affects the location choice of the young and college-educated. This is a coarse test of our hypothesis. A better test would exploit spatial variation in the timing of the introduction of key applications or platforms (e.g., Yelp), but such variation is hard to isolate.

8 Discussion

Urban revival currently attracts considerable media attention and interest from the general public. Using census data, we show that this revival is indeed happening in almost all large U.S. cities, and is driven by the location decisions of the young and college-educated. While the rest of the country continues to suburbanize, the young and college-educated flock downtown.

We evaluate the importance of various explanations for this trend. We find that diverging preferences for non-tradable services like restaurants and bars explain the diverging location decisions of the young and college-educated relative to other groups. Travel and expenditure shares of the young and college-educated also diverge from that of other groups, lending further credence to our model's results.

It is, of course, important to identify the source of such changing preference parameters. We have explored a few likely candidates. Rising restaurant quality and diversity drives young college graduates downtown in large cities, but we cannot establish that these improvements explain the changing preference that we estimate. In addition to change in the composition of amenities, changes in the composition of household type and income groups among the young and college-educated might explain their urbanization. For instance, the young and college-

[53] We define independent establishments in the NETS data as having fewer than five other establishments with the same name (see Appendix B.3 for details). The NPD Group, a marketing agency, reports 53.8 percent of independent restaurants in the spring of 2010. We find 49.6 percent with our methodology.

educated are increasingly likely to report living alone, and solos have much stronger demand for urban living and non-tradable service amenities, perhaps because of higher demand for networking, socializing and dating opportunities. Indeed we find that the rising preference of the young and college-educated for non-tradable services depends in large part on the presence of other young college graduates nearby (i.e., homophily). Other explanations, such as a complementarity between urban living and mobile technology that benefits digitally savvy young college graduates, are harder to test and remain speculative.

It is striking that the classic factors used to explain residential location decisions (e.g., jobs, housing, and schooling) struggle to explain urban revival. If the key factor at play is indeed a changing preference for urban non-tradable consumption amenities, then there are important consequences for the sustainability and welfare implications of urban revival. Consumption amenities are endogenous, and diverging preferences mean that while high-quality non-tradable services may compensate the young and college-educated for high housing prices near city centers, these amenities fail to compensate the poorer households already living there. These poorer households may either be displaced or incur high housing costs for downtowns offering fewer of the amenities that suit their less luxurious tastes. We are exploring these welfare implications in complimentary work.

References

Albouy, D. and B. Lue (2015). Driving to opportunity: Local rents, wages, commuting, and sub-metropolitan quality of life. *Journal of Urban Economics 89*, 74–92.

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

Autor, D. H., C. J. Palmer, and P. A. Pathak (2017). Gentrification and the Amenity Value of Crime Reductions: Evidence from Rent Deregulation. Working Paper 23914, National Bureau of Economic Research.

Bartik, T. J. (1991). Who benefits from state and local economic development policies? *Upjohn Press*.

Baum-Snow, N. (2007). Did Highways Cause Suburbanization? *The Quarterly Journal of Economics 122*(2), 775–805.

Baum-Snow, N. and D. Hartley (2016). Causes and Consequences of Central Neighborhood Change, 1970-2010.

Bayer, P., F. Ferreira, and R. McMillan (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy 115*(4), 588–638.

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal 75*(299), 493–517.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841–890.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics 25*(2), 242–262.

Boustan, L. P. (2010). Was Postwar Suburbanization White Flight? Evidence from the Black Migration. *The Quarterly Journal of Economics 125*(1), 417–443.

Card, D. and A. B. Krueger (1992). School Quality and Black-White Relative Earnings: A Direct Assessment. *The Quarterly Journal of Economics 107*(1), 151–200.

Carlino, G. A. and A. Saiz (2008). Beautiful City: Leisure Amenities and Urban Growth. SSRN Scholarly Paper ID 1280157, Social Science Research Network, Rochester, NY.

Cortright, J. (2014). The Young and the Restless and the Nation's Cities. Technical report, City Observatory.

Couture, V. (2013). Valuing the Consumption Benefits of Urban Density. *University of California, Berkeley. Processed*.

Davidoff, T. (2016). Supply Constraints Are Not Valid Instrumental Variables for Home Prices Because They Are Correlated with Many Demand Factors. *Critical Finance Review 5*(2), 177–206.

Diamond, R. (2016). The determinants and welfare implications of U.S. workers' diverging location choices by skill: 1980-2000. *American Economic Review 106*(3), 479–524.

Duranton, G. and D. Puga (2015). Chapter 8 - Urban Land Use. In J. V. H. a. W. C. S. Gilles Duranton (Ed.), *Handbook of Regional and Urban Economics*, Volume 5 of *Handbook of Regional and Urban Economics*, pp. 467–560. Elsevier.

Edlund, L., C. Machado, and M. Sviatchi (2015). Bright Minds, Big Rent: Gentrification and the Rising Returns to Skill. Working Paper 21729, National Bureau of Economic Research.

Ellen, I. G., K. M. Horn, and D. Reed (2017). Has Falling Crime Invited Gentrification? Technical Report CES 17-27, Center for Economic Studies.

Epple, D. and H. Sieg (1999). Estimating Equilibrium Models of Local Jurisdictions. *Journal of Political Economy 107*(4), 645–681.

Ferreira, F. and J. Gyourko (2011). Anatomy of the beginning of the housing boom: US neighborhoods and metropolitan areas, 1993-2009. Working Paper 17374, National Bureau of Economic Research.

Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition Methods in Economics. In *Handbook of Labor Economics*, Volume 4. Elsevier B.V.

Frey, W. H. (1993). The new urban revival in the United States. *Urban Studies 30*, 741–741.

Glaeser, E. L., J. Gyourko, and R. E. Saks (2006). Urban growth and housing supply. *Journal of Economic Geography 6*(1), 71–89.

Glaeser, E. L., M. E. Kahn, and J. Rappaport (2008). Why do the poor live in cities? The role of public transportation. *Journal of Urban Economics 63*(1), 1–24.

Glaeser, E. L., J. Kolko, and A. Saiz (2001). Consumer city. *Journal of Economic Geography 1*(1), 27–50.

Glaeser, E. L., A. Saiz, G. Burtless, and W. C. Strange (2004). The Rise of the Skilled City [with Comments]. *Brookings-Wharton Papers on Urban Affairs*, 47–105.

Graham, M. R., M. J. Kutzbach, and B. McKenzie (2014). Design Comparison of LODES and ACS Commuting Data. Technical report, United States Census Bureau, Center for Economic Studies.

Gyourko, J., C. Mayer, and T. Sinai (2013). Superstar cities. *American Economic Journal: Economic Policy 5*(4), 167–199.

Holian, M. J. and M. E. Kahn (2012). The Impact of Center City Economic and Cultural Vibrancy on Greenhouse Gas Emissions from Transportation. MTI Report 11-13, Mineta Transportation Institute.

Igami, M. and N. Yang (2016). Unobserved Heterogeneity in Dynamic Games: Cannibalization and Preemptive Entry of Hamburger Chains in Canada. *Quantitative Economics 7*(2), 483–521.

Jaffee, D. and J. M. Quigley (2011). The Future of the Government Sponsored Enterprises: The Role for Government in the U.S. Mortgage Market. Working Paper 17685, National Bureau of Economic Research.

Johnson, K. and N. Wingfield (2013). As Amazon Stretches, Seattle's Downtown Is Reshaped. *The New York Times*.

Kneebone, E. and S. Raphael (2011). City and Suburban Crime Trends in Metropolitan America. Technical report, Brookings Institution.

Kotkin, J. and W. Cox (2011). Cities and the Census. *City Journal*.

Lee, S. and J. Lin. (Forthcoming). Natural Amenities, Neighborhood Dynamics, and Persistence in the Spatial Distribution of Income. *The Review of Economic Studies*.

LeRoy, S. F. and J. Sonstelie (1983). Paradise lost and regained: Transportation innovation, income, and residential location. *Journal of Urban Economics 13*(1), 67–89.

Levitt, S. D. (2004). Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not. *Journal of Economic Perspectives 18*(1), 163–190.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.

McFadden, D. (1978). Modelling the Choice of Residential Location. In A. Karlqvist (Ed.), *Spatial Interaction Theory and Planning Models*. New York: Elsevier North-Holland.

Mills, E. (1969). The Value of Land. *The Quality of the Urban Environment*, 231–253.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2015). Commuting, Migration and Local Employment Elasticities. Working Paper 21706, National Bureau of Economic Research.

Moretti, E. (2012). *The New Geography of Jobs*. Houghton Mifflin Harcourt.

Murphy, D. (2017). Home Production, Expenditure, and Economics Geography. *University of Virginia Darden School of Business*.

Muth, R. (1969). Cities and Housing: The Spatial Patterns of Urban Residential Land Use. *University of Chicago, Chicago 4*, 114–123.

Neumark, D., J. Zhang, and B. Wall (2007). Employment dynamics and business relocation: New evidence from the National Establishment Time Series. *Research in Labor Economics 26*, 39–83.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review 14*(3), 693–709.

Parrott, J. and M. Zandi (2013). Opening the Credit Box. Technical report, Moodys Analytics and the Urban Institute, New York and Washington, DC.

Rappaport, J. (2009). The increasing importance of quality of life. *Journal of Economic Geography 9*(6), 779–804.

Rappaport, J. (2015). Millennials, Baby Boomers, and Rebounding Multifamily Home Construction. Working Paper 2637622, Federal Reserve Bank of Kansas City.

Roback, J. (1982). Wages, Rents, and the Quality of Life. *Journal of Political Economy 90*(6), 1257–78.

Rosen, S. (1979). Wage-Based Indexes of Urban Quality of Life. In P. M. Mieszkowski and M. R. Straszheim (Eds.), *Current Issues in Urban Economics*. John Hopkins University Press.

Sanderson, E. and F. Windmeijer (2016). A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables. *Journal of Econometrics 190*, 212–221.

Small, K. and E. Verhoef (2007). *The Economics of Urban Transportation* (2 edition ed.). New York: Routledge.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear IV regression. pp. 80–108. New York: Cambridge University Press.

Stock, K. and V. Wong (2015). *Chipotle: The Definitive Oral History*.

Teulings, C. N., I. V. Ossokina, and H. L. F. d. Groot (2014). Welfare Benefits of Agglomoration and Worker Heterogeneity. Technical Report 289, CPB Netherlands Bureau for Economic Policy Analysis.

Toivanen, O. and M. Waterson (2005). Market Structure and Entry: Where's the Beef? *The RAND Journal of Economics 36*(3), 680–699.

Waddell, P., C. Bhat, N. Eluru, L. Wang, and R. M. Pendyala (2007). Modeling interdependence in household residence and workplace choices. *Transportation Research Record: Journal of the Transportation Research Board 2003*(1), 84–92.

Tables and Figures

Table 1: Nested-Logit Residential Location Choice Regression Results (OLS)

Panel A: College Educated

	25-34 Year-Olds		35-44 Year-Olds		45-65 Year-Olds	
	Change	Level	Change	Level	Change	Level
Variable	[1]	[2]	[3]	[4]	[5]	[6]
Job Density (Low Wage)	-0.070***	-0.047***	-0.084***	-0.189***	-0.022***	-0.160***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Job Density (Med. Wage)	0.049***	0.110***	0.120***	0.259***	0.077***	0.258***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Job Opportunities (High Wage)	0.033***	-0.054***	-0.015***	-0.045***	-0.015***	-0.059***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
House Price Index	0.062***	-0.001	0.002*	0.028***	0.000	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Restaurant Density	0.012***	0.012***	0.008***	-0.001	0.006***	-0.007***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Food Store Density	0.007***	-0.008***	0.013***	-0.001	0.015***	-0.016***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Population Density		-0.007***		-0.012***		-0.031***
		(0.002)		(0.002)		(0.002)
Share of Same Type		-0.036***		-0.031***		-0.043***
		(0.001)		(0.002)		(0.002)
Within-CBSA Share	0.887***		0.888***		0.853***	
	(0.005)		(0.005)		(0.006)	
Observations	33,941		33,892		34,700	

Panel B: Non-College Educated

	25-34 Year-Olds		35-44 Year-Olds		45-65 Year-Olds	
	Change	Level	Change	Level	Change	Level
Variable	[1]	[2]	[3]	[4]	[5]	[6]
Job Density (Low Wage)	-0.141***	-0.135***	-0.210***	-0.221***	-0.126***	-0.090***
	(0.007)	(0.008)	(0.004)	(0.004)	(0.003)	(0.003)
Job Density (Med. Wage)	0.070***	0.173***	0.322***	0.401***	0.171***	0.205***
	(0.009)	(0.009)	(0.005)	(0.005)	(0.004)	(0.004)
Job Opportunities (High Wage)	0.015**	-0.129***	-0.088***	-0.132***	-0.054***	-0.115***
	(0.006)	(0.005)	(0.003)	(0.003)	(0.003)	(0.002)
House Price Index	0.003	0.000	-0.002*	0.040***	0.001	0.011***
	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Restaurant Density	0.014***	-0.008	0.007***	-0.040***	0.008***	-0.020***
	(0.003)	(0.006)	(0.001)	(0.003)	(0.001)	(0.003)
Food Store Density	0.044***	0.034***	0.033***	0.045***	0.022***	0.004
	(0.003)	(0.007)	(0.001)	(0.003)	(0.001)	(0.003)
Population Density		-0.081***		-0.017***		-0.030***
		(0.006)		(0.003)		(0.003)
Share of Same Type		-0.006**		0.020***		-0.045***
		(0.003)		(0.002)		(0.002)
Within-CBSA Share	0.537***		0.820***		0.854***	
	(0.017)		(0.006)		(0.004)	
Observations	35,030		35,084		35,177	

Notes: * – 10%; ** – 5%; ***–1%. The change in the share of type d individuals within CBSA c who live in tract j is instrumented. Each regression is weighted by the share of type d in tract j in year 2000.

Table 2: Nested-Logit Residential Location Choice Regression Results (IV)

Panel A: College-Educated

Variable	25-34 Year-Olds Change [1]	Level [2]	35-44 Year-Olds Change [3]	Level [4]	45-65 Year-Olds Change [5]	Level [6]
Job Density (Low Wage)	-0.114***	-0.035***	-0.166***	-0.171***	-0.038***	-0.130***
	(0.015)	(0.008)	(0.013)	(0.007)	(0.014)	(0.008)
Job Density (Med. Wage)	-0.090***	0.036***	0.162***	0.236***	0.103***	0.212***
	(0.017)	(0.011)	(0.014)	(0.009)	(0.016)	(0.011)
Job Opportunities (High Wage)	0.212***	0.012	0.103***	0.042***	0.039***	0.019***
	(0.017)	(0.008)	(0.013)	(0.006)	(0.015)	(0.007)
House Price Index	0.019***	-0.009***	-0.039***	0.020***	-0.018***	0.012***
	(0.006)	(0.003)	(0.005)	(0.003)	(0.005)	(0.003)
Restaurant Density	0.385***	0.394***	0.231***	0.232***	0.249***	0.215***
	(0.027)	(0.029)	(0.018)	(0.020)	(0.022)	(0.023)
Food Store Density	-0.047***	-0.215***	0.003	-0.117***	0.082***	-0.019
	(0.009)	(0.021)	(0.007)	(0.015)	(0.009)	(0.018)
Population Density		-0.049***		-0.098***		-0.274***
		(0.011)		(0.010)		(0.014)
Share of Same Type		-0.099***		-0.074***		-0.087***
		(0.007)		(0.005)		(0.005)
Within-CBSA Share	0.753***		0.754***		0.655***	
	(0.014)		(0.013)		(0.016)	
Observations	33,941		33,892		34,700	

Panel B: Non-College Educated

Variable	25-34 Year-Olds Change [1]	Level [2]	35-44 Year-Olds Change [3]	Level [4]	45-65 Year-Olds Change [5]	Level [6]
Job Density (Low Wage)	-0.317***	-0.111***	-0.456***	-0.303***	-0.318***	-0.202***
	(0.024)	(0.017)	(0.016)	(0.011)	(0.011)	(0.008)
Job Density (Med. Wage)	-0.202***	-0.064***	0.539***	0.419***	0.328***	0.201***
	(0.030)	(0.021)	(0.020)	(0.014)	(0.013)	(0.010)
Job Opportunities (High Wage)	0.362***	-0.010	-0.002	0.012	-0.091***	-0.037***
	(0.027)	(0.014)	(0.019)	(0.009)	(0.014)	(0.007)
House Price Index	-0.055***	0.014**	-0.007	0.012***	0.030***	-0.002
	(0.009)	(0.006)	(0.006)	(0.004)	(0.005)	(0.003)
Restaurant Density	0.288***	0.245***	0.349***	0.298***	0.241***	0.195***
	(0.049)	(0.045)	(0.032)	(0.032)	(0.023)	(0.023)
Food Store Density	0.081***	0.035	0.012	-0.090***	0.050***	-0.013
	(0.018)	(0.036)	(0.012)	(0.025)	(0.009)	(0.018)
Population Density		-0.337***		-0.184***		-0.202***
		(0.020)		(0.013)		(0.010)
Share of Same Type		0.104***		0.121***		-0.028***
		(0.010)		(0.008)		(0.006)
Within-CBSA Share	0.261***		0.641***		0.774***	
	(0.038)		(0.019)		(0.011)	
Observations	35,030		35,084		35,177	

Notes: * – 10%; ** – 5%; ***–1%. The change in house prices, consumption amenity density, job density, within-CBSA share, and the levels of local demographic shares and house prices are instrumented. Each regression is weighted by the share of type d in a tract j in year 2000.

Table 3: Nested-Logit Residential Location Choice Regression Results Including More Amenities and Distance to City Center for the Young and College-Educated (OLS and IV)

Panel A: OLS Regression Results

	4 amenities		9 amenities		With Dist. to City Center	
	Change	Level	Change	Level	Change	Level
Variable	[1]	[2]	[3]	[4]	[5]	[6]
Job Density (Low Wage)	-0.069***	-0.046***	-0.069***	-0.045***	-0.067***	-0.043***
Job Density (Med. Wage)	0.047***	0.107***	0.045***	0.106***	0.043***	0.102***
Job Opportunities (High Wage)	0.035***	-0.052***	0.039***	-0.050***	0.038***	-0.049***
House Price Index	0.062***	-0.001	0.060***	-0.004***	0.061***	-0.003***
Museum and Library Density			-0.002***	-0.008***	-0.002***	-0.009***
Parks and Golf Density			-0.001**	0.003***	-0.001*	0.003***
Gym and Sports Density			0.011***	0.018***	0.011***	0.019***
Restaurant Density	0.010***	0.012***	0.006***	0.006***	0.007***	0.007***
Bar Density	0.002**	0.002*	0.001*	0.004***	0.001*	0.004***
Personal Services Density			0.011***	0.003	0.012***	0.004
Merchandise Stores			-0.004***	-0.013***	-0.004***	-0.013***
Food Store Density	0.005***	-0.008***	0.001	-0.007**	0.002	-0.007**
Apparel Store Density	0.008***	0.000	0.005***	0.002	0.005***	0.001
Distance to City Center						-0.005***
Population Density		-0.009***		-0.004**		-0.007***
Share of Same Type		-0.037***		-0.038***		-0.040***
Within-CBSA Share	0.887***		0.907***		0.897***	
Observations	33,941		33,941		33,941	

Panel B: IV Regression Results

	4 amenities		9 amenities		With Dist. to City Center	
	Change	Level	Change	Level	Change	Level
Variable	[1]	[2]	[3]	[4]	[5]	[6]
Job Density (Low Wage)	-0.138***	-0.034***	-0.131***	-0.037***	-0.120***	-0.031***
Job Density (Med. Wage)	-0.057***	0.036***	-0.012	0.053***	-0.001	0.047***
Job Opportunities (High Wage)	0.211***	0.023***	0.159***	0.011**	0.139***	0.013**
House Price Index	0.020***	-0.011***	0.041***	-0.014***	0.049***	-0.006**
Museums and Libraries			-0.018***	-0.020***	-0.034***	-0.036***
Parks and Golf Density			-0.014***	-0.006**	-0.010*	-0.003
Gym and Sports Density			0.025**	0.029**	0.030***	0.037**
Restaurant Density	0.323***	0.329***	0.133***	0.129***	0.144***	0.138***
Bar Density	0.035***	0.018*	0.046***	0.046***	0.042***	0.039***
Personal Services Density			0.080***	0.062***	0.095***	0.083***
Merchandise Stores			-0.026***	-0.025***	-0.018**	-0.013
Food Store Density	-0.049***	-0.198***	-0.057***	-0.145***	-0.045***	-0.128***
Apparel Store Density	0.005	-0.002	0.002	0.012	-0.026**	-0.022
Distance to City Center						-0.024***
Population Density		-0.041***		-0.011		-0.029***
Share of Same Type		-0.095***		-0.074***		-0.086***
Within-CBSA Share	0.760***		0.810***		0.763***	
Observations	33,941		33,941		33,941	

Notes: * – 10%; ** – 5%; ***–1%. In Panel A the change in the share of type d individuals within CBSA c who live in tract j is instrumented. In Panel B the change in house prices, consumption amenity density, job density, within-CBSA share, and the levels of local demographic shares and house prices are instrumented. Each regression is weighted by the share of type d in tract j in year 2000.

Table 4: Share of Non-Tradable Services' Urbanizing Contribution Across Specifications for the Young and College-Educated

Specification	Rank	Share of Contribution to Urbanization
Panel A: Basic Set of Controls with 9 Amenities		
Base IV Specification	1 (+)	83%
Base OLS Specification	1 (+)	50%
Non-nested IV with CBSA Fixed Effects	1 (+)	75%
Panel B: Base IV Specification with Additional Controls		
Distance to City Center Control	1 (+)	73%
School, Crime and Transit Controls	1 (+)	82%
Homophily Control	1 (+)	49%
Panel C: Base IV Specification with Alternative Housing Index		
Ferreira/Gyourko Hedonic Index	1 (+)	72%
Zillow 2 Bedroom Index	1 (+)	83%
Housing on the Left Hand Side	1 (+)	78%
HUD Fair Market Rents	1 (+)	77%
Housing Age	1 (+)	30%
Panel D: Base IV Specifications with 2 Amenities (Food Store and Restaurant)		
Without Amenity Quality and Diversity	1 (+)	88%
With Amenity Quality and Diversity	1 (+)	65%

Table 5: Nested-Logit Residential Location Choice Regression Results Including School, Crime and Transit (IV)

	School Quality		Violent Crime Rate		Transit Time	Regression
	Change	Level	Change	Level	Level (2014)	Obs.
Demographic Group	[1]	[2]	[3]	[4]	[5]	[6]
College-Educated:						
25-34 Year-Olds	0.001	-0.004	0.002	0.034***	0.035***	10,452
35-44 Year-Olds	0.012***	-0.004*	-0.011***	0.023***	0.005*	10,353
45-65 Year-Olds	0.023***	0.036***	-0.007**	0.022***	-0.013**	10,648
Non-College Educated:						
25-34 Year-Olds	-0.002	-0.014	-0.023***	0.012	0.061***	10,807
35-44 Year-Olds	0.003	-0.038***	-0.032***	0.017**	0.052***	10,812
45-65 Year-Olds	0.005	-0.033***	-0.018***	-0.005	0.050***	10,833

Notes: * – 10% significance level; ** – 5% significance level; ***–1% significance level. Regressions also include all variables in the base specification, as outlined in Table 2.

Table 6: Nested-Logit Residential Location Choice Regression Results Including Homophily Control (IV)

Variable	Base Specification Change [1]	Base Specification Level [2]	Homophily Controls Change [3]	Homophily Controls Level [4]
Job Density (Low Wage)	-0.114***	-0.035***	-0.135***	-0.013**
Job Density (Med. Wage)	-0.090***	0.036***	-0.050***	0.006
Job Opportunities (High Wage)	0.212***	0.012	0.170***	-0.004
House Price Index	0.019***	-0.009***	0.005	-0.006***
Restaurant Density	0.385***	0.394***	0.060***	0.056***
Food Store Density	-0.047***	-0.215***	-0.005	-0.037***
Population Density		-0.049***		
Share of Same Type		-0.099***		
Nearby Pop Density			0.066***	0.000
Nearby Same Type Share			0.067***	-0.005**
Within-CBSA Share	0.753***		0.728***	
Observations	33,941		33,941	

Notes: * – 10%; ** – 5%; ***–1%. The change in house prices, consumption amenity density, job density, within-CBSA share, and the levels of local demographic shares and house prices are instrumented. Each regression is weighted by the share of type d in a CBSA c in year 2000.

Table 7: Nested-Logit Residential Location Choice Regression Results Homophily Interaction

Variable	IV Change [1]	IV Level [2]	OLS Change [3]	OLS Level [4]
Job Density (Low Wage)	-0.109***	-0.026**	-0.068***	-0.047***
Job Density (Med. Wage)	-0.133***	0.006	0.047***	0.108***
Job Opportunities (High Wage)	0.242***	0.022**	0.033***	-0.053***
House Price Index	0.009	-0.001	0.062***	-0.001
Restaurant Density	0.347***	0.183***	0.010***	0.005**
Food Store Density	-0.099***	-0.039	0.009***	-0.003
I.Restaurants	-0.057	0.493***	0.006***	0.024***
I.Food Stores	0.168***	-0.354***	-0.003***	-0.015***
Population Density		-0.082***		-0.010***
Share of Same Type		-0.199***		-0.043***
Within-CBSA Share	0.681***		0.877***	
Observations	33,941		33,941	

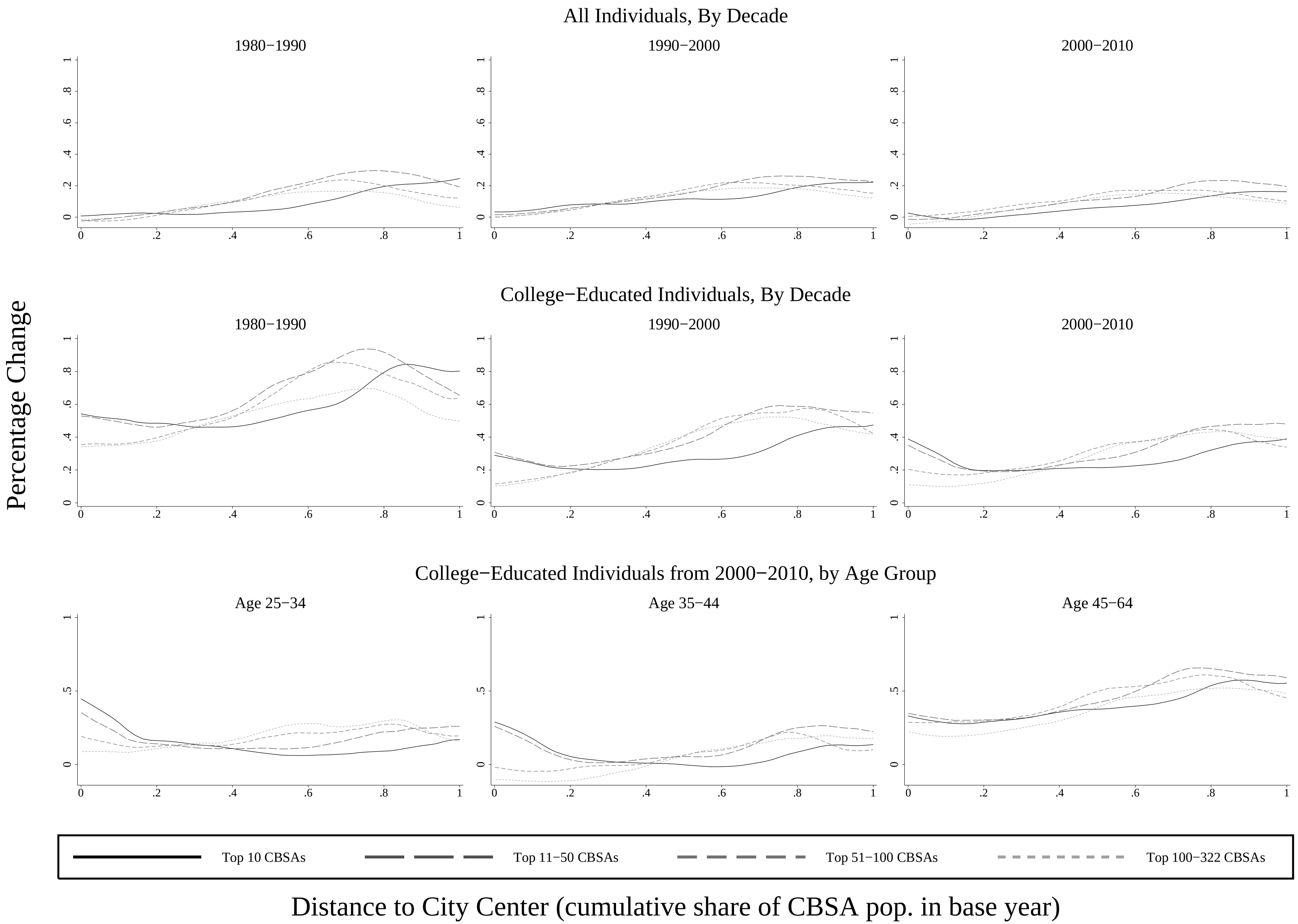
Notes: * – 10%; ** – 5%; ***–1%. In OLS the change in the share of type d individuals within CBSA c who live in tract j is instrumented. In IV the change in house prices, consumption amenity density, consumption amenity density interaction, job density, within-CBSA share, and the levels of local demographic shares, consumption amenity density interaction, and house prices are instrumented. Each regression is weighted by the share of type d in tract j in year 2000.

Table 8: Nested-Logit Residential Location Choice Regression Results
Including Amenity Quality and Diversity (IV)

	Quality and Diversity		Same Sample without Quality and Diversity	
	Change	Level	Change	Level
Variable	[1]	[2]	[3]	[4]
Job Density (Low Wage)	-0.094***	-0.029***	-0.137***	-0.041***
Job Density (Med. Wage)	-0.075***	0.048***	-0.050***	0.062***
Job Opportunities (High Wage)	0.203***	0.006	0.210***	0.009
House Price Index	0.002	-0.025***	0.027***	-0.013***
Restaurant Density	0.380***	0.280***	0.272***	0.215***
Food Store Density	-0.044***	-0.159***	-0.051***	-0.148***
Restaurant Quality	0.053***	0.018***		
Food Stores Quality	0.009	0.009		
Restaurants Diversity	0.174***	0.055***		
Food Stores Diversity	0.000	-0.023		
Population Density		-0.031***		-0.006
Share of Same Type		-0.111***		-0.109***
Within-CBSA Share	0.799***		0.756***	
Observations	21,365		21,365	

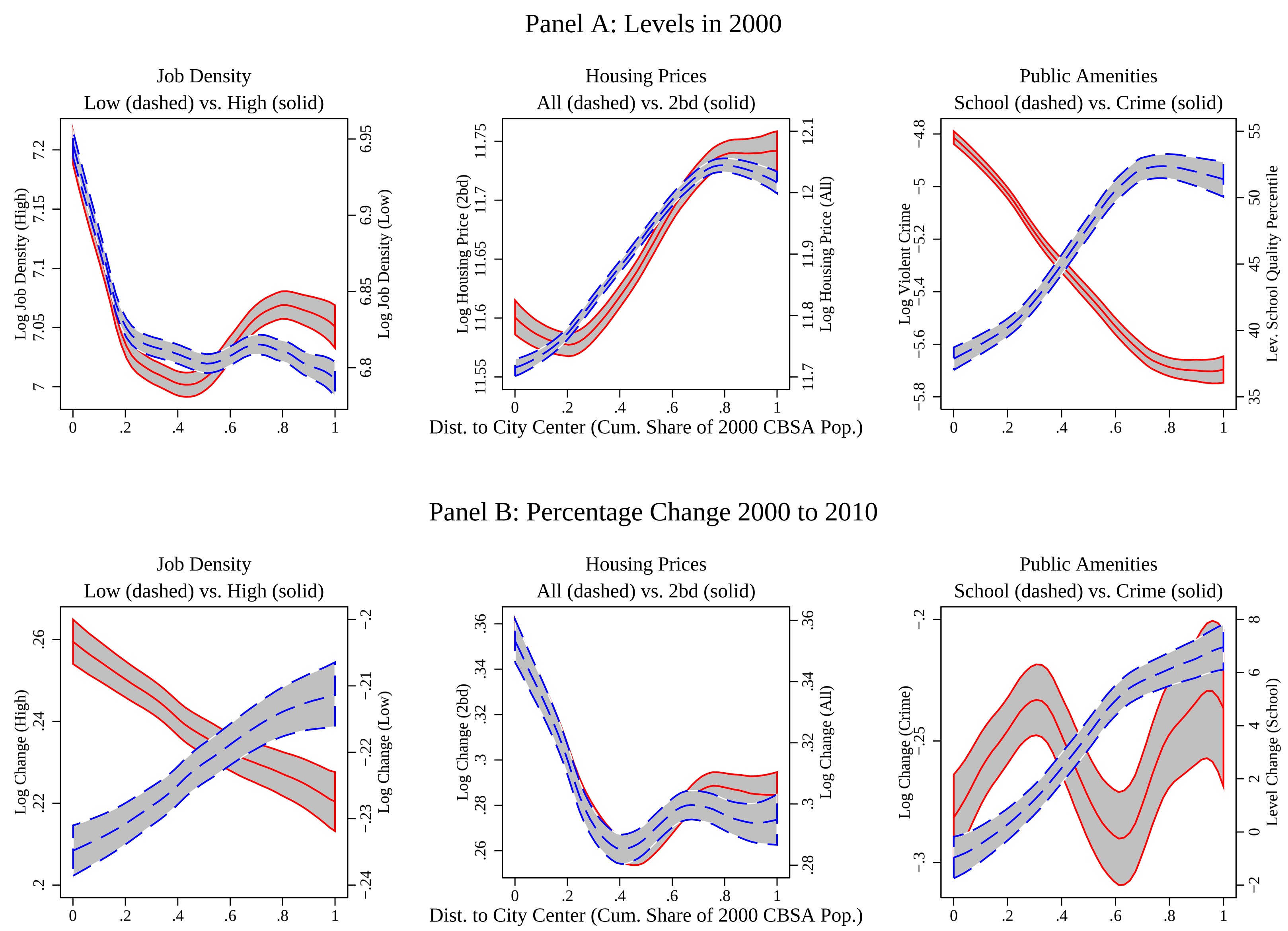
Notes: * – 10%; ** – 5%; ***–1%. The change in house prices, consumption amenity density and quality, job density, within-CBSA share, and the levels of local demographic shares and house prices are instrumented. Each regression is weighted by the share of type d in a CBSA c in year 2000.

Figure 1: Population Growth at Various Distance from the City Center by CBSA Size



Notes: Data from decennial censuses 1980–2000 and ACS 2008–2012. Each figure shows a non-parametric kernel fit of percent change in tract population at different distances from the city center. Distance from the city center is measured as the cumulative share of CBSA population in the base year. The top row of plots presents kernels for the percent change in total population between 1980 and 1990, 1990 and 2000, and 2000 to 2010. The second row is similar but for the college-educated population. The final row presents kernels for the 2000-2010 percent change in the college-educated population in three age brackets. The four lines show the percent change for the top 1-10 CBSAs, top 11-50 CBSAs, top 51-100 CBSAs and all remaining CBSAs, as ranked by CBSA population in the pre-period of each plot.

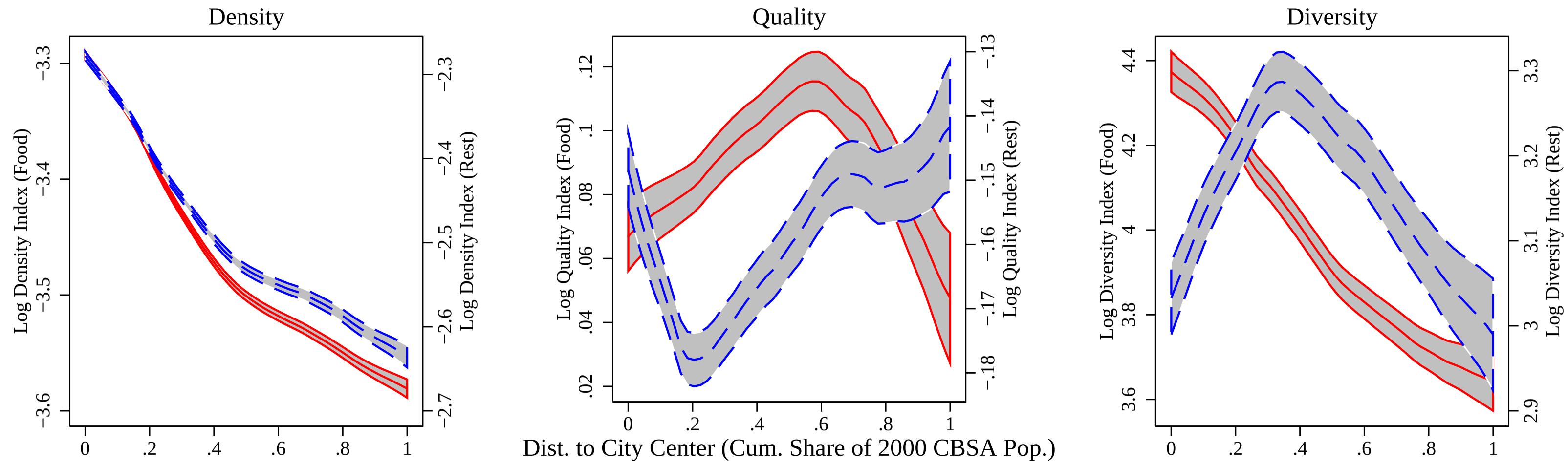
Figure 2: Initial Level and Change in Job Density, Housing Prices, and Public Amenities at Various Distance from the City Center



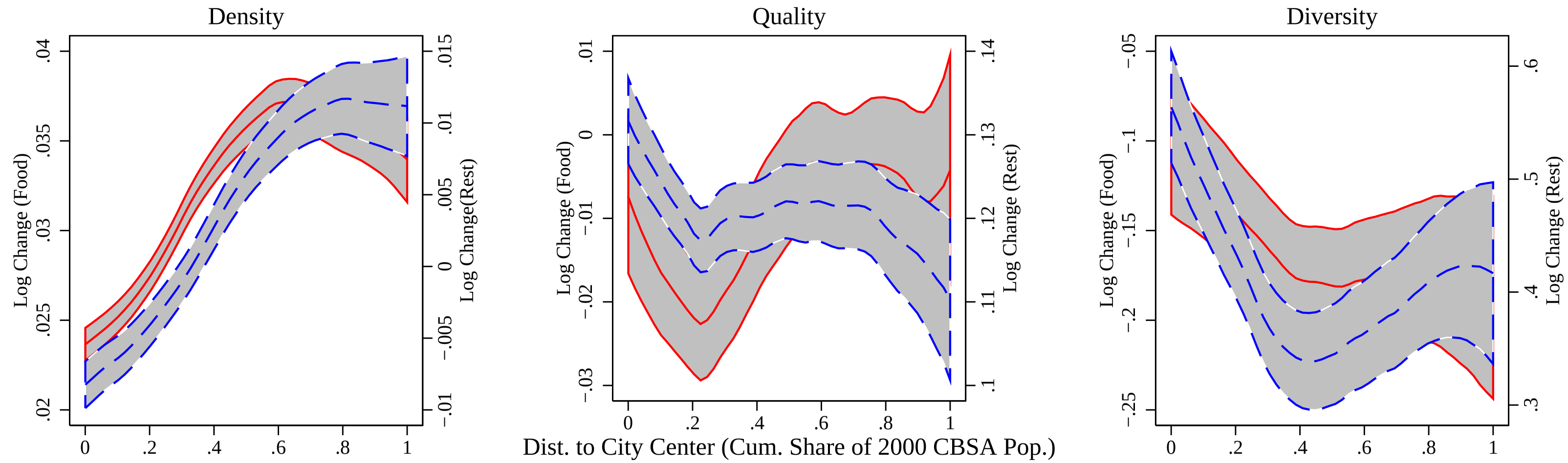
Notes: All confidence intervals are 95% intervals. Panel A shows a non-parametric kernel fit of the log of job density, log house price, log per capita violent crime and school ranking in 2000 plotted against the young-college population-weighted distance from the city center for all tracts in our estimation sample. Panel B shows a similar kernel of percent changes from 2000 to 2010. The job data is from LODES in 2002 and 2011, the house price data is from Zillow.com, the crime data is from UCR, and school data is from SchoolDigger.com (see Appendix A for details on data sources).

Figure 3: Initial Level and Change in Restaurant and Food Store Density, Quality, and Diversity Indexes at Various Distance from the City Center

Panel A: Levels in Restaurants (dashed) and Food Stores (solid) in 2000

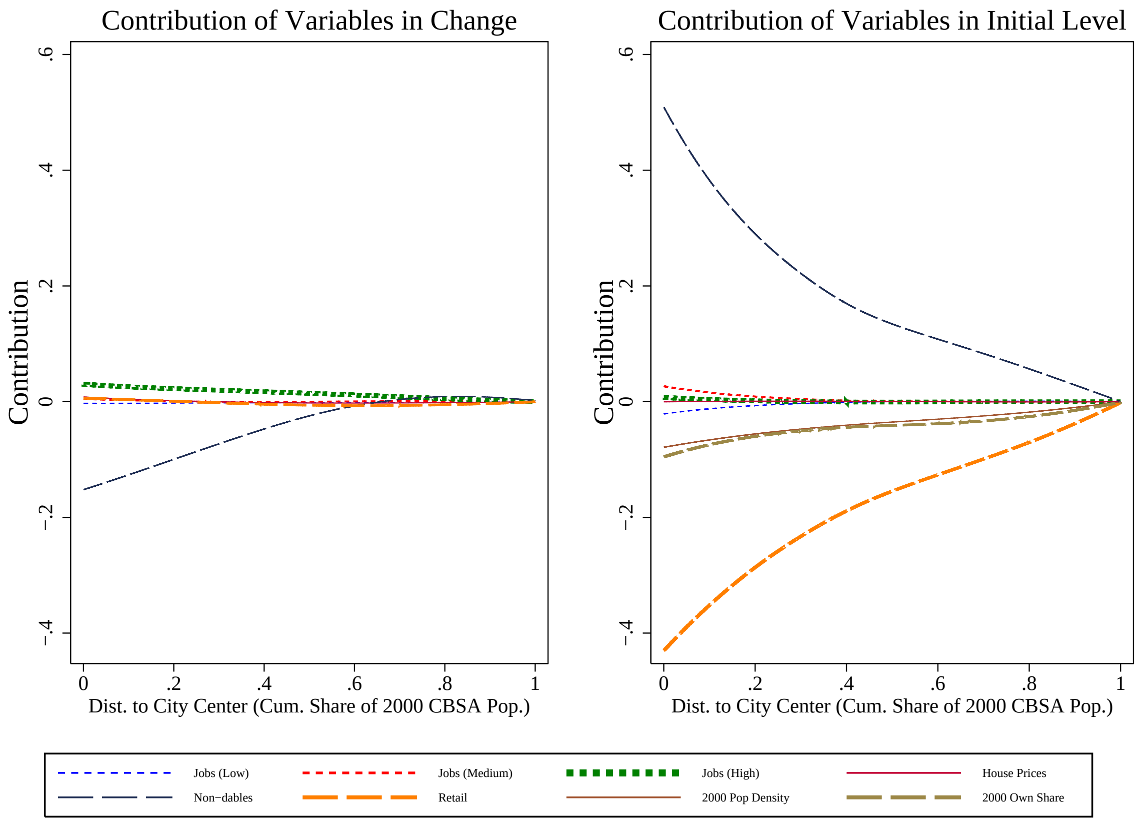


Panel B: Percentage Change in Restaurants (dashed) and Food Stores (solid) 2000 to 2010



Notes: All confidence intervals are 95% intervals. Panel A shows a non-parametric kernel fit of the log of amenity density, quality, and diversity indexes in 2000 for restaurants and food stores, plotted against the young-college population-weighted distance from the city center for all tracts in our estimation sample. Panel B shows a kernel of the 2000 to 2010 percent change in these indexes . See Appendix B for details on consumption amenity index construction.

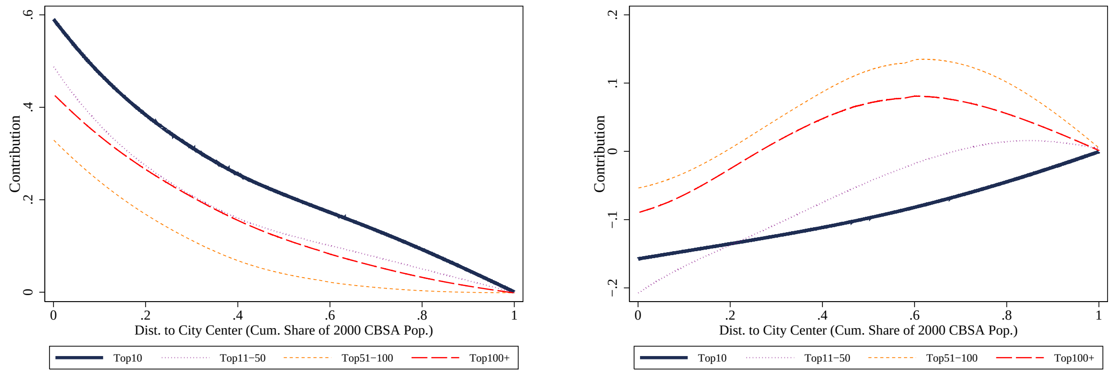
Figure 4: Variables Contributing to Young College-Educated Growth at Various Distance from the City Center



Notes: We use coefficients from the base IV specification in Table 2. See description in Section 4.3.

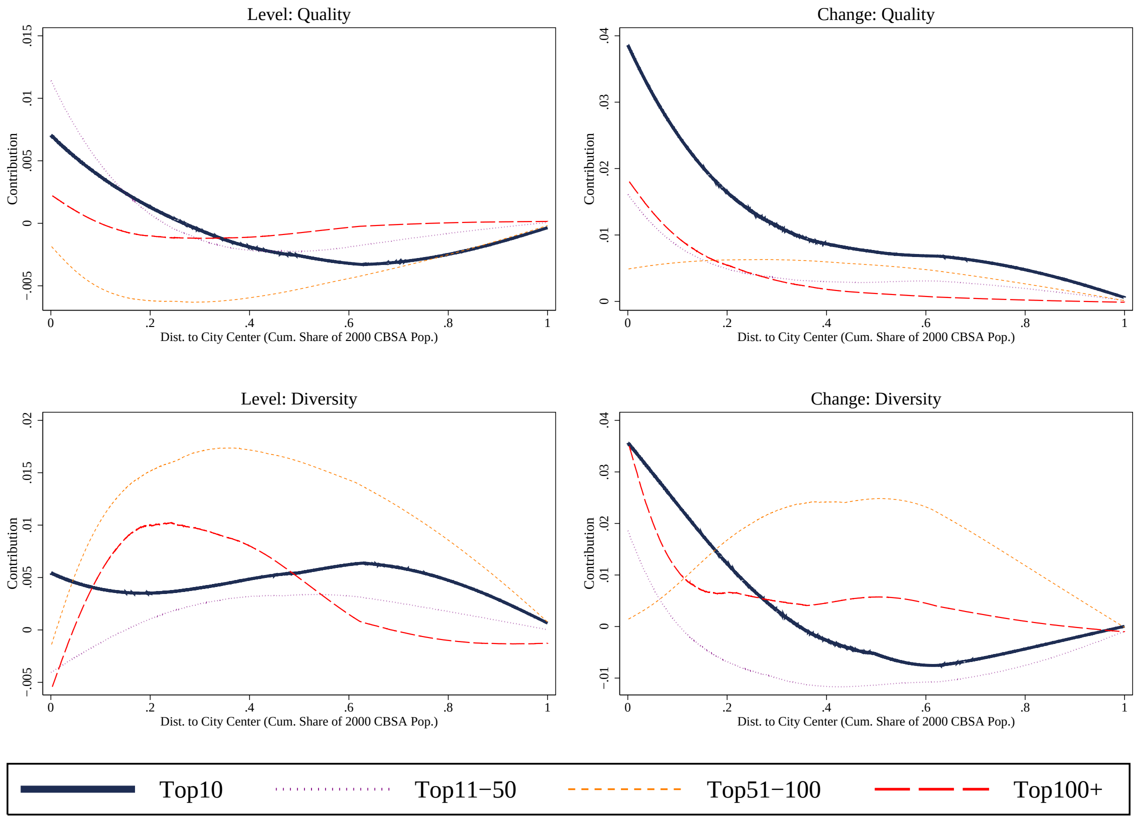
Figure 5: Contribution of Restaurant Density to Young College-Educated Growth at Various Distance from the City Center by CBSA Size

(a) Level (b) Change



Notes: We use coefficients from the base IV specification in Table 2. See description in Section 4.3.

Figure 6: Contribution of Restaurant Quality and Diversity to Young College-Educated Growth at Various Distance from the City Center by CBSA Size



Notes: We use coefficients from the IV specification including amenity quality and diversity indexes, shown in Table 8. See description in Section 4.3.

Figure 7: Change in Commute Patterns

Panel A: All Workers in All CBSAs

Residence-City Center Dist. \ Workplace-City Center Distance (miles)	[0, 1)	[1,2)	[2,4)	[4, 8)	[8, 16)	[16, 32)	32+
[0, 1)	-15.02	-15.02	-12.92	-3.84	6.21	9.02	14.76
[1,2)	-14.08	-12.85	-14.67	-7.36	3.12	6.88	7.22
[2,4)	-11.56	-9.70	-10.68	-6.38	0.67	4.14	10.25
[4, 8)	-2.81	0.17	-3.39	-3.93	1.46	6.16	6.15
[8, 16)	8.88	13.82	8.60	8.00	2.54	10.27	14.76
[16, 32)	20.75	27.81	22.38	22.62	16.36	3.33	15.59
32+	32.28	41.13	33.81	37.91	40.84	31.67	10.69

Panel B: High-Wage Workers in All CBSAs

Residence-City Center Dist. \ Workplace-City Center Distance (miles)	[0, 1)	[1,2)	[2,4)	[4, 8)	[8, 16)	[16, 32)	32+
[0, 1)	48.50	57.94	56.53	68.80	80.03	67.23	80.11
[1,2)	42.10	43.55	35.84	49.18	60.61	57.19	66.97
[2,4)	33.58	41.49	33.63	38.97	46.45	48.02	62.66
[4, 8)	38.48	48.31	33.84	33.99	38.82	40.87	43.56
[8, 16)	45.42	61.98	45.57	42.76	35.51	40.58	42.34
[16, 32)	56.01	71.58	55.56	54.37	44.16	36.48	47.12
32+	75.52	83.88	73.26	76.90	76.83	61.85	51.69

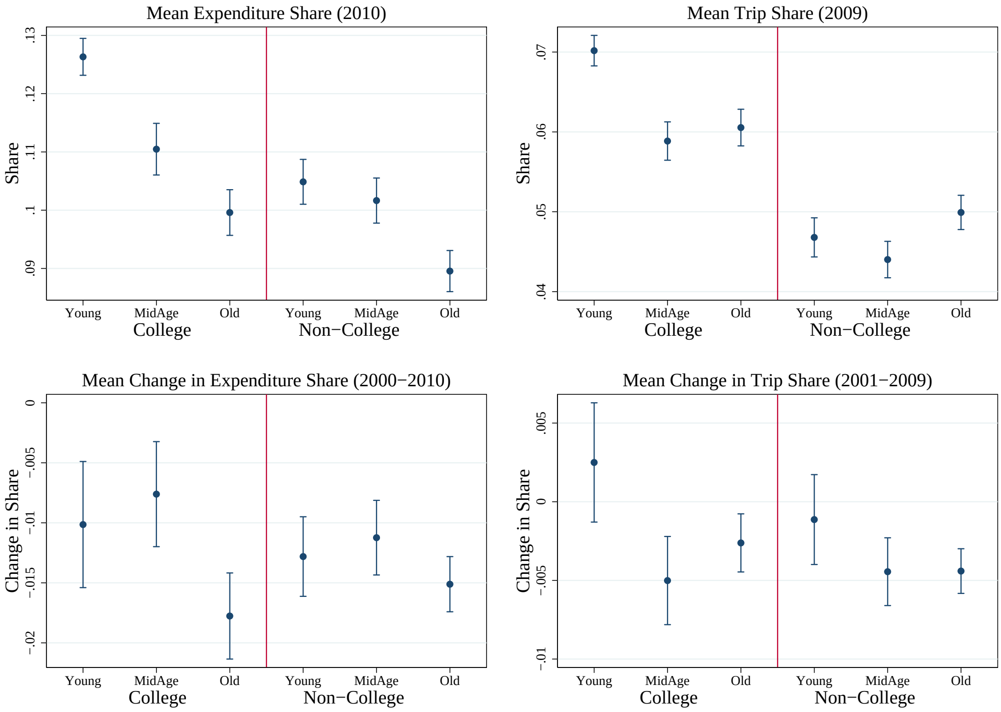
Panel C: High-Wage Workers in Largest 10 CBSAs

Residence-City Center Dist. \ Workplace-City Center Distance (miles)	[0, 1)	[1,2)	[2,4)	[4, 8)	[8, 16)	[16, 32)	32+
[0, 1)	78.44	97.24	110.60	105.37	80.36	65.39	69.00
[1,2)	93.22	69.63	62.09	82.21	62.21	55.85	68.90
[2,4)	81.85	95.44	60.68	69.70	49.80	39.92	59.27
[4, 8)	68.30	103.79	43.34	42.56	34.18	24.82	30.15
[8, 16)	47.63	81.92	34.83	28.67	19.68	25.19	29.50
[16, 32)	35.92	62.33	30.79	30.58	23.24	25.76	36.93
32+	67.69	96.85	53.77	56.41	54.16	46.80	40.21

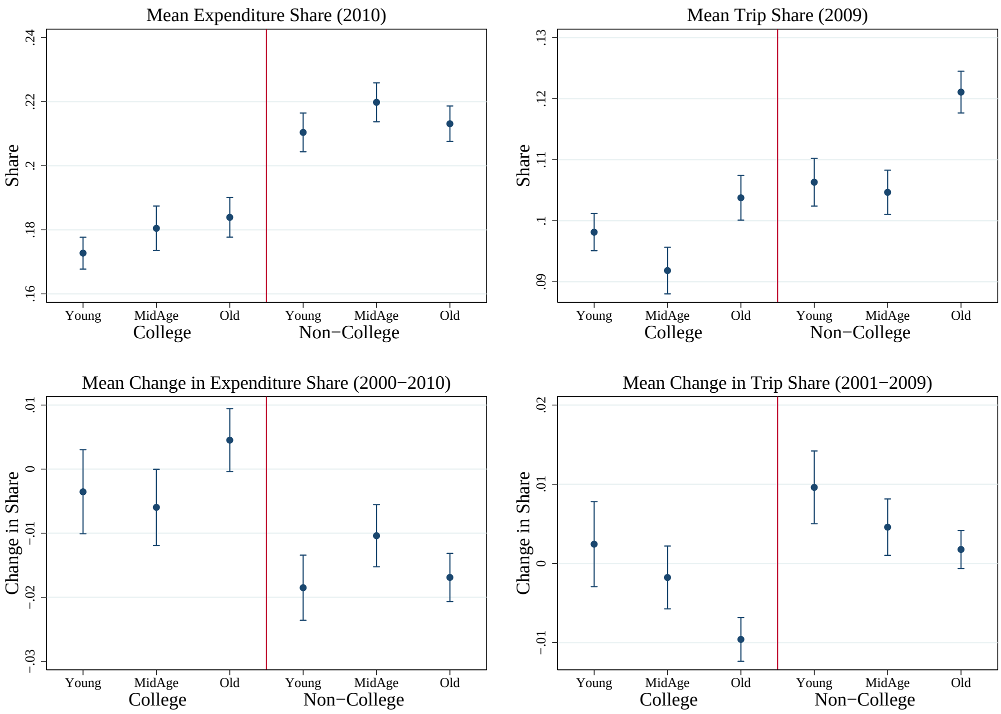
Notes: Data from LODES 2002 and 2011. Each cell shows the percentage change from 2002 to 2011 in the number of workers living and working at given distance intervals from the city center. High-wage workers earn more than $3333/month in nominal dollars.

Figure 8: Expenditure and Trip Shares on Non-Tradable Services and Tradable Retail in the CEX and NHTS

Panel A: Restaurants



Panel B: Food Expenditures and Trips to Buy Goods

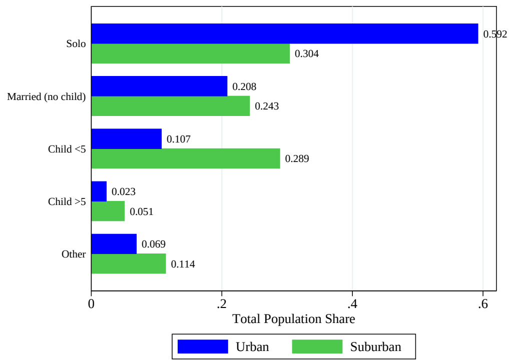


Notes: All confidence intervals are 95% intervals. Data from the Consumer Expenditure Survey (CEX) and the National Household Travel Survey (NHTS). The left-hand column of each panel shows mean CEX expenditure shares for each age-education group and the right-hand column shows mean NHTS trip shares. Trip shares to buy goods include food, apparel, and hardware.

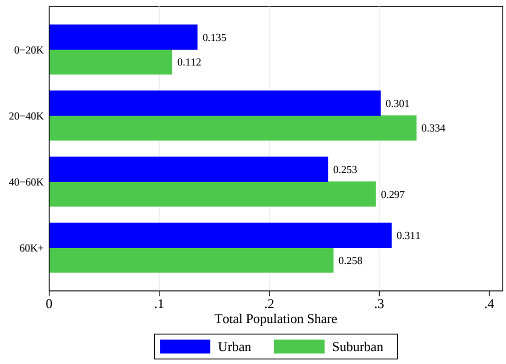
Figure 9: Share of Young College-Educated Individuals by Household Type and Income Bracket

Panel A: Share in Urban vs. Suburban Area in 2000

By Household Types

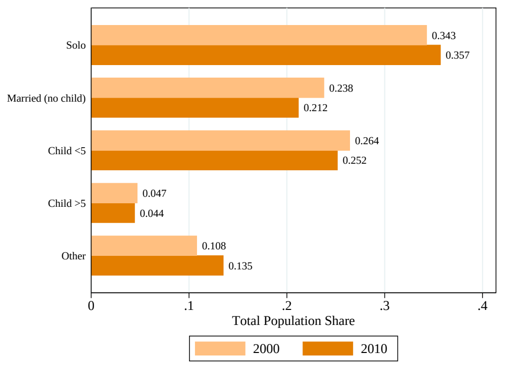


By Income Brackets

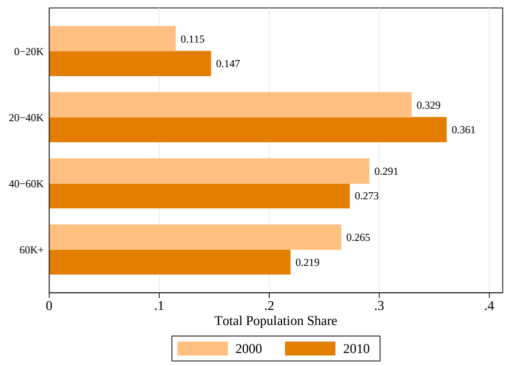


Panel B: Share in 2000 vs. 2007-2011

By Household Types



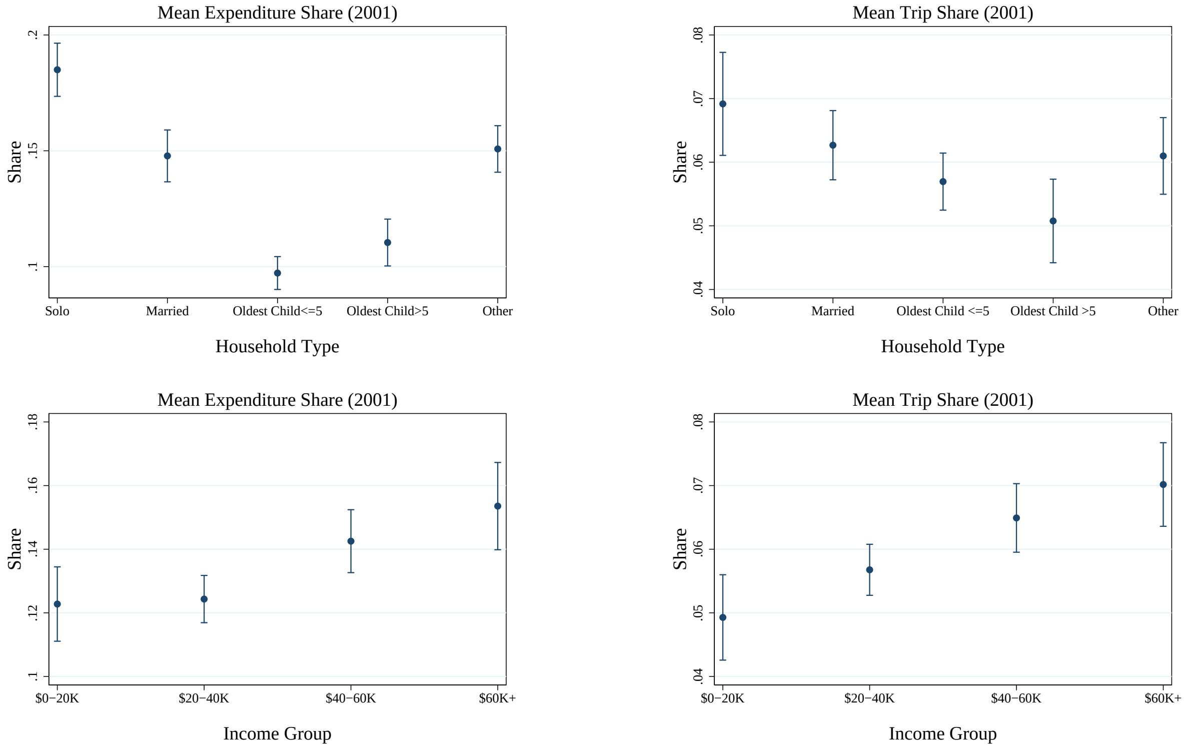
By Income Brackets



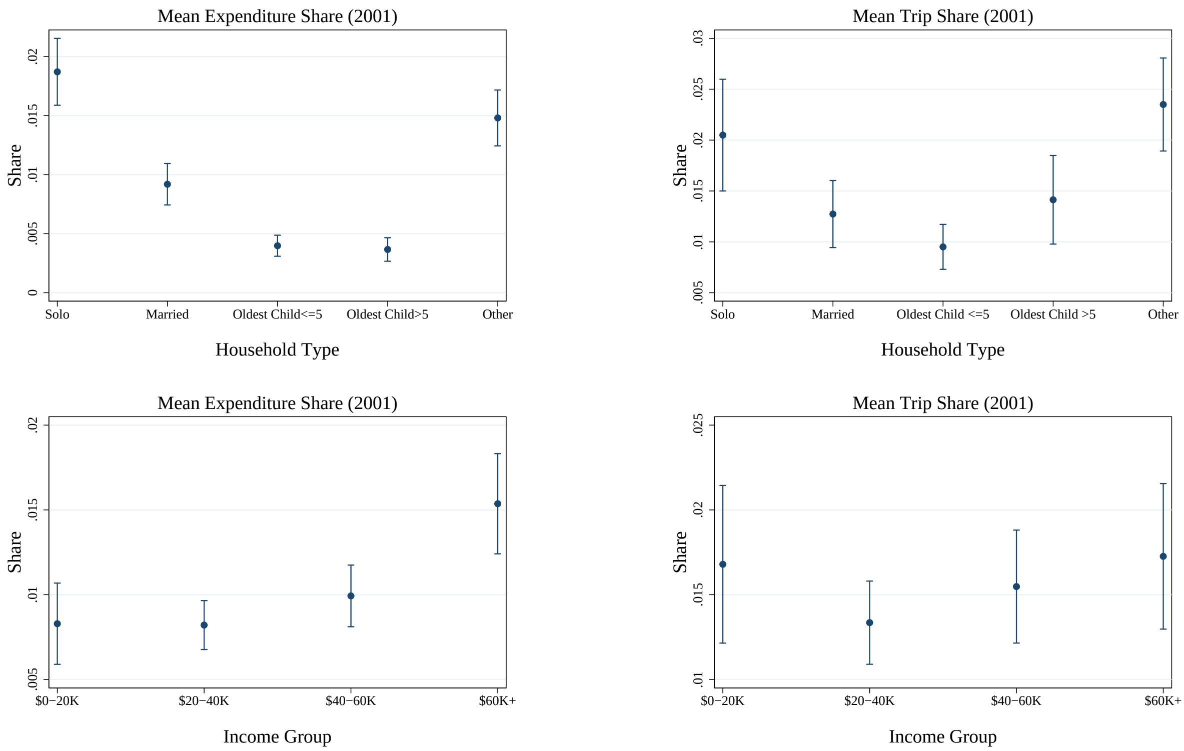
Notes: Shares computed out of all 25-34 year-old college-educated individuals in the 50 largest CBSAs from the 5 percent sample of IPUMS in 2000 and an aggregate of five 1 percent samples from 2007 to 2011. 2007-2011 income is discounted to 1999 dollars by a factor 0.741 (recommmended with IPUMS data) and all income is adjusted for household size using the OECD equivalence scale. The urban area of each CBSA is defined by sequentially adding PUMAs closest to the city center until it accounts for 10 percent of the population.

Figure 10: Expenditure and Trip Shares on Non-Tradable Services by Household Types and Income Brackets for the Young and College-Educated

Panel A: Restaurants



Panel B: Bar Expenditures and Trips to Go Out/Hangout



All confidence intervals are 95% intervals. Data from the Consumer Expenditure Survey (CEX) and the National Household Travel Survey (NHTS) for all college-educated individuals 25-34 year old. Mean CEX expenditure shares on the left and mean NHTS trip shares on the right.

Appendices

A Data Appendix

The following appendix provides detailed information of all data sources.

A.1 Census Data and ACS Data

Census Tract Data and Definitions For our stylized facts on recent urban growth, we assemble a database of constant 2010 geography census tracts using the Longitudinal Tract Data Base (LTDB) and data from the National Historical Geographic Information System (NHGIS) for the 1980-2000 censuses and the 2008-2012 ACS. In each of the censuses from 1970 to 2000, some tracts are split or consolidated and their boundaries change to reflect population change over the last decade. The LTDB provides a crosswalk to transform tract level variable from 1970 to 2000 censuses into 2010 tract geography. This reweighting relies on population and area data at the census block level, which is small enough to ensure a high degree of accuracy. We combine these reweighted data with the 2008-2012 ACS data, which already uses 2010 tract boundaries.

CBSA Definitions Core Based Statistical Areas (CBSAs) refer collectively to metropolitan and micropolitan statistical areas. CBSAs consist of a core area with substantial population, together with adjacent communities that have a high degree of economic and social integration with the core area. We assign 2010 census tracts to CBSAs based on 2013 CBSA definitions. Our model estimation sample consists of all 355 metropolitan area CBSAs.

A.2 LODES Data

The LODES data comes from the Longitudinal Employer-Household Dynamics (LEHD) data. The LODES data consists of three parts: origin-destination (OD), workplace area characteristics (WAC), and residence area characteristics (RAC). The WAC data provides counts of workers in each census block by wage groups and 20 NAICS sectors that we use to compute our job density indexes and wage group-specific Bartik instruments. We use the OD data for the residence-workplace model of subsection 5.2. The OD data provides counts of workers working and living in a census block pair by age and income groups (but not for age-income interactions). For each census block pair, counts are available for three age groups (29 or younger, 30 to 54, and 55 or older) and three nominal wage groups ($1,250/month or less, $1,251/month to $3,333/month, and greater than $3,333/month). We aggregate the OD data at the tract level and exclude federal workers.

The LODES data for general public use is processed to protect the workers' confidentiality (Graham et al., 2014).[54] There are two aspects to confidentiality protection in the LODES data.[55] First, the residential location of workers is synthesized. That is, the residential census block of a worker is "coarsened" and drawn from a distribution of blocks within the same census tract, PUMA or Super-PUMA. Graham et al. (2014) note that only 10 percent of residences are coarsened above the census tract level, so synthesis has no impact on 90 percent of our sample, which is aggregated at the tract level. Moreover, only residential-workplace pairs with very small shares – generally for long commutes - have residences coarsened at a geography larger than a census tract. Our weighted regressions ensure that small cells have little impact on our estimation results. Second, the workplace location of residents is subject to noise infusion and small cell imputation. These procedures again have the most impact on block-pairs with very small worker counts, and both tract-level aggregation and weighted regressions ensure a minimal impact of these procedures on our estimates.[56]

A.3 NETS Data

The 2012 National Establishment Time-Series (NETS) Database includes 52.4 million establishments with time-series information about their location, industries, performance and headquarters from 1990-2012. The NETS dataset comes from annual snapshots of U.S. establishments by Duns and Bradstreet (D&B). D&B collects information on each establishment through multiple sources such as phone surveys, Yellow Pages, credit inquiries, business registrations, public records, media, etc. Walls & Associates converts D&B's yearly data into the NETS time-series. The NETS data records the exact address for about 75 percent of establishments. In the remaining cases, we observe the establishments zip code and assign it's location to the zip code centroid.

Neumark et al. (2007) assess the NETS reliability by comparing it to other establishment datasets (i.e., QCEW, CES, SOB and BED data). Their conclusions support our use of the NETS data to compute a long 10-year difference in establishment density. They report that NETS has better coverage than other data sources for very small establishments (1-4 persons), which is often the size of consumption amenity establishments.

We further assess the precision of the NETS by considering aggregate growth of chain estab-

[54]The complexity and opacity of these procedures may discourage academic use of the data. We share these concerns, but argue that too much caution is unwarranted in many empirical contexts including ours.

[55]Another source of measurement error comes from the LEHD source data, in which 40 percent of jobs are at multi-establishment employers. The state of Minnesota reports establishment level data, so the LEHD uses Minnesota data to impute an establishment to workers at multi-establishment employers in other states. For instance, workers are more likely imputed to establishments closer to their residence.

[56]See Graham et al. (2014) for additional technical details on these procedures, comparison with the ACS commute data, and further references on the LEHD and LODES data creation.

lishments. For instance, according to Stock and Wong (2015), Chipotle had nearly 100 stores in 2000 and grew to about 1000 stores in 2010. The NETS reports 21 Chipotle stores in 2000 and around 800 in 2010. These numbers show that the NETS data captures general growth patterns, but we struggle to identify all chains due to merging on inconsistent establishment names and lags in D&B recording new locations. Table 5 reports the number of establishments nationally in 2000 and 2010 in each of our nine consumption amenity type, as well as the SIC codes used to define these types.[57]

A.4 Zillow House Price Indexes

Our main house price index comes from Zillow.com.[58] Our "all home" index is the Zillow House Value Index (ZHVI) for all homes (i.e., single family, condominium, and cooperative), which is available monthly for 10,452 zip codes in 2000 and 11,118 zip codes in 2010. In robustness checks, we use the Zillow House Value Index for 2-Bedroom Homes, which is available monthly for 7,423 zip codes in 2000 and 8,941 zip codes in 2010 and HUD's Fair Market Rent Series (FMR) for one bedroom, two bedroom and three bedroom homes, which is calculated annually for 3,038 counties in 2000 and 3,042 counties in 2010. For each zip code in the Zillow data, we compute a yearly index by averaging over all months of the year. We map zip codes to tracts with a crosswalk from HUD. We compute the tract-level index as the weighted average of the home value index across all zip codes overlapping with the tract, using as weights the share of residential address in the tract falling into each each zip code. For tracts falling partly into missing zip codes, we normalize the residential share in zip codes with available data to one. If a tract does not fall into a zip code with available data, but instead falls into a tract grouping defined in Ferreira and Gyourko (2011) in which some other tracts have available data, we assign to this tract the average index of these other tracts in the group. The final data set contains home value indexes for 51,165 tracts in 2000 (9,478 tracts inferred from tract group average) and 53,784 tracts in 2010 (8,685 tracts inferred from tract group average).

A.5 UCR Crime Data

The crime data comes from the Uniform Crime Reporting Program (UCR) from 1990, 2000 and 2010. As in Ellen et al. (2017), we use data on violent crimes, which include murder, rape, robbery, and aggravated assault. UCR relies on each city's police district to self-report their crime statistics to the FBI. Therefore, we lack coverage for police districts that did not report.

[57] The NPD Group, a marketing agency, reports 579,416 restaurants in the spring of 2010. Couture (2013) reports 273,000 restaurants on Google Local in States accounting for 50 percent of the U.S. population, suggesting close to 550,000 restaurants nationally. By comparison, the NETS reports 416,807 restaurants nationally in 2010.

[58] The index and methodology are available at: http://www.zillow.com/research/data/.

Table A.1: NETS Establishment Counts and SIC codes

Category	Description [1]	00 Estab. Counts [2]	10 Estab. Counts [3]	SIC Codes [4]
Non-Tradable Services				
Restaurants	full service, fast food, etc.	437570	416807	581200 - 581209
Bars	bar, clubs, lounge, etc.	64948	75261	581300 - 581302
Personal Services	nails, hair, beauty, etc.	385745	544486	723, 724, 729901, 729902
Gym and Sports	gyms, tennis courts, etc.	134613	193238	7991, 7997, 7999
Stores				
Food Stores	grocery stores, markets, bakeries, etc.	281269	335802	54
Apparel Stores	apparel stores	197909	239863	56
General Merchandise	department, variety stores	43468	54797	53
Activities				
Museums, Galleries and Libraries	museums, art galleries, libraries, etc.	35972	52961	84120000, 84120100 - 84120102, 84129901 - 84129903, 842200 - 842202, 823100 - 823104
Parks and Golf	amusement parks, golf courses	10438	9727	7992, 7996

Notes: Columns 2 and 3 show the total number of establishments in each type of consumption amenity in the U.S. in 2000 and 2010 respectively.

There are multiple police districts within each CBSA. In 1990, there were 9,222 police districts reporting, which increased to 11,044 in 2010, partially because new cities were incorporated. To impute district-level data to census tracts, we use GIS software to map every 2010 census tract into the corresponding district or districts that it overlaps with. We then assign the crime total for each district to the tracts that overlap with it (population-weighted overlap) assuming that population and crime are uniformly distributed within tracts and within districts. The final data set contains crime data for 54,745 tracts in 1990 and 57,095 tracts in 2010 after discarding tracts that do not overlap with any districts.

A.6 Consumption Expenditure Survey (CEX) Data

The Consumer Expenditure Survey (CEX) is conducted by the U.S. Department of Labor's Bureau of Labor Statistics. We use the public-use micro-data from the CEX Diary Survey for years 1998 to 2002 and 2008 to 2012. These surveys record all expenditures for each respondent, including expenditures on small, frequently purchased items over two consecutive one-week periods, as well as characteristics, income and weights for the consumer unit (household). Each CEX expenditure receives a Universal Classification Code (UCC) that we match to our amenity types as follows:

1. Restaurants (UCC 190111 - 190926, "Food away from home" (excluding beer, wine and other alcohol))
2. Bars (UCC 200511 - 200536, Beer, wine and other alcohol in "Food away from home")
3. Food Stores (UCC 10110 - 180720, "Food")
4. Apparel stores (UCC 360110 - 410901, "Apparel")

To obtain population estimates of mean expenditure shares, we use weights at the consumer unit level (total sample weight). Our sample size for the 24-35 year-old college-educated (smallest group) is 7,166 individuals in 1998-2002 and 7,111 individuals in 2008-2012.

A.7 National Household Transportation Survey (NHTS) Data

The National Household Travel Survey (NHTS) conducted by the Federal Highway Administration (and local partners) provides travel diary data on daily trips taken in a 24-hour period for each individual in participating households. We use the 2001 and 2009 NHTS surveys. Each trip has a WHYTO (trip purpose) code that we match to our amenity types as follow:

1. Restaurants (WHYTO 80, 82, 83, "Meals", "get/eat meal", "coffee/ice cream/snacks")

2. Bars (WHYTO 54, "Go out/hang out: entertainment/theater/sports event/go to bar")

3. Food Stores (WHYTO 41, "Buy goods: groceries/clothing/hardware store")

4. Apparel stores (WHYTO 41, "Buy goods: groceries/clothing/hardware store")

We use weights at the person level to compute population estimates of mean trip shares.[59] Our sample size for the 24-35 year-old college-educated (smallest group) is 6,228 individuals in 2001 and 7,309 individuals in 2009.

B Variable Definitions

This appendix details the computation of the dependent variable in our regression, as well as the measures of job density, and consumption amenity density, quality, and diversity.

B.1 Dependent Variable

The dependent variable, $\Delta \ln \tilde{s}^d_{jc}$, is the 2000 to 2010 log change in the share of age-education group d that lives in tract j of CBSA c relative to a base tract. It comes from tract-level population counts by age and education from the decennial census of 2000 and from the American Community Survey (ACS) 2008-2012 aggregates, as in our stylized facts. Let n^d_{jct} be the number of individuals of group d in tract j in CBSA c. The share of all type d residents who live in tract j in CBSA c at time t is then:

$$s^d_{jct} = \frac{n^d_{jct}}{\sum_c \sum_j n^d_{jct}}.$$

B.2 Job Density Index

We use the LODES data to compute a distance-weighted average of the number of jobs in tracts surrounding each residential tract in 2002 and in 2011. The job density index for a tract j' for wage group g is:

$$\text{job density}^g_{j't} = \sum_j w(d_{j'j}) n^g_{j'jt} \text{ where } w(d_{j'j}) = \frac{1/(d_{j'j}+1)}{\sum_j 1/(d_{j'j}+1)},$$

[59]The NHTS reports household income in brackets. We use the midpoint of each bracket, and $167,000 for the top bracket "$100,000+" , as an estimate for household income. The 2009 survey excludes children under five, but we know the age-range of the youngest child. If any child in a household does not fill the survey and we know that the youngest child in the household is younger than five, then we assume that the child who did not fill the survey is younger than five.

where $n^g_{j'jt}$ is the number of workers who work in tract j but do not live in tract $j^{'}$,and $d_{j'j}$ is the linear distance in miles between the centroids of tract j and j'.

B.3 Amenity Variables

Consumption Amenity Density Indexes We measure the level and change in the availability of different types of establishments around each tract's centroid. The amenity density index for, say, restaurants in tract j is high if there are many restaurants within a short travel time of tract j's centroid. The amenity density index for a given type is the inverse of a CES price index, in which the price of visiting an establishment includes transport cost, as in Couture (2013). We assume an elasticity of substitution of 8.8, estimated by Couture (2013) with restaurant data. The higher the elasticity, the lower the weight on establishments far away from the tract centroid, and the more localized the amenity index. The price of a visit to an establishment is a constant derived from the CEX for that type, plus a transport cost by foot from the tract centroid.[60] So for each type a, the density index in tract j is:

(A.1)
$$A_{aj} = \frac{1}{\left(\sum_{i=1}^{I_j}(p_a + t_{ij})^{1-\sigma}\right)^{1/1-\sigma}},$$

where *p* is the average price of a visit to an establishment in amenity type *a*, *t* is the travel cost of a two-way trip to establishment *i* from the tract centroid *j*, I_j is the set of all NETS establishments in type a within 50 miles of a tract, and σ is the elasticity of substitution equal to 8.8. To compute travel costs, we start with the linear distance from tract j's centroid to an establishment i.[61] To go from linear to actual travel distance, we use an average ratio of actual to linear travel distance computed from each tract's centroid to a random sample of 100 NETS establishments on Google Maps. To go from travel distance to travel time, we use Google Maps' constant walking speed of 20 minutes per mile.

[60]Using CEX expenditures that most closely match our amenity types, we set a price of $34.8 for museums, $36.7 for parks/golf and gyms/sports, $10.2 for restaurants, $12.4 for bars, $18.9 for personal services, $60.4 for general merchandise stores, $36.5 for food stores and $60.4 for apparel stores. Transport costs assume a value of time equal to $12 dollars per hour (equal to 50 percent of the average U.S. wage as suggested in Small and Verhoef, 2007).

[61]When there are no establishments within 50 miles of a tract centroid, a tract receives a top code for that amenity type equal to the highest non-missing value in the tract sample. Usually around 5-10 percent of tracts are top-coded depending on the type, although amusement parks and golf are top-coded in more than 50 percent of cases.

Consumption Amenity Quality, Diversity, and Independent Establishment Index

The restaurant and food store quality indexes of section 7.2 are weighted averages of Esri's MPI ratings for all 61 rated restaurant chains and all 24 food store chains near a tract. The CES weights are exactly as in equation A.1, so the quality index for type a in tract j is:

$$\text{Q}_j = \frac{\sum_{i=1}^{I_j} MPI_i \times (p + t_{ij})^{(1-\sigma)}}{\sum_{i=1}^{I_j} (p + t_{ij})^{(1-\sigma)}}. \tag{A.2}$$

The amenity diversity indexes of section 7.2 are inverse Herfindahl indexes, which capture the diversity of the 70 restaurant SIC8 types and 66 food store SIC8 types:

$$\text{H}_{ji} = \frac{1}{\sum_i {m_{ij}}^2}, \tag{A.3}$$

where m_{ij} is the market share of SIC8 code i within 50 miles of tract j. When computing market shares, each restaurant receives the same CES weight as in equations A.1 and A.2.

The independent establishment index of section 7.4 is a similarly weighted average of the share of independent establishments near a tract. We compute it using equation A.2, after replacing MPI_i with a dummy equal to 0 for all establishments part of chains with at least five establishments with the same name, and 1 for all other "independent" establishments.

Predicted Establishment Entry and Exit for Consumption Amenity Density and Quality Instrument Table A.2 uses regression results from equation 4 aggregated over all 559 SIC8 codes for establishments included in our nine consumption amenity density indexes. The table provides the percentage of SIC8 codes for which an explanatory has a positive (column 1) or negative (column 2) and significant coefficient at the 10 percent level. Table A.3 presents similar results for entry and exit regressions on the 61 restaurant chains for which we have MPI ratings (patterns are similar for food stores).[62]

[62] A negative entry prediction prevents computation of the chain-instrument. To limit the occurrence of negative MPI weights in our vector of "predicted" establishments, we aggregate chains into ten MPI deciles. If a bin with negative MPI weight persists, we remove it.

Table A.2: Tract-level Predicted Establishment Entry at the SIC8 Level

	Percentage of SIC8-Specific Coefficients		
	Negative and Significant	Positive and Significant	Not Significant at 10% Level
Same SIC8			
Within 0-1 miles	93%	1%	6%
Within 1-2 miles	56%	7%	37%
Within 2-4 miles	28%	21%	52%
Within 4-8 miles	16%	32%	51%
Same SIC6, Different SIC8			
Within 0-1 miles	6%	47%	47%
Within 1-2 miles	12%	27%	61%
Within 2-4 miles	13%	20%	67%
Within 4-8 miles	19%	22%	59%
Same SIC4, Different SIC6			
Within 0-1 miles	10%	52%	38%
Within 1-2 miles	13%	25%	62%
Within 2-4 miles	19%	14%	67%
Within 4-8 miles	26%	19%	55%

Notes: The table lists each explanatory variable in the entry and exit regression in Equation 4 at the SIC8 level, and provides the percentages of significant variables at the 10 percent level out of all 559 SIC8 codes within our nine consumption amenity types.

Table A.3: Tract-level Predicted Restaurant Entry at the Chain Level

	Percentage of Chain-Specific Coefficients		
	Negative and Significant	Positive and Significant	Not Significant at 10% Level
Same Chain			
Within 0-1 miles	89%	2%	10%
Within 1-2 miles	64%	2%	34%
Within 2-4 miles	18%	34%	48%
Within 4-8 miles	2%	66%	33%
Same SIC8, Different Chain			
Within 0-1 miles	7%	52%	41%
Within 1-2 miles	9%	24%	67%
Within 2-4 miles	17%	11%	72%
Within 4-8 miles	17%	11%	72%
Same SIC6, Different SIC8			
Within 0-1 miles	11%	39%	50%
Within 1-2 miles	17%	19%	64%
Within 2-4 miles	17%	8%	75%
Within 4-8 miles	39%	8%	53%
Same SIC4, Different SIC6			
Within 0-1 miles	2%	70%	28%
Within 1-2 miles	7%	33%	61%
Within 2-4 miles	18%	13%	69%
Within 4-8 miles	46%	11%	43%

Notes: The table lists each explanatory variable in the entry and exit regression in Equation 4 at the chain level, and provides the percentages of significant variable at the 10 percent level out of the 61 restaurant chains for which we have MPI data.

C Additional Tables and Results

Table A.4: First Stage for Base IV Specification in Table 2

Endogeneous Variable Name [1]	**Reduced-Form F Stat** [2]	**Conditional SW F Stat** [3]	**Under-ID SW Chi-2** [4]
Change in Job Density (Low Wage)	1823.31	68.40	2054.64
Change in Job Desnity (Med. Wage)	1888.26	69.60	2090.80
Change in Job Desnity (High Wage)	1641.77	79.14	2377.45
Change in House Price Index	873.95	82.59	2480.91
Change in Restaurant Density	231.34	9.17	275.57
Change in Food Store Density	350.17	24.90	748.06
Population Density	2853.82	145.46	4369.67
Share of Same Type	635.09	18.49	555.53
Change in Within CBSA share	53.23	22.70	681.89

Notes: Column 2 reports the reduced-form first-stage statistics, column 3 reports the first-stage SW conditional F-statistic from Sanderson and Windmeijer (2016), and column 4 reports an under-identification test, also from Sanderson and Windmeijer (2016). Sanderson and Windmeijer (2016) do not report critical values for their F-statistic and recommend the use of Cragg-Donald critical values from Stock and Yogo (2005), which are unavailable for regressions with more than two endogenous variables. The standard rule of thumb is that an F-statistic smaller than ten is weak, in the sense that either the bias of the IV estimator is larger than 10 percent of the bias of the OLS estimator at the 5 percent confidence level or else that a 5-percent Wald test rejects hypotheses at more than the 10-percent level (Stock and Yogo, 2005).

Table A.6: Nested Logit Residential Location Choice Regression Results for the Young and College-Educated: Comparison with Change Only Regression (IV)

	Base Spec		Changes Only	Changes with Level Controls	
	Change	Level	Change	Change	Level
Variable	[1]	[2]	[5]	[3]	[4]
Job Density (Low Wage)	-0.114***	-0.035***	-0.202***	-0.131***	
Job Density (Med. Wage)	-0.090***	0.036***	-0.034***	-0.035***	
Job Opportunities (High Wage)	0.212***	0.012	0.197***	0.137***	
House Price Index	0.019***	-0.009***	0.022***	0.030***	
Restaurant Density	0.385***	0.394***	0.131***	0.174***	
Food Store Density	-0.047***	-0.215***	-0.041***	-0.048***	
Population Density		-0.049***			0.024***
Share of Same Type		-0.099***			-0.039***
Within-CBSA Share	0.753***		0.887***	0.808***	
Observations	33,941		33,941	33,941	

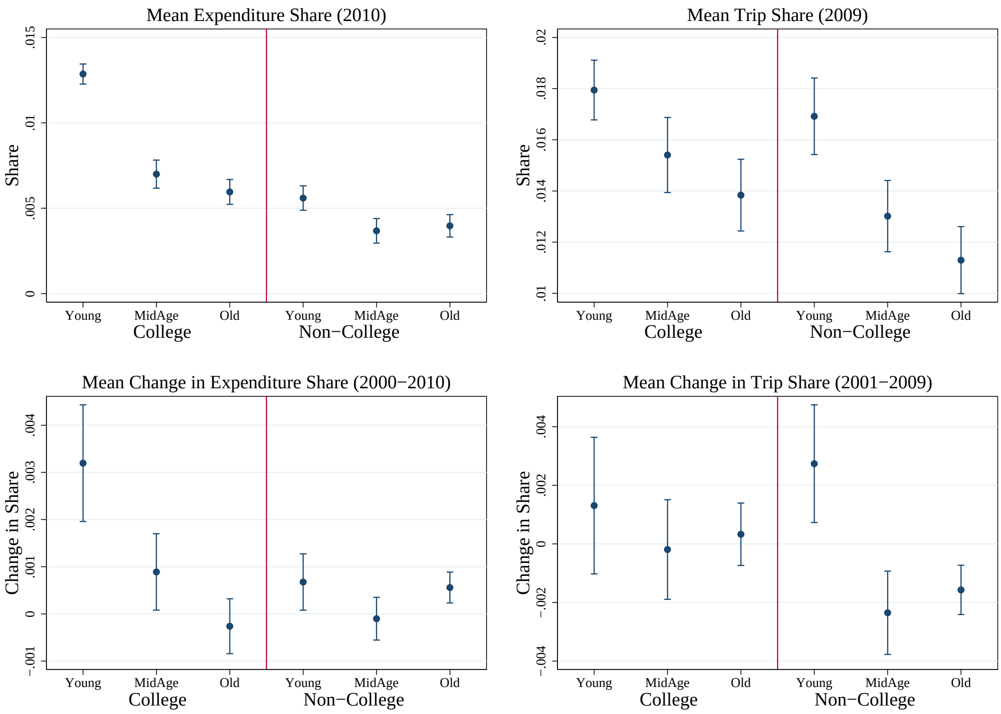
Notes: * – 10%; ** – 5%; ***–1%. The change in house prices, consumption amenity density, job density, within-CBSA share, and the levels of local demographic shares and house prices are instrumented. Each regression is weighted by the share of type d in tract j in year 2000.

Table A.5: Ranking of Restaurant Density's Urbanizing Contribution Relative to Other Variables and of Restaurant Density's Coefficient Size Relative to Other Age Education Groups for the Young and College-Educated

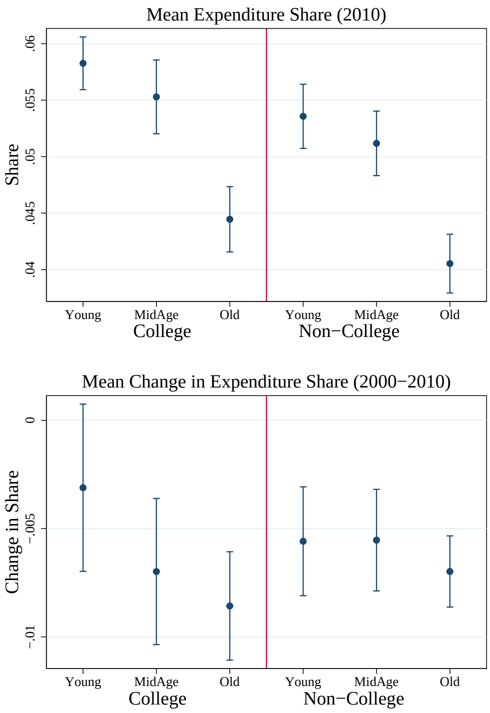
	IV				OLS (except for within-CBSA share)			
	Rank of Contribution		Rank of Young College Coeff (/6)		Rank of Contribution		Rank of Young College Coeff (/6)	
Specification	Level	Change	Level	Change	Level	Change	Level	Change
Panel A: Base Specification, 2 Amenities								
Base Specification	1 (+)	13 (-)	1 (+)	1 (+)	3 (+)	8 (-)	1 (+)	2 (+)
Non-nested with CBSA Fixed-Effects	1 (+)	14 (-)	1 (+)	1 (+)	6 (+)	10 (-)	1 (+)	1 (+)
Panel B: with Additional Controls								
Distance to City Center	1 (+)	14 (-)	1 (+)	1 (+)	3 (+)	9 (-)	1 (+)	2 (+)
School, Crime and Transit	1 (+)	18 (-)	2 (+)	3 (+)	3 (+)	10 (-)	1 (+)	3 (+)
Homophily Control	1 (+)	14 (-)	2 (+)	4 (+)	5 (+)	10 (-)	1 (+)	4 (+)
Amenity Quality and Diversity	1 (+)	20 (-)	1 (+)	1 (+)	3 (+)	17 (-)	1 (+)	1 (+)
Panel C: with Alternative Housing Index								
Ferreira/Gyourko Hedonic Index	1 (+)	13 (-)	1 (+)	1 (+)	3 (+)	7 (-)	1 (+)	2 (+)
Zillow 2 Bedroom Index	1 (+)	13 (-)	3 (+)	3 (+)	3 (+)	8 (-)	1 (+)	2 (+)
Housing on the LHS	1 (+)	10 (-)	4 (+)	5 (+)	3 (+)	7 (-)	1 (+)	6 (+)
HUD Fair Market Rents	1 (+)	13 (-)	1 (+)	2 (+)	3 (+)	9 (-)	1 (+)	2 (+)
Housing Age	1 (+)	12 (-)	1 (+)	1 (+)	3 (+)	10 (-)	1 (+)	1 (+)

Figure A.1: Expenditure and Trip Shares on Bars and Apparel Stores in the CEX and NHTS

Panel A: Bar Expenditures and Trips to Go Out/Hangout



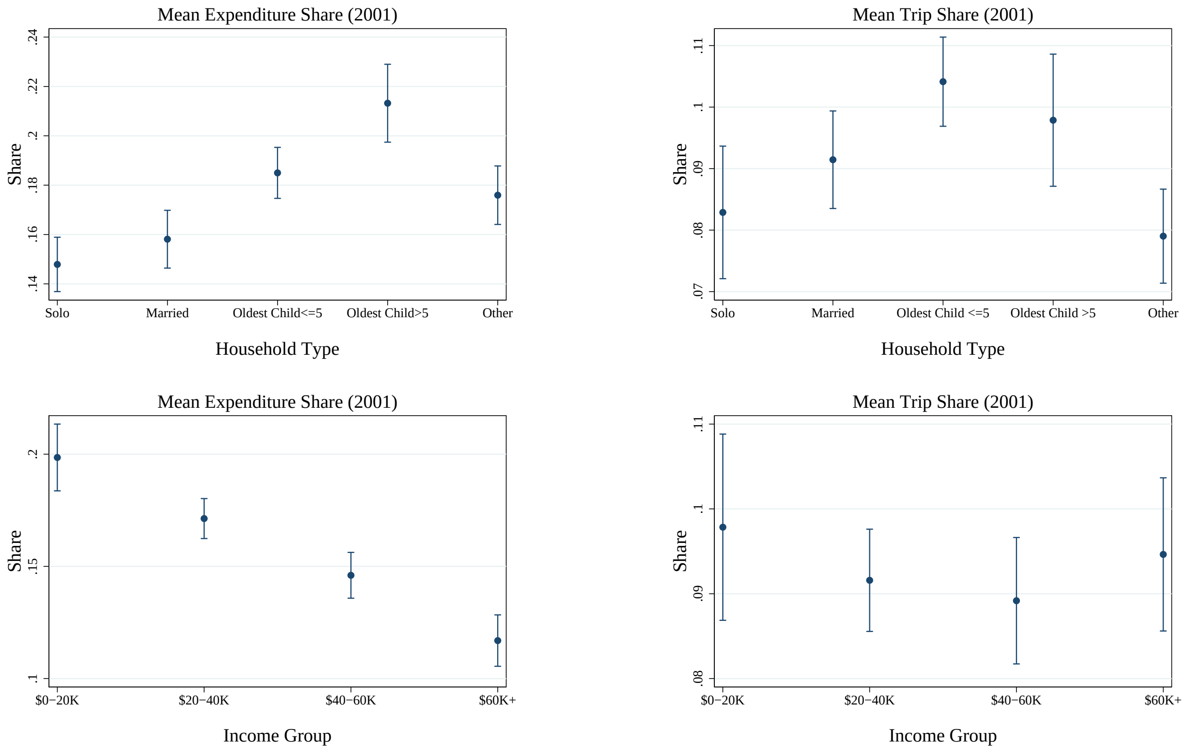
Panel B: Apparel Store Expenditures and Trips to Buy Goods



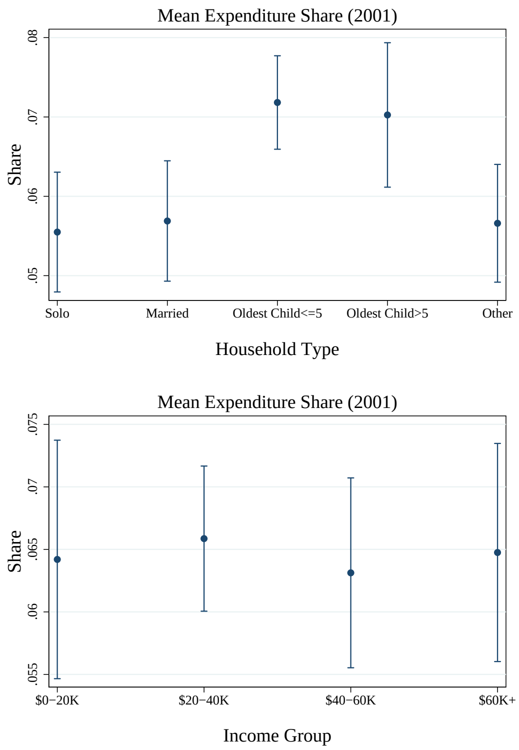
Notes: All confidence intervals are 95 percent intervals. Data from the Consumer Expenditure Survey (CEX) and the National Household Travel Survey (NHTS). The left hand column of each panel shows mean CEX expenditure shares for each age-education group and the right hand column shows mean NHTS trip shares. Trip shares to buy goods include food, clothes and hardware.

Figure A.2: Expenditure and Trip Shares on Tradable Retail by Household Types and Income Bracket for Young College-Educated

Panel A: Food Store Expenditures and Trip Shares to Buy Goods



Panel B: Apparel Store Expenditures



Notes: All confidence intervals are 95% intervals. Data from the Consumer Expenditure Survey (CEX) and the National Household Travel Survey (NHTS) for all college-educated individuals 25-34 years old. Mean CEX expenditure shares on the right and mean NHTS trip shares on the left. Trip shares to buy goods includes food, clothes and hardware.

D Commute Model

D.1 Specification and Identification

The commute model is the same as the residential choice model described in section 3, except that each individual i chooses both its residential location j and workplace location k in year t to maximize its indirect utility function V^i_{jkt}:

$$\text{(A.4)} \qquad \max_{j,k} V^{id}_{jkct} = \alpha^d_t \mathbf{X_{jct}} + \beta^d_t \mathbf{X_{kct}} - \omega^d d_{jkc} + \xi^d_{jct} + \chi^d_{kct} + \mu^d_{jkc} + \theta^d_{ct} + \varepsilon^{id}_{jkct},$$

where $\mathbf{X_{jt}}$ and $\mathbf{X_{kt}}$ are vectors of observable time-varying characteristics of residences and workplaces, d_{jk} is the travel distance from residence j to workplace k, and ω^d reflects group d's marginal disutility from commuting. μ^d_{jkc} is a time-invariant unobserved quality that will cancel out in first-difference, ξ^d_{jct} and χ^d_{kct} are unobserved group-specific, time-varying qualities of each residential and workplace location, and θ^d_{ct} is an unobserved time-varying quality of CBSA c for individuals in group d.[63] We solve the model as in section 3 and the resulting estimating equation is:

$$\text{(A.5)} \qquad \Delta \ln\left(s^d_{jk}\right) = \alpha^d_{2011} \Delta \tilde{\mathbf{X}}_{\mathbf{jc}} + \Delta \alpha^d \tilde{\mathbf{X}}_{\mathbf{jc,2002}} + \sigma^d_{kc} + \Delta \omega^d d_{jkc} + \Delta \tilde{\xi}^d_{jc} + \epsilon^d_{jkct},$$

where we included a workplace fixed-effect σ^d_{kc} capturing both observed and unobserved group-specific and time-varying workplace characteristics. The key identifying assumption is that the workplace fixed-effect captures the entire impact on residential choice of any changes in job location. To understand this assumption, note that as in Monte et al. (2015), individuals select both their place of work and place of residence simultaneously and get a joint logit residential-workplace preference shock. The joint logit shock implies that for any group d, job growth in work tract k reallocates residents to a given residential tract j in direct proportion to the initial share of group d workers in tract k who lived in tract j. This reallocation pattern, a consequence of the logit's IIA property, is reasonable because one expects faster residential growth in tracts initially providing more commuters to fast-growing work tracts. The commute distance control relaxes this reallocation assumption by allowing for group-specific changes in distaste for long commute. As in section 3, the identification strategy relies on first-differencing and a rich set of controls, but it exploits workplace fixed effects instead of instrumental variables to estimate the

[63] We use the CBSA fixed-effect instead of the nested-logit structure, because adding both the workplace fixed-effect and the nested-logit "within-CBSA" term restricts the identifying variation to the very small share of individuals who live and work in two different CBSAs in our sample.

impact of residential characteristics free of simultaneity with job locations.

To clearly show the impact of removing the bias due to endogenous job location, we compare estimates from the workplace fixed-effects model of equation A.5 to those from a pure residential choice model estimated with the LODES data. To do so, we collapse the LODES data to the residential tract level and, starting from equation A.5, we obtain the following estimating equation for a pure residential choice model:

$$\widetilde{\Delta \ln s_{jk}^d} = \alpha_{2011}^d \Delta \tilde{\mathbf{X}}_{\mathbf{jc}} + \Delta \alpha^d \tilde{\mathbf{X}}_{\mathbf{jc},\mathbf{2002}} + \tilde{\theta}_c^d + \Delta \tilde{\xi}_{jc}^d + \epsilon_{jkct}^d, \tag{A.6}$$

where $\tilde{\theta}_c^d$ is a CBSA fixed-effect and $\Delta \tilde{\xi}_{jc}^d$ are unobserved time-variant tract characteristics included in the error term of the regression.

D.2 Variable Definition

Before estimating the model, we describe the variables in equation A.5 that are not in the residential choice model of section 3.

Commute Shares

The dependent variable in the workplace fixed-effect model is the log change in the share of residents of group d living and working in a residential-workplace tract pair, between 2002 and 2011, relative to a base tract pair. Let n_{jkct}^d be the number of group-d people who live in tract j and work in tract k in CBSA c in year t. Let c be the CBSA of tract k and let L_c be the set of tracts located in CBSA c. The share of workers who live in tract j and work in tract k in CBSA c in year t is therefore:

$$s_{jkct}^d = \frac{n_{jkct}^d}{\sum_c \sum_j \sum_{k \in L_c} n_{jktc}^d}.$$

To estimate a residential choice model using LODES data, we simply aggregate s_{jkct}^d over all work tracts in all CBSAs to obtain:

$$s_{jct}^d = \sum_c \sum_{k \in L_c} s_{jkct}^d.$$

Commute Distance

We proxy for the commute time between the workplace and residence tract by controlling for the Haversine distance d_{jkc} between workplace tract k and residential tract j in the estimating equation.

Residence Tract Characteristics

Residential characteristics are the same as in the residential choice model of Section 3. Note that with a workplace fixed-effect, the variables for job density become purely residential characteristics. They capture the possibility that households value living near employment locations other than their own, and such job opportunities may be relevant to dual-career households or to future career events.

D.3 Results

The estimation results are in Table A.7. We include the full set of nine consumption amenity densities to highlight the impact of the workplace fixed-effects on the residential characteristic coefficients. Column 1 provides, for comparison, results for the young college-educated population from our base OLS specification in Table 1. Column 2 and 3 show LODES results for high-wage workers, with the pure residential choice model of equation A.6 in column 2 and the workplace fixed-effect model of equation A.5 in column 3. High-wage workers are the only group in the LODES data urbanizing in large cities between 2002 and 2011, albeit much less strikingly than the young and college-educated. Introducing the workplace fixed-effect in column 3 reduces the size of the coefficient on high-wage job density to near zero, which shows that once we control for where they work, high-wage workers do not value living near jobs. Other residential characteristics stay remarkably constant from the residential to the workplace fixed-effect model. For instance, the coefficients on changes in restaurants, bars, gyms, and personal services, are positive, significant and of similar magnitude in both models. This is the key result of this section, suggesting that the simultaneous determination of workplace and residential location is likely not an important source of bias on residential characteristics in our within-city residential choice model.[64]

[64] We find the same result if we run IV regressions using the set of instrumental variables from section 4.2.1.

Table A.7: Commute Model Regression Results

Variable	25-34, College-Educated Change [1]	Level [2]	High Wage (Residential) Change [3]	Level [4]	High Wage (Work FE) Change [5]	Level [6]
Job Density (Low Wage)	-0.069***	-0.045***	0.121***	0.019	0.042***	0.000
Job Density (Med. Wage)	0.045***	0.106***	-0.319***	0.116***	-0.114***	-0.007*
Job Opportunities (High Wage)	0.039***	-0.050***	0.196***	-0.088***	-0.015***	-0.016***
House Price Index	0.060***	-0.004***	0.046***	0.009***	0.039***	0.015***
Museum and Library Density	-0.002***	-0.008***	0.000	0.013***	-0.001***	0.007***
Golf and Park Density	-0.001**	0.003***	0.000	-0.001	0.002***	0.005***
Gym and Sports Density	0.011***	0.018***	0.017***	-0.004	0.019***	0.007***
Restaurant Density	0.006***	0.006***	0.009***	-0.035***	0.008***	-0.022***
Bar Density	0.001*	0.004***	0.008***	0.023***	0.005***	0.018***
Personal Services Density	0.011***	0.003	0.027***	0.002	0.027***	0.007***
Merchandise Store Density	-0.004***	-0.013***	0.007***	0.054***	-0.002***	0.024***
Food Store Density	0.001	-0.007**	0.019***	0.009*	0.015***	-0.001
Apparel Store Density	0.005***	0.002	0.005***	-0.020***	0.003***	-0.025***
Population Density		-0.004**		-0.175***		-0.144***
Share of Same Type		-0.038***		-0.072***		-0.043***
Observations	33,941		45,309		3,292,773	

Notes: * – 10% significance level; ** – 5% significance level; ***–1% significance level. Column 1 and 2 shows main OLS specification of Table 1 for the young and college-educated. Column 3 to 4 shows the residential choice model using LODES data for high-wage workers, and column 5 and 6 shows the workplace fixed effects model for these workers. Each regression is weighted by the share of type d in tract in year 2000. Standard error clustered at the tract level in the workplace fixed effects model.

Non-tradable service levels make the second-most important contribution to urbanizing high-wage workers, even though most workers earning more than \$3,333 per month are not among the young and college-educated group. The most important contribution comes from the negative coefficient on average distance between workplace and residential tract. This indicates that the aversion of high-wage individuals to long commutes increased from 2002 to 2011. We are cautious in interpreting this result for two reasons. First, this variable is also the most important contributor to *suburbanizing* young workers (< 30 years old). Non-tradable services levels, however, urbanize both young and high-wage workers. Second, Figure 7 shows that commute length *increased* for high-wage workers from 2002 to 2011, especially for those living closest to the city center, due to reverse commuting. Tellingly, the only other LODES group urbanized by rising aversion to long commutes is older workers (>54 years old), which is consistent with census results in Table 2, in which middle-aged and old college-educated individuals have the most positive change in taste for proximity to high-wage jobs (this does not hold in the OLS). All of this suggests that while the aversion of older college-educated individuals for long commutes may have risen, this trend is not strong enough to urbanize this group. Finally, we note that our estimates of rising aversion to commute length in high-wage workers support the hypothesis in Edlund et al. (2015) that a taste for shorter commutes - through higher value of time - contributes to downtown gentrification. The rising value of time for high-wage workers is consistent with the rising preference for proximity to both jobs and consumption amenities, and could be the topic of promising empirical work.

Online Appendix

This document contains supplementary material for the paper "Urban Revival in America, 2000 to 2010", by Victor Couture and Jessie Handbury. Appendix E contains additional stylized facts documenting growth across CBSAs of different sizes from 1980 to 2010. Appendix F contains additional results not shown in the paper. Appendix G contains detailed derivation for the estimating equation and the urbanization contribution plots in the paper. Appendix H discusses the hypothesis that reduced access to homeownership following the housing crisis explains the urbanization of young college graduates.

E Additional Stylized Facts

Tables A.8 and A.9 present an alternative way of tabulating the data from Figure 1 in the main text.

Table A.8: Counts of CBSAs: Urban Population Growth Greater than Suburban Population Growth

		All Individuals			College-Educated Indviduals			College-Educated Individuals 2000-2010		
Urban Def.	CBSA Pop. Rank	1980-1990 [1]	1990-2000 [2]	2000-2010 [3]	1980-1990 [4]	1990-2000 [5]	2000-2010 [6]	Age 25-34 [7]	Age 35-44 [8]	Age 45-64 [9]
5 pct.	1-10	1	1	3	4	4	9	10	9	4
5 pct.	11-50	3	1	3	5	6	19	33	18	5
5 pct.	51-100	4	4	3	9	2	12	23	15	9
5 pct.	101-355	67	46	36	84	49	66	104	97	93
10 pct.	1-10	2	0	2	4	3	7	9	9	3
10 pct.	11-50	1	2	0	4	3	11	29	14	2
10 pct.	51-100	5	3	3	9	3	9	21	10	6
10 pct.	101-355	64	28	21	61	28	41	78	72	73
15 pct.	1-10	2	0	2	3	1	7	8	8	1
15 pct.	11-50	2	2	1	2	3	9	24	12	2
15 pct.	51-100	7	3	4	8	2	5	18	6	4
15 pct.	101-355	60	30	23	65	21	33	79	63	50
2 Miles	1-10	1	2	5	3	6	10	10	9	5
2 Miles	11-50	1	2	7	4	10	26	36	23	10
2 Miles	51-100	5	3	3	9	2	9	20	10	6
2 Miles	101-355	76	37	38	85	28	40	100	66	60
3 Miles	1-10	1	1	4	4	3	10	10	10	5
3 Miles	11-50	2	2	1	5	6	17	31	20	3
3 Miles	51-100	6	3	4	9	2	6	19	9	4
3 Miles	101-355	85	48	45	94	30	40	100	56	55
5 Miles	1-10	2	0	2	4	3	9	10	9	5
5 Miles	11-50	2	0	0	1	3	10	30	16	1
5 Miles	51-100	4	3	5	6	2	6	15	6	3
5 Miles	101-355	96	66	58	102	41	43	94	59	51
Central	1-10	1	0	1	1	0	8	8	8	0
Central	11-50	3	2	1	3	4	8	19	10	4
Central	51-100	8	5	6	11	4	9	17	8	5
Central	101-355	79	42	23	75	31	28	75	46	41

Notes: Columns 1 to 3 give the number of CBSAs for which the total urban population grew faster than the total suburban population for the given decade. Columns 4 to 6 give the number of CBSAs for which the college-educated population grew faster in urban areas relative to suburban areas for the given decade. Finally, columns 7 to 9 give the number of CBSAs for which the college-educated population grew faster in urban areas relative to suburban areas from 2000 to 2010 by age group. These counts are given for various urban definitions and CBSA population rankings. The X pct. urban area is the set of all tracts closest to the city center accounting for X percent of a CBSA's population, and the X miles urban area is the set of all tracts with centroids within X miles of the city center.

Table A.9: Population Growth in Urban and Suburban Areas by CBSA Size

		All Individuals						College-Educated Individuals						College-Educated Individuals, 2000-2010					
	CBSA	1980-1990		1990-2000		2000-2010		1980-1990		1990-2000		2000-2010		Age 25-34		Age 35-44		Age 45-64	
Urban	Pop.	Urb	Sub	Urb	Sub	Urb	Sub	Urb	Sub	Urb	Sub	Urb	Sub	Urb	Sub	Urb	Sub	Urb	Sub
Def.	Rank	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]
5 pct.	1-10	0.00	0.14	0.04	0.14	0.05	0.09	0.56	0.70	0.31	0.32	0.46	0.29	0.53	0.16	0.38	0.09	0.38	0.46
5 pct.	11-50	0.01	0.26	0.00	0.14	-0.05	0.11	0.48	0.76	0.29	0.42	0.37	0.34	0.43	0.18	0.23	0.15	0.33	0.50
5 pct.	51-100	-0.05	0.19	-0.01	0.17	-0.03	0.13	0.41	0.72	0.08	0.36	0.18	0.33	0.22	0.23	-0.02	0.10	0.24	0.46
5 pct.	101-355	-0.02	0.18	-0.02	0.14	-0.03	0.13	0.51	0.87	0.11	0.37	0.11	0.31	0.09	0.23	-0.10	0.08	0.22	0.40
10 pct.	1-10	0.00	0.15	0.03	0.15	0.02	0.10	0.53	0.71	0.29	0.33	0.39	0.29	0.45	0.14	0.31	0.08	0.35	0.47
10 pct.	11-50	-0.03	0.21	0.01	0.18	-0.03	0.14	0.48	0.78	0.25	0.44	0.30	0.35	0.31	0.18	0.19	0.15	0.32	0.51
10 pct.	51-100	-0.01	0.20	0.00	0.15	-0.01	0.14	0.41	0.74	0.11	0.37	0.18	0.34	0.17	0.23	-0.03	0.10	0.26	0.46
10 pct.	101-355	0.05	0.27	0.01	0.15	-0.02	0.12	0.55	0.89	0.11	0.39	0.11	0.33	0.08	0.24	-0.09	0.09	0.20	0.41
15 pct.	1-10	0.02	0.16	0.04	0.16	0.01	0.11	0.51	0.73	0.26	0.34	0.33	0.29	0.37	0.14	0.24	0.08	0.32	0.47
15 pct.	11-50	0.06	0.29	0.02	0.16	0.00	0.12	0.56	0.92	0.13	0.41	0.12	0.34	0.26	0.19	0.14	0.16	0.31	0.52
15 pct.	51-100	-0.02	0.23	0.02	0.19	-0.02	0.15	0.49	0.81	0.24	0.46	0.26	0.36	0.15	0.24	-0.04	0.11	0.27	0.48
15 pct.	101-355	0.00	0.21	0.01	0.16	0.01	0.15	0.42	0.77	0.13	0.39	0.17	0.35	0.09	0.25	-0.08	0.10	0.20	0.43
2 Miles	1-10	0.01	0.13	0.05	0.14	0.12	0.09	0.57	0.69	0.31	0.32	0.51	0.29	0.61	0.17	0.37	0.10	0.41	0.46
2 Miles	11-50	-0.05	0.18	0.00	0.17	0.01	0.13	0.50	0.75	0.33	0.42	0.41	0.34	0.44	0.19	0.32	0.15	0.36	0.50
2 Miles	51-100	-0.01	0.19	0.00	0.14	-0.01	0.14	0.45	0.72	0.08	0.37	0.17	0.33	0.18	0.23	-0.03	0.10	0.26	0.46
2 Miles	101-355	0.08	0.29	0.03	0.16	0.01	0.12	0.61	0.91	0.13	0.42	0.12	0.34	0.11	0.25	-0.09	0.11	0.20	0.43
3 Miles	1-10	0.00	0.14	0.03	0.14	0.07	0.09	0.59	0.70	0.31	0.32	0.49	0.29	0.58	0.15	0.40	0.09	0.39	0.46
3 Miles	11-50	-0.04	0.19	-0.01	0.17	-0.02	0.14	0.49	0.77	0.28	0.43	0.32	0.34	0.34	0.18	0.23	0.15	0.32	0.50
3 Miles	51-100	0.01	0.21	0.01	0.16	0.00	0.15	0.43	0.76	0.11	0.39	0.16	0.35	0.16	0.24	-0.04	0.11	0.26	0.47
3 Miles	101-355	0.10	0.31	0.05	0.17	0.03	0.13	0.64	0.97	0.18	0.45	0.15	0.37	0.14	0.26	-0.08	0.13	0.22	0.46
5 Miles	1-10	0.00	0.14	0.02	0.15	0.04	0.09	0.59	0.70	0.29	0.33	0.44	0.29	0.52	0.14	0.35	0.09	0.37	0.46
5 Miles	11-50	-0.03	0.21	0.00	0.18	-0.03	0.14	0.48	0.79	0.25	0.44	0.29	0.35	0.15	0.27	-0.05	0.15	0.24	0.49
5 Miles	51-100	0.01	0.23	0.02	0.17	0.01	0.16	0.44	0.80	0.13	0.41	0.17	0.36	0.30	0.18	0.19	0.15	0.31	0.51
5 Miles	101-355	0.11	0.34	0.07	0.17	0.05	0.14	0.67	1.01	0.21	0.48	0.17	0.40	0.14	0.25	-0.04	0.12	0.27	0.49
Central	1-10	0.05	0.17	0.07	0.16	0.02	0.12	0.51	0.77	0.22	0.36	0.28	0.31	0.29	0.15	0.14	0.09	0.30	0.50
Central	11-50	0.06	0.23	0.08	0.19	0.02	0.17	0.59	0.83	0.30	0.47	0.25	0.38	0.20	0.20	0.11	0.17	0.35	0.54
Central	51-100	0.06	0.21	0.07	0.16	0.04	0.16	0.50	0.80	0.19	0.41	0.15	0.38	0.12	0.27	-0.07	0.14	0.27	0.51
Central	101-355	0.09	0.31	0.06	0.16	0.03	0.13	0.59	0.98	0.18	0.44	0.13	0.37	0.11	0.27	-0.10	0.14	0.21	0.46

Notes: Columns 1 to 6 give the percentage change in the total population for a given CBSA group in urban areas and suburban areas by decade. Columns 7-12 give the percentage change in the college-educated population for a given CBSA group in urban areas and suburban areas by decade. Columns 13 to 18 give the percentage change in the college-educated population from 2000 to 2010 for a given CBSA group in urban areas and suburban areas by age group. The X pct. urban area is the set of all tracts closest to the city center accounting for X percent of a CBSA's population, and the X miles urban area is the set of all tracts with centroids within X miles of the city center.

F Additional Results

Figure A.3 shows the urbanization contribution kernel plots for all explanatory variables by age-education groups. We use coefficients from the base IV specification, given in Table 2 of the main text. The left-hand plots shows the contribution of variables in changes, and the right-hand plot shows that of variables in initial levels.

G Additional Derivations

G.1 Derivation of Estimating Equation in Section 4.2

The utility maximization problem of each individual i of type d is to choose its residential location tract j in CBSA c in year t to maximize its indirect utility function V^i_{jct}:

$$\text{(A.7)} \qquad \max_j V^i_{jct} = \beta^{d(i)}_{wt} \ln w^{d(i)}_{jct} - \beta^{d(i)}_{At} \ln p_{Ajct} - \beta^{d(i)}_{Ht} \ln p_{Hjct} + \beta^{d(i)}_{at} \ln a_{jct} + \mu^{d(i)}_{jc} + \xi^{d(i)}_{jct} + \psi^i_{ct}(\sigma^{d(i)}) + (1-\sigma^{d(i)})\nu^i_{jct}.$$

This equation is described in the main text and, as outlined in Berry (1994), it yields a linear equation for the share $\tilde{s}^d_{jct}$ of individuals in group d who choose tract j in CBSA c relative to a base tract $\bar{j}$. We can write the share of type d individuals living in residential location j in year t as the product of the within-CBSA share of individuals living in location j in year t and the CBSA share of individuals in year t:

$$s^d_{jct} = s^d_{j|ct} s^d_{ct},$$

where

$$s^d_{j|ct} = \frac{\exp\left(V^d_{jct}/(1-\sigma^d)\right)}{D^d_{ct}},$$

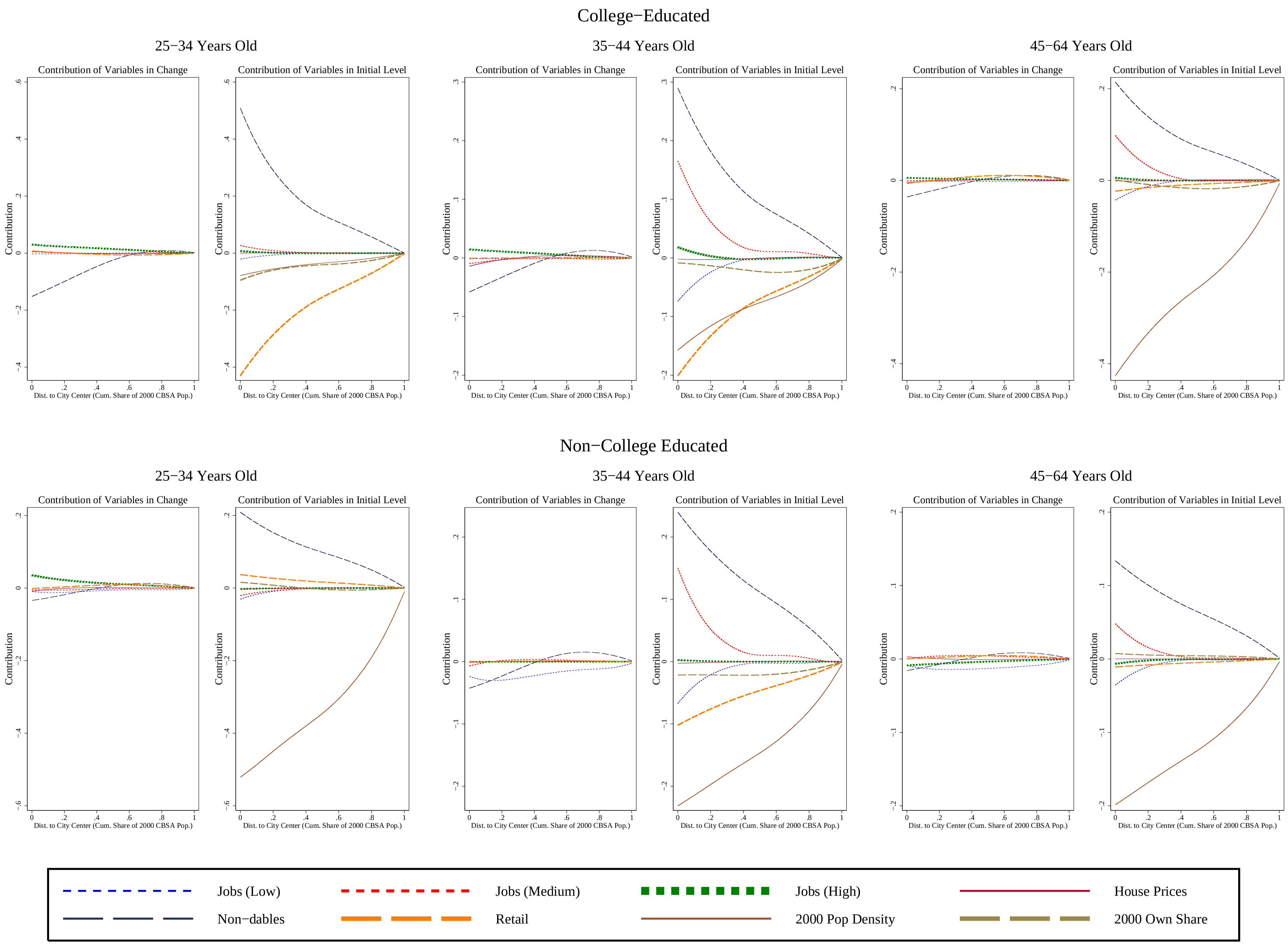
$$s^d_{ct} = \frac{\left(D^d_{ct}\right)^{1-\sigma^d}}{\sum_{c\in C}\left(D^d_{ct}\right)^{1-\sigma^d}},$$

and

$$D^d_{ct} = \sum_{j\in J_c} \exp\left(V^d_{jct}/(1-\sigma^d)\right),$$

where J_c denotes the set of residential locations in CBSA c, C denotes the universe of CBSAs, and $V^d_{jct} = \beta^d_{wt} \ln w^d_{jct} - \beta^d_{At} \ln p_{Ajct} - \beta^d_{Ht} \ln p_{Hjct} + \beta^d_{at} \ln a_{jct} + \mu^d_{jc} + \xi^d_{jct}$ denotes the mean utility for an individual of type d from residential location j in year t. Following Berry (1994),

Figure A.3: Variables Contributing to Growth at Various Distance from the City Center by Age-Education Groups



Notes: See description in text.

this collapses to:

$$s^d_{jct} = \frac{\exp\left(V^d_{jct}/(1-\sigma^d)\right)}{\left(D^d_{ct}\right)^{\sigma^d}\sum_{c\in C}\left(D^d_{ct}\right)^{1-\sigma^d}}. \tag{A.8}$$

Fixing some tract $\bar{j}$ in CBSA $\bar{c}$ as the base residential location, we have that the log expected share of type-d people who reside in location j in CBSA c in year t relative to the log expected share that reside in location $\bar{j}$ in CBSA $\bar{c}$ in year t is equal to:

$$\ln s^d_{jct} - \ln s^d_{\bar{j}\bar{c}t} = \frac{V^d_{jct} - V^d_{\bar{j}\bar{c}t}}{1-\sigma^d} - \sigma^d\left(\ln D^d_{ct} - \ln D^d_{\bar{c}t}\right). \tag{A.9}$$

Substituting $D^d_{ct} = \sum_{j\in J_c}\exp\left(V^d_{jct}/(1-\sigma^d)\right)$ and $\ln s^d_{jct} = \ln s^d_{ct} + \ln s^d_{j|ct}$ into (A.9) and rearranging terms, we have that:

$$\ln s^d_{jct} - \ln s^d_{\bar{j}ct} = \left(V^d_{jct} - V^d_{\bar{j}\bar{c}t}\right) + \sigma^d\left(\ln s^d_{j|ct} - \ln s^d_{\bar{j}|\bar{c}t}\right).$$

From this, we obtain equation 2 from the main text:

$$\ln\tilde{s}^d_{jct} = \beta^d_{wt}\ln\tilde{\mathbf{w}}_{\mathbf{jct}} + \beta^d_{At}\ln\tilde{\mathbf{A}}_{\mathbf{jct}} - \beta^d_{Ht}\ln\tilde{p}_{Hjct} + \mu^d_{jc} + \tilde{\xi}^d_{jct} + \tilde{\xi}^d_{wjct} + \sigma^d\ln\tilde{s}^d_{j|c}, \tag{A.10}$$

where $\tilde{X}_j = X_j - X_{\bar{j}}$ and we normalize $\mu_{\bar{j}c}$ to equal zero.

G.2 Derivation of Contribution Plots in Figure A.3

We first outline how these fitted moments are calculated. We start with the fitted value from equation 3 for the expected change in the share of the total national population in a given demographic group d that resides in tract j relative to the share of that demographic group that resides in our base tract l:

$$\widehat{\Delta\ln\tilde{s}^d_{jc}} = \sum_k \widehat{\delta}^d_k X_k,$$

where $\sum_k \widehat{\delta}^d_k X_k = \widehat{\beta^d_{w,2010}}\ln\Delta\tilde{\mathbf{w}}_{\mathbf{jc}} + \widehat{\Delta\beta^d_w}\ln\tilde{\mathbf{w}}_{\mathbf{jc,2000}} + \widehat{\beta^d_{A,2010}}\Delta\ln\tilde{\mathbf{A}}_{\mathbf{jc}} + \widehat{\Delta\beta^d_A}\ln\tilde{\mathbf{A}}_{\mathbf{jc,2000}} + \widehat{\beta^d_{p_H,2010}}\Delta\ln\tilde{p}_{Hjc} + \widehat{\Delta\beta^d_{p_H}}\ln\tilde{p}_{Hjc,2000}$, and as usual we exclude the within-CBSA share. We un-difference this fitted value from the change in the observed share of demographic group d in the base tract l and from the observed share of demographic group d residing in tract j in 2000 to get a fitted value for the log share of demographic group d that resides in tract j in 2010:

$$\widehat{\ln s^d_{jc,2010}} = \widehat{\Delta\ln\tilde{s}^d_{jc}} + \Delta\widetilde{\ln s^d_{lc}} + \ln s^d_{jc,2000}.$$

We take the exponent of this fitted 2010 log share and multiply it by the population of demographic group d in 2010 to get the fitted value for the population of demographic group d in tract j in 2010:

$$\widehat{pop^d_{jc,2010}} = \exp\left(\ln \widehat{s^d_{jc,2010}}\right) * pop^d_{2010}.$$

We divide this fitted population for demographic group d by the observed total population of tract j in 2010 to arrive at the share of tract j's population in demographic group d in 2010. We then difference this 2010 fitted level from the observed value of this share in 2000 to get a fitted prediction of the change in population share that is represented in the contribution plots:

$$\Delta \widehat{s_{d|jc,2010}} = \frac{\widehat{pop^d_{jc,2010}}}{pop^{all}_{jc,2010}} - \frac{pop^d_{jc,2000}}{pop^{all}_{jc,2000}}.$$

Putting this together and rearranging terms we have that:

$$\text{(A.11)} \qquad \Delta\widehat{s_{d|jc,2010}} = \left(\frac{pop^d_{jc,2000}}{pop^{all}_{jc,2000}}\right)\left(\underbrace{\left[\prod_k \exp\left(\widehat{\delta^d_k} X_{jc,k}\right)\right]}_{\text{Estimated Scaling Factors}} \underbrace{\exp\left(\Delta \ln \tilde{s}^d_{jc}\right)\left(\frac{pop^d_{2010}}{pop^d_{2000}}\right)\left(\frac{pop^{all}_{jc,2010}}{pop^{all}_{jc,2000}}\right)^{-1}}_{\text{Observed Changes in Population Shares and Levels}} - 1\right).$$

Equation A.11 shows that the contribution of any single regression factor, X_k, to the spatial distribution of demographic group d across tracts j (and, therefore, to the change in each demographic group's population share in each tract) depends on a scaling factor $\exp\left(\widehat{\delta^d_k} X_{jc,k}\right)$, where $\widehat{\delta^d_k}$ is the estimated non-standardized regression coefficient on tract characteristic $X_{jc,k}$. The change in demographic group d's population share of tract j from 2000 to 2010 is determined by the product of these estimated scaling factors multiplied by the product of various observed demographic changes, including the change in demographic group d's population share of the base tract l from 2000 to 2010, the change in the aggregate population of demographic group d from 2000 to 2010, and the inverse of the change in tract j's population from 2000 to 2010.

Figure A.3 plots estimated kernels of these scaling factors, $\exp\left(\widehat{\delta^d_k} X_{jc,k}\right)$, in each tract j in all CBSAs c, for each age-education group d, and for each regression factor k against the distance of tract j from the city center of CBSA c. The plots on the left-hand side for each age-education group depict the kernel of the scaling factors associated with the change in environment ($\widehat{\beta^d_{w,2010}}\Delta \ln \tilde{\mathbf{w}}_{\mathbf{jc}}$, $\widehat{\beta^d_{A,2010}}\Delta \ln \tilde{\mathbf{A}}_{\mathbf{jc}}$, and $\widehat{\beta^d_{H,2010}}\Delta \ln \tilde{p}_{Hjc}$), while those on the right-hand side depict the scaling factors associated with the change in preferences ($\widehat{\Delta\beta^d_w}\ln \tilde{\mathbf{w}}_{\mathbf{jc,2000}}$, $\widehat{\Delta\beta^d_A} \ln \tilde{\mathbf{A}}_{\mathbf{jc,2000}}$, $\widehat{\Delta\beta^d_A} \ln \tilde{\mathbf{A}}_{\mathbf{jc,2000}}$ and $\widehat{\Delta\beta^d_H} \ln \tilde{p}_{Hjc,2000}$).

H Additional Hypothesis

One prominent hypothesis to explain the recent urbanization of certain population groups is reduced access to homeownership following the housing crisis and recession of 2007-2009. Given that rental units (almost always multifamily) are more urbanized than owner-occupied units (generally single-family homes), a decline in accessibility to homeownership that disproportionately affects young college graduates could pull them out of the suburbs and push them into urban areas.

In the aftermath of the housing crisis, credit score requirements for access to mortgage credit became more stringent. For instance, the average FICO credit score of mortgages acquired by Fannie Mae and Freddie Mac rose from 725 in 2007 to more than 760 by 2010 (Parrott and Zandi 2013).[65] Presumably, this reduction in credit availability has been disproportionately harmful to younger individuals about to enter the housing market, and may have driven them away from homeownership and toward rental options. Consistent with this story, Rappaport (2015) documents the rapid increase in multifamily construction starting in 2010, and the increased propensity of young adults to live in multifamily units as opposed to single-family homes following the housing crisis.

The main flaw in this hypothesis is the timing of the housing crisis: the 2000s include more years of historically easy mortgage credit than of restricted credit. Using IPUMS data and a methodology similar to that in Section 7.3, we decompose the growth of the young and college-educated by tenure type (owners and renters) from 2000 to 2010. Renters are indeed more prevalent in urban areas. However, homeowners have grown *faster* nationally than renters.[66] Therefore, the premise of the housing market hypothesis that young college graduates have been forced into renting from 2000 to 2010 is not supported by the data. In fact, further analysis reveals that homeownership rates among young college graduates increased in both urban and suburban areas over that period.

To provide additional support for this conclusion, we replicate our stylized facts using the earliest available ACS data, from 2005-2009 (not shown). We find patterns of urban revival very similar to those observed in later years. Given that the housing crisis only covers half of the 2005-2009 time period, this result again challenges to notion that reduced access to mortgage credit drives urban revival.

[65] In 2010, Fannie and Freddie acquired 61 percent of total new home mortgage originations (Jaffee and Quigley, 2011).

[66] The number of young and college-educated homeowners has grown by 19 percent from 2000 to 2010, versus 8 percent for renters.