Urban Revival in America, 2000 to 2010*

Victor Couture† and Jessie Handbury‡

July 2017

Abstract

This paper documents and explains the rising proclivity of college-educated individuals to locate near city centers. We show that this recent trend is driven entirely by younger cohorts in larger cities, and represents a striking reversal of fortune for urban America. With a residential choice model, we quantify the role of amenities, jobs, and house prices in explaining this urban revival. In our model, these changing preferences of young professionals for non-tradable serviceslike restaurant, bars, gyms, and beauty salons account for between 50 to 80% of their growth near city centers. Complementary datasets on expenditures and travel confirm that non-tradable services are rising in importance for the young and college-educated. Our investigation into the causes of these changing preferences highlights the expanding role of non-tradable services in generating socializing opportunities with other young professionals (homophily), but also indicates roles played by delayed family formation, rising incomes and improvements in the quality and diversity of non-tradable services.

Keywords: Residential Choice, Consumption Amenities, Job Location, Gentrification
JEL Classification: R23

*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, Gilles Duranton, Ben Faber, Fernando Ferreira, Ed Glaeser, Joe Gyourko, Andy Haughwout, Jeffrey Lin, Hal Martin, Christopher Palmer, Jordan Rappaport, Jesse Shapiro, and Matt Turner as well as participants at the 2014 Urban Economics Association Meeting, the Duke-ERID Conference, the Stanford-SITE conference, and other seminars and conferences 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 Econonics for generous financial support.
†Haas School of Business, UC Berkeley
‡Wharton School, University of Pennsylvania and NBER

Mounting anecdotal evidence indicates that urban areas in American cities have experienced a reversal of fortunes since 2000. In this paper, we document the extent of urban revival in the U.S. and seek to explain it. We first show that urban revival is characterized by rapid growth in the young and college-educated population near the city center of almost all large US cities. We then use a rich database at a fine spatial scale to estimate a residential choice model and quantify the importance of various factors explaining the urbanization of young professionals. We find that changes in the availability of jobs and amenities rarely favored downtowns over the last decade. Instead, the initial levels of non-tradable service amenities like restaurants, bars, gyms and beauty salons have the most power to explain the recent urbanization of young professionals. In the context of our model, this empirical result indicates that changing tastes for proximity to highly urbanized non-tradable service amenities was key to the urban revival that we document. Prior work by Glaeser et al. (2004), Moretti (2012) and Diamond (2016) establishes the primacy of college-educated location choice as a determinant of economic success *across* cities since 1980. The new *within*-city trends that we uncover may have similarly far-reaching implications for urban America, as issues like rapid gentrification and rising spatial inequality capture the attention of urban policy-makers and the general public.

We present a set of novel stylized facts that document the size and scope of urban revival. Though the mid-to-late 20th century suburbanization of the aggregate population documented and explained by Glaeser et al. (2004), Baum-Snow (2007), Boustan (2010), and others continued from 2000 to 2010, the college-educated started urbanizing during this decade. This reversal was driven by a sharp change in the location decisions of the younger cohort of college-educated, and in spite of continued suburbanization of older college-educated cohorts. This young college-educated cohort, whose population grew three times as fast near city centers as it did elsewhere in the largest 50 CBSAs, urbanized so fast as to reverse the poor college-growth performance of downtowns in earlier decades.[1]

Many competing hypotheses can explain this recent urbanization of young, college-educated Americans. We test the most prominent of these hypotheses by estimating a nested-logit residential choice model. In this model, individuals choose a residential tract to live in based on tract characteristics like jobs, house prices, and amenities, as well as idiosyncratic taste parameters that are correlated across tracts within the same CBSA. We allow for the weights that individuals place on tract characteristics to vary across age-education groups and over time. Given this flexibility, our estimated model explains the differential growth of various age-education groups across tracts and, crucially, distinguishes the impact of recent changes in tract character-

[1]Some preliminary trends, notably in gateway cities like New York, Chicago, Boston and San Francisco are already apparent in the 1990s and before. 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 then. Our finding is that urban revival really emerges as a widespread phenomenon in the 2000s, and is restricted to areas smaller than the central city.

istics from that of recent changes in group-specific collective tastes for these characteristics.

Our estimates indicate that changes in well-studied tract characteristics such as proximity to jobs and public amenities play an important role in explaining changing location decisions across all tracts and all populations. We, however, are specifically interested in explaining the shift in the young-college population along the urban-suburban dimension. In this dimension, our estimates reveal a decisive role for the changing collective tastes for non-tradable service amenities. This conclusion depends both on preference parameters from our residential choice model and on the distribution of explanatory variables across the urban-suburban dimension. Downtowns of large cities are characterized by a high density of non-tradable service amenities like restaurants, which rose in quality and diversity over the last decade. Our residential choice model estimates suggest that these special characteristics have not only increasingly attracted young professionals, but attracted them more than their non college-educated and older college-educated counterparts. These changing tastes for non-tradable service amenities account for between 50 to 80% of the urbanization of the young and college-educated from 2000 to 2010. Other factors, like the slight urbanization of high-income jobs, turn out to be less important.

In our model, the shift in the tastes of the young and college-educated towards non-tradable service amenities is inferred from the estimated weights on the levels of these amenities. We assess the external validity of this structural interpretation using expenditure data from the Consumer Expenditure Survey (CEX) and trips data from the National Household Travel Survey (NHTS). The patterns in these data confirm, for example, that the young and college-educated have a stronger taste for living near restaurants than other age-education groups, and that they have experienced the most positive change in this taste over the last decade. Very similar patterns hold for expenditures on and trips to other non-tradable amenities, but not for expenditures on or trips to tradable retail.

We conclude with evidence supporting three hypotheses for the changing collective tastes of the young and college-educated for living near service amenities. First, what we interpret as a changing taste for high amenity density may be due to changing amenity composition in high density areas, towards a more diverse set of establishments that cater to the tastes of young professionals. We do find that both rising diversity and improving quality attract young professionals downtown, but these variables do not account for changing taste for service density. Second, we find that the young and college-educated have only increased their propensity to move to locations with *both* lots of non-tradable services and high shares of their own type. This suggests that evolving tastes for restaurants or bars derive from opportunities that these establishments provide to socialize and network. Finally, we show that richer and solo living young professionals devote a larger fraction of their expenditures and travel to non-tradable services. Therefore, delayed family formation in the 2000s mechanically shifted the travel and expenditure shares of the young and college-educated towards non-tradable service establishments. Rising disposable income would have a similar impact, but the Great Recession overlaps with our post-period data and obscures these trends.[2,3]

Our analysis contrasts with existing work on residential choice in the U.S. in three important ways. The inclusion of a broad set of competing explanatory factors distinguishes our work from an emerging literature on central city gentrification, which like our paper documents and explains the rising socio-economic status of downtown inhabitants over the last decade. Interestingly, Baum-Snow and Hartley (2016) also identify rising amenity values, in general, as an important driver of downtown gentrification. Edlund et al. (2015), however, focus on the taste for shorter commutes of high-skilled workers, while Ellen et al. (2017) focus on the importance of an initial central city crime drop. Our model reveals that of all factors, it is consumption amenities, in particular non-tradable services, which plays the decisive role in the urbanization of young professionals in the 2000s.

Another contribution of the paper is to measure the role of private consumption amenities in residential choice within CBSAs. Existing work in the US has either included only public amenities like schools (see, e.g., Bayer et al. 2007) or considered the role of consumption amenities within larger geographical units like CBSAs (e.g. Diamond 2016).[4] This last distinction matters, as one may move to the Bay Area primarily for the job opportunities, but choose to live in the center of San Francisco for the consumption amenities. To identify the role of consumption amenities, we build tract-level indexes of proximity to nine different amenities (e.g., restaurants, bars, food stores, apparel stores, museums, etc) and a novel Bartik-type instrument for changes in these indexes. This instrument draws from a recent IO literature on retailer location choices (e.g., Igami and Yang, 2015) that demonstrates the role that the pre-existing local business mix plays in the entry/exit decisions of firms at a highly disaggregated level (chain or SIC8). When including these instrumented consumption amenity indices along with standard controls and instruments for house prices and job growth, we find that private consumption amenities play an important role in within-CBSA location decisions independent of that played by jobs and standard amenities, such as crime, school, and transit, as well as homophily more generally.

Methodologically, our empirical framework is distinct from existing work studying within-CBSA location choices at a fine spatial scale in that we estimate a two-period model using data for all CBSAs, instead of using a cross-section of data from a small survey sample.[5] This

[2]One variant of this explanation for changing tastes is recent income growth amongst the college-educated, which will tend 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).

[3]Recent innovation in mobile technology may complement urbanized amenities, which benefits digitally savvy young professionals. This hypothesis remains speculative and our tests are very indirect.

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

[5]Baum-Snow and Hartley (2016) also estimate a tract-level residential choice mode across CBSAs, but not a

first-difference specification allows us to control for omitted variables that are constant in each location revealing important inter-temporal variation in the factors that drive location choices of different demographic groups. We also show that our results are robust to the inclusion of additional time-varying omitted factors, including changing tastes for distance to the city center and changes in the quality and diversity of consumption amenities. Further, we additionally extend our residential choice model to a residential-workplace choice model with workplace fixed-effects.[6] Adding a workplace fixed-effect convincingly frees estimates of residential characteristics from the main endogeneity problem of simultaneity with job locations. The intuition for this identification strategy is similar to that in Glaeser et al. (2001), who suggest that an increase in reverse commuting - in people who live in the central city but work in the suburbs - signals the importance of central city amenities. Our findings demonstrate that little bias result from using only residential data instead of residence/workplace data.

This paper focuses on documenting and explaining the recent 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 urban revival in complementary work.

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 amenities that we find to drive urban revival. Section 7 explores various hypotheses for these changing preferences and section 8 concludes.

1 Data

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. Geographically-consistent tract-level population counts by age and education are from the decennial censuses of 1980 to 2000 and the American Community Survey (ACS) 2008-

residential-workplace choice model. Albouy and Lue (2015) estimates 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.

[6] This analysis is relegated to robustness since the LODES residence-workplace data is not available age and education groups, but only by either broad income or broad age categories.

2012 aggregates, downloaded from the National Historical Geographic Information System (NHGIS).

The main geographical unit in our analysis is a census tract within a Core-Based Statistical Area (CBSA). We construct CBSAs using constant 2010 tract and CBSA boundaries from the Longitudinal Tract Data Base (LTDB). We define city center for 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.

To explain our stylized facts, we build datasets describing the density of workplace locations in three wage groups, the density of different types of consumption amenities (as well as measures of quality and diversity when available), and house prices in the vicinity of each census tract. The LEHD Origin-Destination Employment Statistics (LODES) for 2002 and 2011 provides counts of people in different wage groups who live and work in a given census block pair, and we use this data to characterize accessibility to job opportunities.[7] We use two datasets to measure access to consumption amenities like restaurants or apparel stores near a tract: (i) a geo-coded census of establishments in 2000 and 2010 from the National Establishment Time-Series (NETS); and (ii) travel times between these establishments and census tract centroids by foot from Google Maps.[8] We measure amenity diversity as an inverse-Herfindahl index using the most refined industry classification available in the NETS (at the SIC8 level, e.g. Korean restaurant). For a few amenity categories, we can also measure quality using ESRI's Market Potential Index (MPI), which measures 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 "All Home Index" which measures median house prices at the zip code level, which we match to 2010 tract geography using a zip code-tract crosswalk from HUD.gov.[9] In robustness checks, we use three alternative house price indices: Zillow's two bedroom index, Zillow's rental index, and a hedonic price index calculated using DataQuick data and the model from Ferreira and Gyourko (2011).

We complement these three main datasets with information on transit times, violent crime per capita, school district rankings, and natural amenities. Our tract-level measure of transit performance comes from Google Maps in 2014, and is the average travel time of a 5 mile trip from a tract centroid to a random set of 100 NETS establishments nearby. Police district-level

[7]To address confidentiality issues, the LODES is partially synthetic. We describe generation of synthetic data in appendix A, and show how aggregation of census block data at the tract level ensures that 90% of the LODES data is unaffected by synthesis.

[8]We also tried computing density indices by car and by transit, but only use indexes by foot because other modes delivered weak instruments. The popularity of the Walk Score hints at the importance of such indexes in location decisions. We also failed to obtain strong instrument for two categories that we exclude from the nine in the paper: Theater' (theater, operas, symphonies, etc) and 'Movie' (movie theater and bowling).

[9]We expand the dataset beyond tracts that Zillow covers by about 30% by spreading the house price indexes across all tracts within a tract-group, a set of three to four neighboring tracts defined in Ferreira and Gyourko (2011), who similarly estimate hedonic price indexes at the tract-group level.

data on violent crime (murder, rape, robbery, and aggravated assault) comes from the Uniform Crime Reporting (UCR) in 2000 and 2010. We obtain data on the within-state rankings of school districts in 2004 and 2010 from SchoolDigger.com.[10] There are typically multiple tracts within a police and school district. We match these areas to 2010 tract boundaries using Census shape files.[11] Data on natural amenities, like the precipitation, hilliness, and coastal proximity of each census tract are from Lee and Lin (2013).

To investigate recent trends in household formation and income growth that can explain the changing preferences of young professionals, we require 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 Consumer Expenditure Survey (CEX) and National Household Transportation Survey (NHTS).

Appendix A provides additional information on data sources and variable construction.

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 at least 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 urban revival in US cities from 2000 to 2010, most of which are new. These facts motivate our empirical analysis.

To establish these facts, 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 in the pre-period for each plot (e.g., 1980 for the 1980-1990 plot). Distance from the city center is in cumulative share of the total pre-period CBSA group population.[12] The first row of Figure 1 shows the continuing suburbanization of the general population from 1980 to 2010. Each column displays population

[10] While we believe that SchoolDigger.com is the most comprehensive database available, we have school ranking data for less than half of our CBSA's sample of tracts.

[11] 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 downtowns experiencing urban revival. In some cases like Houston and Altlanta, police districts are at the county level, so the parts of the 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% of the CBSA population.

[12] The weighting of the kernel regression by initial tract population ensures that local growth estimates are independent of tract size.

growth in a different decade, and all three decades since 1980 feature a clear positive growth gradient from the city center in CBSAs of all sizes.

Slower urban population growth does not preclude urban revival. Downtowns are generally already built up and subject to heavy housing regulations (Glaeser et al., 2006), so their increased desirability triggers rising house prices and gentrification rather than population growth. In fact, many authors argue that the *college-educated* population share is the best indicator of spatial success in recent decades (Glaeser et al. (2004), Moretti (2012)). So 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 11-50th largest cities, college-educated growth is fastest near the city center, with the negative growth gradient flattening between distance percentiles 15 and 20. This trend is specific to the 50 largest cities and, outside of three or four gateway cities like New York and San Francisco, only emerged in the most recent decade. This explains why early claims of urban revival were not backed up by systematic evidence. Defining downtowns as the set of tracts closest to the city center accounting for 5% of a CBSA's population, we find that the college-educated population grew faster downtown than elsewhere in 28 out of the 50 largest CBSAs in the 2000s, compared to only 10 CBSAs in the 1990s and 9 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.[13]

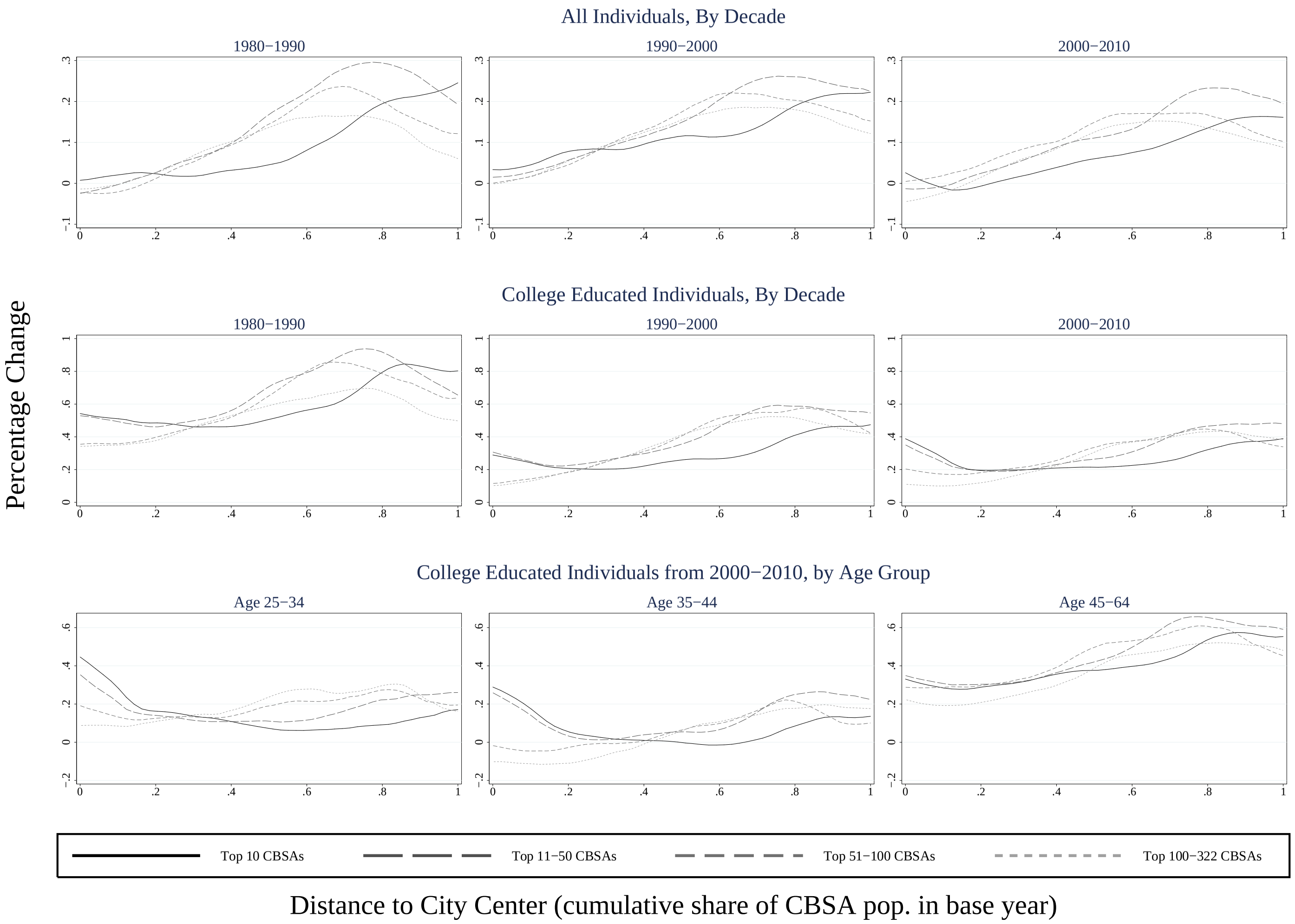
Row 3 further refines our investigation and breaks down college-educated growth in the 2000s by age group. This breakdown is relevant because we expect the residential preferences of the younger generation to differ from that of older Americans. Moreover, the popular press emphasizes the urbanization of both young people and retiring baby-boomers.[14] 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, with the younger group experiencing the sharpest gradient with nearly 40% growth near city centers relative to around 15% growth past the 20th distance percentile. This trend is also most pronounced in the 50 largest cities. Contrary to claims that retiring baby-boomers are urbanizing, the older 45-64 and 65+ year-old (not shown) college-educated groups are still suburbanizing.[15]

[13] In an online Appendix E, we propose different ways of tabulating the data shown in Figure 1. We compare downtown growth - using various downtown definitions - to that in the rest of the CBSA to document the scope of urban revival across cities.

[14] A recent industry report by CEO for Cities (Cortright, 2014) also uses 2000 Census data and 2008-2012 ACS data, and considering only the 51 largest MSAs and the the 25-34 college-educated population, shows that they are growing faster downtowns (defined as 3-mile radius from the city center.)

[15] The popular press also emphasizes the urbanization of "millennials", 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.

Figure 1: Population Growth at Various Distance from the City Center in All CBSAs



Notes: Data from decennial censuses 1980–2000 and ACS 2008–2012. Each figure shows 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.

These local trends are sharp enough to have aggregate impact. 150 million Americans live in the 50 largest CBSAs. In these large cities, downtowns accounting for five percent of the population experienced 24% of the total increase in the young college-educated population and 11.5% of the middle-aged college-educated population between 2000 and 2010. Strikingly, the young and college-educated urbanized in 23 of the 25 largest CBSAs (see online Appendix E for tabulation). The exceptions are Riverside, 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.[16]

Appendix Table A.7 shows that our stylized facts are robust to using a different city center definition (Central Business Districts from the 1982 census of retail trade), income 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. Other current work on central city gentrification support our findings, with Baum-Snow and Hartley (2016) showing that downtowns are becoming richer, more educated and more white, and Edlund et al. (2015) showing that they become more educated and expensive.

The objective of the rest of this paper is to find the factor(s) driving the population growth gradients by age-education group from 2000 to 2010 of Figure 1, with a strong focus on explaining the remarkable growth of the young and college-educated near city centers.

3 Residential Choice Model

To explain the changing residential location choices of different demographic groups, we specify a workhorse discrete choice model (augmented in section 5.4 to study the joint workplace-residential location decision). 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.[17]

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

[17] Our model differs from Bayer et al. (2007)'s important application to residential choice of McFadden (1973) and McFadden (1978)'s random utility model. Unlike Bayer et al. (2007), we derive our indirect utility from a primitive Cobb-Douglas consumer optimization problem, and we add a time dimension. Instead of estimating unobserved neighborhood heterogeneity using the technique developed in Berry et al. (1995), we include many additional neighborhood characteristics into the model and are able to derive simpler linear regressions.

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 wage (net of commuting costs) w_{jct}^{id} between units of housing H, private consumption amenities A, and an freely-traded outside good Z in order to maximize the following Cobb-Douglas utility function:

$$U_{jct}^{id} = \alpha_{jct}^{id} H^{\beta_{Ht}^d} A^{\beta_{At}^d} Z^{\beta_{Zt}^d}$$

subject to a budget constraint:

$$w_{jct}^d = p_{Hjct} H + p_{Ajct} A + Z,$$

where p_{Hjct} is the price of housing, p_{Ajct} is a price index for consumption amenities that varies with transport costs to these amenities, and α_{jct}^{id} 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:

$$\alpha_{jct}^{id} = a_{jct}^{\beta_{at}^d} \exp\left(\mu_{jc}^d + \xi_{jct}^d\right) \exp\left(\psi_{ct}^{id}(\sigma^d) + (1 - \sigma^d)\varepsilon_{jct}^{id}\right).$$

The group-specific tastes for each tract is represented by the sum of two terms: a time-invariant component μ_{jc}^d, and a time-varying tract-specific component, ξ_{jct}^d. The individual-specific tastes take a nested-logit structure with tracts nested by CBSA. Tract taste shocks, ε_{jct}^{id}, are independent draws from the extreme value distribution, while CBSA taste shocks, $\psi_{ct}^{id}(\sigma^d)$, are independent draws from the unique distribution such that $\psi_{ct}^{id} - (1 - \sigma^d)\varepsilon_{jct}^{id}$ is also an extreme value random variable. The parameter $0 \leq \sigma^d < 1$ governs the within-group correlation in the error term $\psi_{ct}^{id}(\sigma^d) + (1 - \sigma^d)\varepsilon_{jct}^{id}$. As σ^d approaches zero, the model collapses to a standard logit model.

After solving the Cobb-Douglas utility maximization problem, each person i chooses its residential tract j to maximize its indirect utility:

$$\max_j V_{jct}^{id} = \beta_{Wt}^d \ln w_{jct}^d(\tau) - \beta_{Ht}^d \ln p_{Hjct} - \beta_{At}^d \ln p_{Ajct} + \beta_{at}^d \ln a_{jct} + \mu_{jc}^d + \xi_{jct}^d + \psi_{ct}^{id}(\sigma^d) + (1 - \sigma^d)\varepsilon_{jct}^{id} \tag{1}$$

Note that $\beta_W = \beta_H + \beta_Z + \beta_A$ and that the wage w is net of commute cost τ. Our empirical implementation of $w(\tau)$ is a vector of time-varying accessibility to jobs in three different wage brackets. So we write $w_{jct}^d(\tau) = w_{jct}(\tau) + \xi_{Wjct}^d$, where $w_{jct}^d(\tau)$ denotes the wage of jobs available to group d from tract j net of commute costs, $w_{jct}(\tau)$ is the observed component of this wage and ξ_{Wjct}^d is unobserved and group-specific.

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 $\bar{j}$:[18]

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

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 private amenity terms, $\beta^d_{At} \ln (1/p_{Ajct}) + \beta^d_{at} \ln a_{jct}$. We drop the dependence of the wage w on commute cost τ from the notation.

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

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

where $\Delta X = X_{2010} - X_{2000}$ for both variables and coefficients.[19] Note that unobserved time-invariant tract characteristics like nice weather or historical architecture cancel out in first-difference. The error term is $\Delta\tilde{\xi}^d_{jc} + \Delta\tilde{\xi}^d_{W,jc} + \epsilon^d_{jc}$, i.e., the sum of any unobserved changes in the perceived quality of a residential location, unobserved changes in the group-specific wage premium of jobs available from a given tract, 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 preference levels of demographic group d (i.e., $\beta^d_{A,2010}$), while coefficients on initial levels of characteristics (e.g., $\tilde{\mathbf{A}}_{\mathbf{j,2000}}$) capture changes in the collective 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 can be driven by either changes in the characteristics of locations or changes in preferences for locations. The young and college-educated might be moving downtown either because characteristics of downtown tracts changed

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

[19]Note that $\beta^d_{A,2010} X_{2010} - \beta^d_{A,2000} X_{2000} = \beta^d_{A,2010} (X_{2010} - X_{2000}) + (\beta^d_{A,2010} - \beta^d_{A,2000}) X_{2000} = \beta^d_{A,2010} \Delta X + \Delta\beta^d_A X_{2000}$

in ways correlated with their collective tastes (i.e., $Corr(\Delta\tilde{X}_{jc}, \beta^d_{X,2010}) > 0$) or because their tastes 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 level and changes in the characteristics of tracts at different distance from the city center, and 2) estimates of the parameters reflecting both the level and change in the collective tastes of the young and college-educated for these characteristics. We now present data summarizing the level 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 factors in explaining the urbanization of young professionals.

4.1 Recent spatial trends in jobs, amenities and house prices

Figures 2 and 3 shows how key tract characteristics vary with distance to city centers. Panel A of each figure shows the kernel density plot of the 2000 logged level of a variable and Panel B shows the kernel density of the log change from 2000 to 2010, with kernel weights based on 2000 tract share of young, college educated individual. The data presented includes all tracts in our estimation sample for which a variable is available. Variable construction is detailed in Appendix B.

Figure 2 presents gradients for jobs in column 1, house prices in column 2, and public amenities (school and crime) in column 3. Proximity to jobs is an inverse distance-weighted average of the number of jobs in tracts surrounding each residential tract in 2002 and in 2011, computed using the LODES data for the three nominal wage groups: high-income jobs paying more than \$3333 per month, middle-income job (\$1000-\$3333) and low-income jobs (<\$1000).[20] Both high wage (hyphenated blue) and low wage (non-hyphenated red) job density are higher near the city center. However, only high wage jobs have grown (slightly) faster near the city center over the last decade. This urbanization of high wage jobs can therefore explain some of the young and college-educated's urbanization, if our model estimates in the next section indeed show that they are attracted to such jobs.

Column 2 shows the gradients for house prices, calculated using zipcode-level data from Zillow.com matched to census tracts.[21] Houses are more expensive away from the city center in 2000, but less so when focusing on two-bedroom homes using the "Two Bedroom Index" (in non-hyphenated red) which better controls for size than the "All Home Index" (in hyphenated blue). House price growth displays a strongly negative gradient from the city center, consistent

[20]These three groups correspond roughly to income terciles in 2002.

[21]In our main regression specification we use the Zillow Home Value Index (that we refer to as "All homes") from 2000 and 2010, which measures the median value of all (non-distressed) properties. The index and methodology are available at: http://www.zillow.com/research/data/.

with the urban revival that we document in this paper.[22]

Column 3 shows that public amenity levels are lower in urban areas, with more violent crime per capita (in non-hyphenated red) and lower ranked schools (in hyphenated-blue, not logged). Urban schools have dropped even further in state district rankings from 2004 to 2010. Violent crime rates are decreasing everywhere as expected, but surprisingly the gradient from the city center is not negative from 2000 to 2010.

Figure 3 presents gradients for two representative consumption amenities: restaurants in hyphenated-blue and food stores in non-hyphenated red. Column 1 shows amenity density, column 2 shows quality and column 3 shows diversity. These indexes are based on the the CES price index methodology in Couture (2013), implemented using the NETS database of all U.S. establishments in 2000 and 2010, travel times from Google, and survey data from ESRI.[23] The density of both restaurants and food stores is highest near the city center, but grew faster in the suburbs, past the median population-weighted distance from the city center. Rising amenity density is therefore unlikely to explain urban revival. We find similar density patterns for our full set of nine consumption amenities: other non-tradable services (bars, gyms and personal services), other stores (apparel and general merchandise), and activity establishments (museums, galleries, and libraries though not for amusement parks and golf courses, which are suburbanized). Unlike density, restaurant and food quality and diversity have increased faster downtown, and moreso for restaurants.

4.2 Estimation

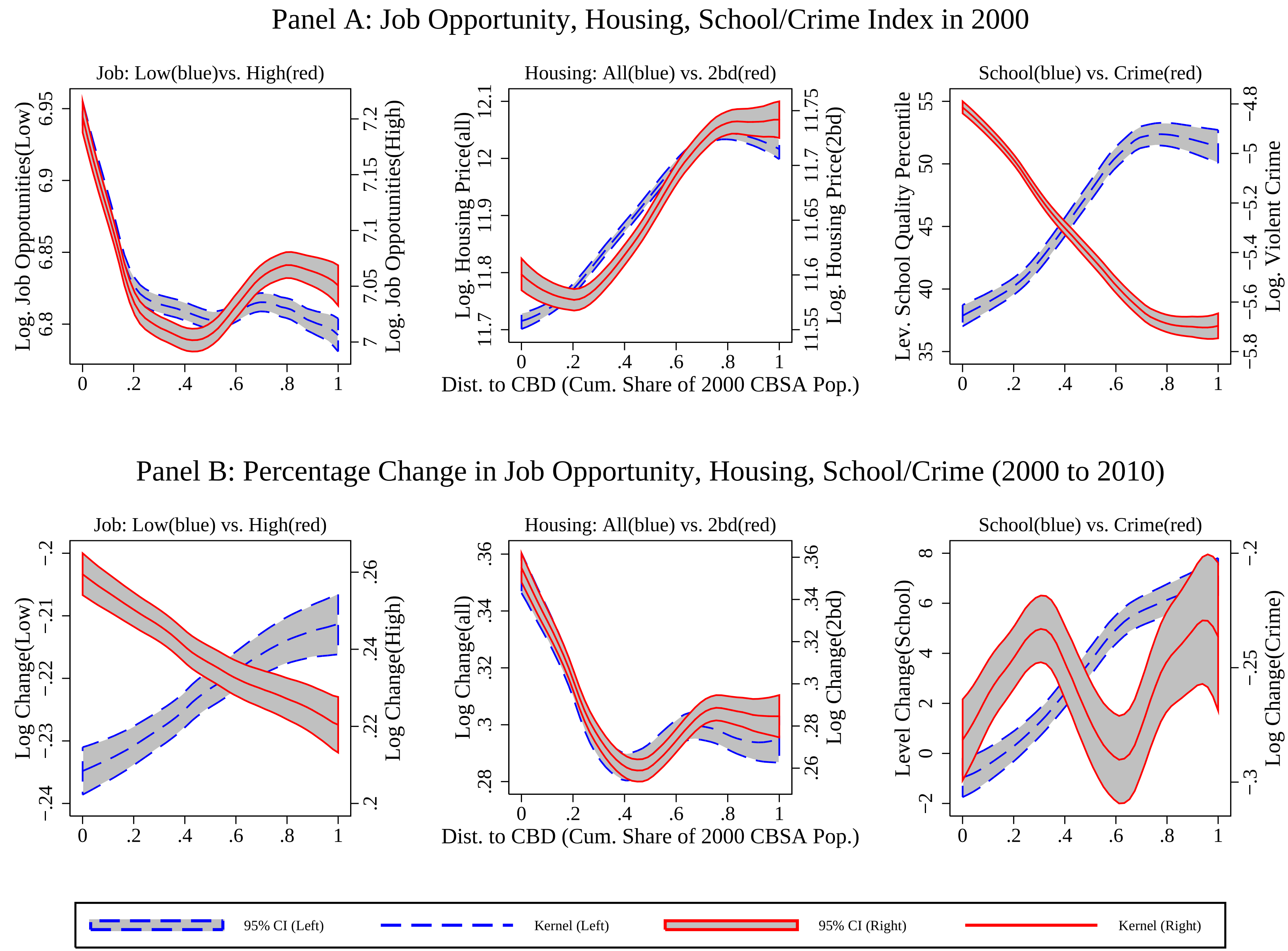
Our main specification will include house price and job opportunity indexes, private amenity densities, as well as own-group tract shares and population density to capture changes in within-group homophily and other unobserved endogenous amenities.[24] In robustness checks, we control for public amenities, such as school quality, crime and transit times. We exclude public amenities from our main specification because we have no instrument for these variables and only have school and crime data for a subset of CBSAs. We also exclude amenity quality and diversity from our main specification, because quality is only available for a subset of establishments and we have no instrument for diversity (our instrument for amenities themselves relies on a measure of local establishment diversity).

[22]Generally the spatial gradients in 2000 and 2010 are similar and the trends in large CBSAs are visible in all CBSAs. House prices are a notable exception: in the 10 largest CBSAs, the spatial house price gradient *reverses* from 2000 to 2010.

[23]The exact methodology is described in Appendix B.2.

[24]Controlling for additional 2000 demographic shares – such as share college 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.

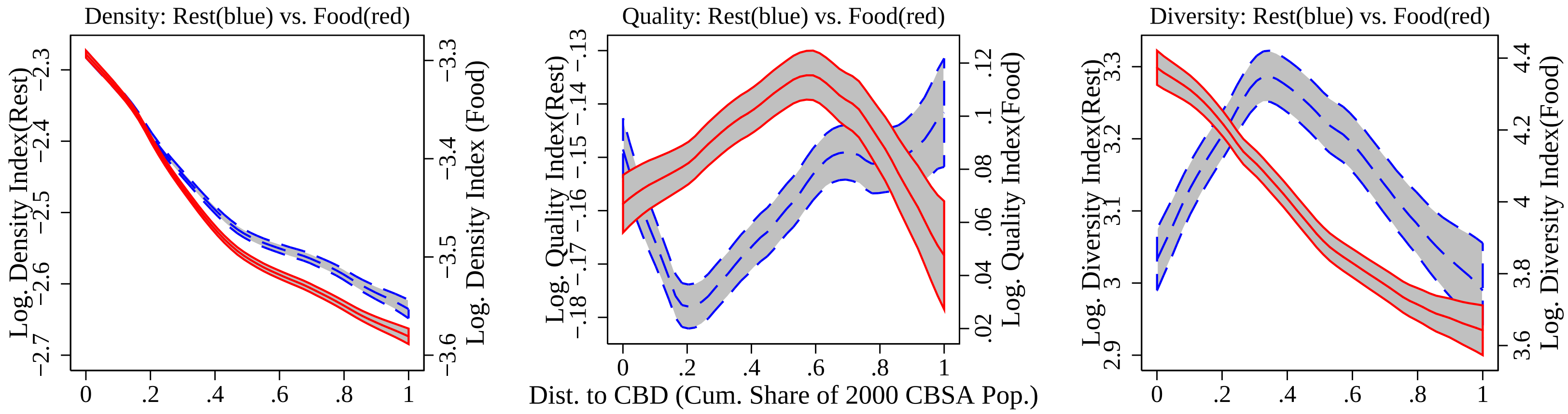
Figure 2: Tract Characteristic Gradients: Job Opportunities, Housing, and School/Crime



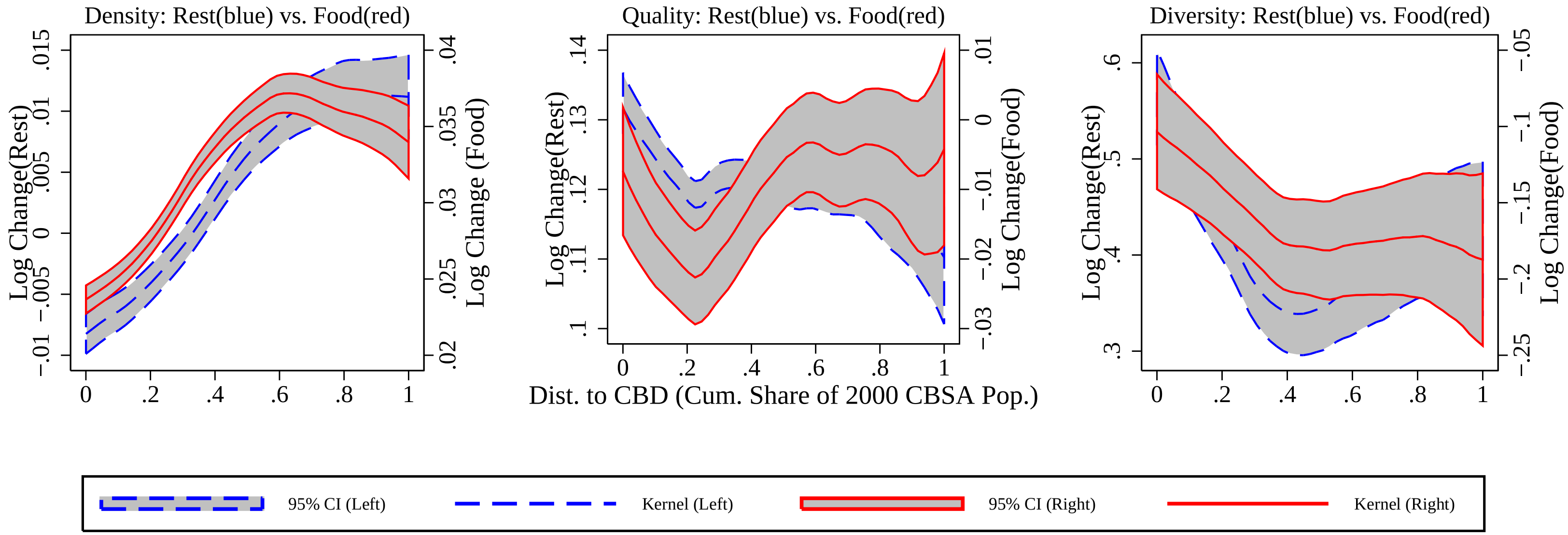
Notes: Panel A shows a non-parametric kernel fit of the log of job opportunities, log house price, log per capita violent crime and school ranking in 2000 plotted against the population-weighted distance from the city center for all tracts in our estimation sample. Panel B shows a similar kernel of percent 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 from UCR and school data from SchoolDigger.com (see Appendix A for detail on variable contructions).

Figure 3: Tract Characteristic Gradients: Restaurant and Food Density, Quality, Diversity Index

Panel A: Tract Composition Gradients: Restaurant and Food Density, Quality, Diversity Index



Panel B: Percentage Change in Density, Quality, Diversity Index (2000 to 2010)



Notes: Panel A shows anon-parametric kernel fit of the log of amenity density, quality, and diversity indices in 2000 for restaurants and food stores, plotted against the 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 indices . See Appendix B for details on amenity index construction.

4.2.1 Identification

There are challenges to identifying the effect of neighborhood characteristics on residential choice. The first-difference regression controls for time-invariant tract characteristics that could be correlated with our regressors, but for neither unobserved changes in tract quality ($\Delta\tilde{\xi}^d_{jc}$) or local wage premia ($\Delta\tilde{\xi}^d_{W,jc}$), each of which could be correlated with our regressors. In robustness checks, we therefore test for omitted variable bias on our level coefficients with an array of controls, including changes in own-group shares in nearby tracts, to distinguish the part of our level coefficients due to endogenous amenities from that directly due to changes in collective tastes. Neither first-differencing nor adding controls, however, resolves reverse causality, which affects variables in changes. For instance, an influx of young professionals in response to unobserved shocks to tract quality or nearby wages may attract amenities and jobs, and raise house prices. We describe instruments for these variables below, devoting more time to the most novel of these, the instrument for the change in amenity density.

In section 5.4 we present results based on an alternative identification strategy using commute data. This data allows us to introduce a workplace fixed-effect in a workplace-residential choice model, thereby convincingly isolating the effect of changes in residential characteristics from that of changes in job location.

Instruments for Consumption Amenity Density Indexes To design an instrument for the change in 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 category variation in national growth rates across finely-defined industries or chains in conjuction with spatial variation in business expansion paths, relying on the fact that these paths depend, in part, on the pre-existing business landscape. This strategy 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 and preemption concerns in determining retailer entry (Igami and Yang, 2015; Toivanen and Waterson, 2005).

The instrument's computation proceeds in two steps. First, we model business expansion paths by regressing SIC8-level establishment entry from 2000 to 2010 in each tract on variables capturing the pre-existing commercial environment in 2000 at different distances from the tract centroid.[25] Then, we predict net establishment entry in each tract by summing the fitted values

[25]We expect entry to decrease with the concentration of establishments in very close proximity offering similar services, 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 SI8 or chain, since these indicate proximity to the chain's upstream suppliers or distribution centers and some pre-existing market knowledge. In addition to within-SIC 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

of these regressions over all SIC8 codes in an amenity category, and aggregate to obtain the predicted change in each 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, and our dependent variable is $n_{j10}^{sic8} - n_{j00}^{sic8}$. We model entry and exit as a function of the business environment, more precisely of $n_{j00,dist}^{sic8}$, $n_{j00,dist}^{sic6|8}$ and $n_{j00,dist}^{sic4|6}$, which represent be 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:

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

The estimation results indicate that competition and cannibalization concerns are important predictors of entry and exit.[26] In 93% 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% of cases and negative and significant in less than 10%.

Our amenity quality indexes are based only on establishments that are part of a 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, in Table A.3 of appendix B.2, highlight the strength of within-chain spatial economies of scale. That is, the effect of proximity to same chain reverts from a negative to a positive sign as distance increases, implying 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 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" 2010 establishments, we compute a "predicted" amenity density indices 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.[27] A valid instrument must

as natural or regulatory supply constraints.

[26] See Table A.2 in appendix B.2 for aggregate results on the predictors of entry and exit across all 1078 SIC8 codes used to define our consumption amenity indexes.

[27] See Table 2 of appendix C.

also be exogenous, i.e., uncorrelated with the error term in equation 3 conditional on other regressors. If only supply factors drive national expansion strategies (e.g., within-chain spatial economies, sharing suppliers) then this exclusion restriction holds. The instrument is robust to changes in local demand because entry predictions draw from a national sample. The exclusion restriction would however be violated by changes in national demand for specific amenity types that, in 2000, were systematically absent or present from tracts that experienced unobserved demand shocks ($\Delta\tilde{\xi}^d_{jc}$ or $\Delta\tilde{\xi}^d_{W,jc}$) in the following decade. This would require that the locations characterized by a specific amenity mix in 2000 experienced correlated location shocks between 2000 and 2010 and that these location shocks were, in turn, correlated with the location-specific growth predicted by the national growth rates of that mix of amenities along with the sensitivity of those establishment types to cannibalization, competition, agglomeration, and within-chain economies. The latter correlation, in particular, seems unlikely: the stronger cannibalization concern of Korean relative to Pizza restaurants is unlikely correlated to changes in relative national demand for these restaurant types.

Instruments for Housing Prices To overcome the endogeneity of house price changes and levels, we exploit the correlation between housing prices and exogenous natural amenities identified by Lee and Lin (2013). We expect natural features (oceans, mountains, etc.) to act like anchors imposing supply constraints on land, whereby 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 natural amenity measures 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 for flooding, and, finally, the logs of the annual precipitation, July maximum, and January minimum temperatures in the tract averaged over 1971 and 2000.[28] As in Bayer et al. (2007), our instrument for tract j uses natural features of tracts one to three miles away, controlling for the average natural features of tracts within one mile. The key exclusion restriction is that natural features further than one mile away from a tract do not impact demand for living in that tract, conditional on the natural features within one mile.

As an additional instrument for housing prices (and for local demographic shares), we include historical tract-level 1970 population shares, by age and education group (college/non-college).

[28] 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.

In a robustness check, we exploit the Cobb-Douglas preference structure to simply difference out the group-specific CEX housing expenditure share 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 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 preferences.

Instruments for Job Opportunity Index We use standard Bartik instruments for changes in the job opportunity indexes. The LODES data includes jobs in our three wage groups by 20 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 the national industry growth.

Instruments for change in the share of type d individual within CBSA c who live in tract j Instrumenting the nested-logit share $\Delta s^d_{j|c}$ requires exogenous factors affecting the attractiveness of tract j relative to all other tracts in its CBSA 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 OLS regression results for the nested-logit model in equation 3 using, for the sake of parsimony, two representative amenity indexes - one for restaurants and one for food stores.[29] Table 2 follows with the same specification in IV (see Table A.5 in Appendix C for detailed first-stage statistics for all instrumented variables).[30] Panel A of both tables shows coefficients for the college-educated by three age groups. Panel B shows the same coefficients for the non-college educated age groups. The first two columns of Panel A display parameter estimates explaining changes in the share of 25-34 year olds (i.e., young) college-educated in

[29]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.

[30]The reduced-form first-stage statistics all easily reject that the instruments are irrelevant. The SW conditional F-statistics - Sanderson and Windmeijer (2016) - 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.

a tract. Column 1 shows coefficients for variables in first-difference (i.e., the change in that variable from 2000 to 2010) and column 2 shows coefficients for the 2000 levels of the same variables. Most coefficients are significant at the 1% level. The OLS coefficients are generally of the same sign as their IV counterparts but often smaller in magnitude, likely as the result of attenuation bias. Recall that we are interested in the relative importance of different factors in explaining urban revival *within* a given specification. So, while striking, these magnitude differences do not impact our main results, which are invariant to scaling all coefficients up or down and, in particular, robust to whether we use the IV or OLS coefficient estimates.[31]

The structural interpretation of the coefficient on a first-difference variable is that of a preference parameter in 2010, $\beta^d_{X,2010}$. A positive sign denotes attraction to this tract characteristic. The coefficient on a 2000 level variable has an interpretation as a change in preference from 2000 to 2010, $\Delta\beta^d_X$. We adopt this interpretation in our discussion of the results. In section 6, we use data on changes in expenditure and trip shares from 2000 to 2010 by age-education groups to provide external validity for this interpretation.

Most coefficients in Table 2 have the expected sign. Considering coefficients on the variables in first-difference, all three college-educated groups have a preference for proximity to high-income jobs and all six groups have a preference for proximity to restaurants, 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).[32]

Turning to the coefficients on variables in levels highlights a key result of the paper, which we show next, on the importance of the shift in collective tastes of the young and college-educated for non-tradable service amenities. This shift in tastes is reflected in the coefficient on the 2000 level of restaurant density, and is signficantly larger for the young and college-educated than for other groups. It is important to demonstrate that this restaurant result generalizes to other non-tradable services. Column 1 of Table 3 adds bars and apparel stores to the set of amenities in our main specification, and column 2 includes all nine consumption amenities. 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 retail stores and activities like museums and amusement parks.[33]

[31] All of the presented coefficients are standardized. So, 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.

[32] 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 excludes housing, as in Baum-Snow and Hartley (2016).

[33] The four non-tradable service amenity density indexes are highly correlated and the restaurant coefficients decrease in size when we add other amenities. This is consistent with endogenous consumption amenities and

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 Opportunities (LI)	-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 Opportunities (MI)	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 (HI)	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)
Restaurants	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 Stores	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 Opportunities (LI)	-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 Opportunities (MI)	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 (HI)	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)
Restaurants	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 Stores	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% significance level; ** – 5% significance level; ***–1% significance level. 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

	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 Opportunities (LI)	-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 Opportunities (MI)	-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 (HI)	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)
Restaurants	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 Stores	-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

	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 Opportunities (LI)	-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 Opportunities (MI)	-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 (HI)	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)
Restaurants	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 Stores	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% significance level; ** – 5% significance level; ***–1% significance level. The change in house prices, level of local demographic share, change in amenity density, change in job opportunities, and change in the share of type d individuals within CBSA c who live in tract j are considered endogenous variables and instrumented in first stage regressions. Each regression is weighted by the share of type d in a tract j in year 2000.

These stark preference patterns do not hold for other groups. OLS regressions equivalent to the IV specification of Table 3 show similar results.

One may worry that positive coefficients on the level of urbanized service amenities originate from fortuitous correlations with the urbanization of the young and college-educated. It is worth noting that other amenity types (e.g., stores) are also urbanized but *do not* have positive level coefficients for young professionals.[34] Further, our coefficients on non-tradable services are robust to including a direct control for tract distance to the city center in column 3 of Table 3 .

Table 3: Nested-Logit Residential Location Choice IV Regression Results Including More Amenities and Distance to City Center

College Educated, 25-34 years old

	4 amenities		9 amenities		With Dist. to CBD	
	Change	Level	Change	Level	Change	Level
Variable	[1]	[2]	[3]	[4]	[5]	[6]
Job Opportunities (LI)	-0.138***	-0.034***	-0.131***	-0.037***	-0.120***	-0.031***
Job Opportunities (MI)	-0.057***	0.036***	-0.012	0.053***	-0.001	0.047***
Job Opportunities (HI)	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***
Golf and parks			-0.014***	-0.006**	-0.010*	-0.003
Gym and sports			0.025**	0.029**	0.030***	0.037**
Restaurants	0.323***	0.329***	0.133***	0.129***	0.144***	0.138***
Bars	0.035***	0.018*	0.046***	0.046***	0.042***	0.039***
Personal Services			0.080***	0.062***	0.095***	0.083***
Merchandise Stores			-0.026***	-0.025***	-0.018**	-0.013
Food Stores	-0.049***	-0.198***	-0.057***	-0.145***	-0.045***	-0.128***
Apparel Stores	0.005	-0.002	0.002	0.012	-0.026**	-0.022
Dist2CBD						-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% significance level; ** – 5% significance level; ***–1% significance level. The change in house prices, level of local demographic share, change in amenity density, change in job opportunities, and change in the share of type d individuals within CBSA c who live in tract j are considered endogenous variables and instrumented in first stage regressions. Each regression is weighted by the share of type d in tract j in year 2000.

dynamics that amplify any exogenous increase in one amenity. We explore issues of endogenous amenities and homophily further in subsection 5.3.

[34]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 urban-suburban distribution of each variable from section 4.1 to identify the main factors explaining urban revival. Any variable in our model contributes to the urbanization of the young and college-educated if (i) its gradient from the city center is negative (positive), and (ii) it has a positive (negative) regression coefficient for young college-educated individuals. To visualize these contributions, we again use kernel plots, which illustrate the contribution of each variable to log changes in the share of a given age-education group living at a given distance from the city center.

We start with our regression equation 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 population-weighted distance of this tract from the city center. Figure 4 shows these "contribution plots" for all explanation variables and all age-education groups, computed with coefficients from the specification in Table 2. As an example of how to interpret these plots, consider the contribution of the variable "change in high income jobs." Recall from Figure 2 that this variable shows a mild negative gradient from the city center. In the contribution plot, this gradient determines the "shape" of the contribution, and the gradient combined with the regression coefficient determines its scale and slope. In this case, the change in high-income jobs has a large standardized coefficient, so it is an important determinant of location choice for the young and college-educated. However, this variable's contribution is not nearly as important *along the urban/suburban dimension*, because high income jobs have only increased slightly faster near city centers relative to outlying areas. To make comparisons of contribution across variables easier, we normalize the contribution of each variable at the edge of a CBSA to 0. As a result, the intercept of each plot with the city center provides a ranking of each variable according to its contribution to urbanizing a given group.

Figure 4 shows the key result of the paper: The 2000 level of restaurants is the most important contributor to the urbanization of the young and college-educated, and this contribution is larger for the young and college-educated than for any other groups. We think of restaurants as representative of non-tradable services more generally, and we replicate this exercise including all nine 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, we conclude that the main contributing factor to the rising share of young professionals near city centers is an increasing collective tastes for urbanized non-tradable service amenities.

Table 4 quantifies these results. Column 1 provides the rank of the contribution of non-tradable service levels in urbanizing the young and college-educated across 11 categories of variables (jobs, house prices, non-tradable services, retail stores and activities in both levels and changes, and demographic shares in levels). The first panel shows that non-tradable service levels rank first in both IV and OLS in our base specification with 9 consumption amenities. Column 2 provides the non-tradable service levels' share of the total contribution to urbanizing the young and college-educated in the model. For the IV specification this share is 83%, with other variables making a positive contribution accounting for the remaining 17%.[35] A CBSA fixed-effect instead of a nested-logit specification delivers the same conclusion. In OLS this share is 50%. Other rows of Table 4 show the robustness of these conclusions for different specifications discussed later in the paper. In Appendix C we provide a similar ranking for all specifications in the paper, in both IV and in OLS. We also show how the size of the young professionals' coefficients on non-tradable services compares to that of other age-education groups.

4.3.1 Can non-tradable services levels explain urban revival in large cities?

Our stylized facts document that the urbanization of the young and college-educated is primarily a large city phenomenon. Figure 5 shows that non-tradable service levels can explain this. The top plot in Column 1 shows the contribution of 2000 restaurant level on urbanizing young professionals for three groups of CBSAs ranked by population: top 10 , top 11-50, and top 50-100, all other CBSAs. We find that non-tradable service amenity levels indeed provide a stronger urbanizing push in large CBSAs, because they have the highest 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 we have only limited data, and therefore choose not to include in our main analysis.

[35] These numbers capture the relative importance of different factors along the urban-suburban dimension in the model. In absolute terms, the contribution of non-tradable service levels to young-college growth near city centers is 1.8 times *larger* than the actual growth. However, other factors are also pushing against this urbanization. Aggregating over all variables' contributions, the model correctly fits the urbanization of the young and college-educated and to a lesser extant of the middle-age and college-educated, and the suburbanization of every other group. Fitted values from the model, however, generate less urbanization than what actually happened.

Figure 4: Factors Contributing to Predicted Urban-Suburban Tract Composition Change Gradients

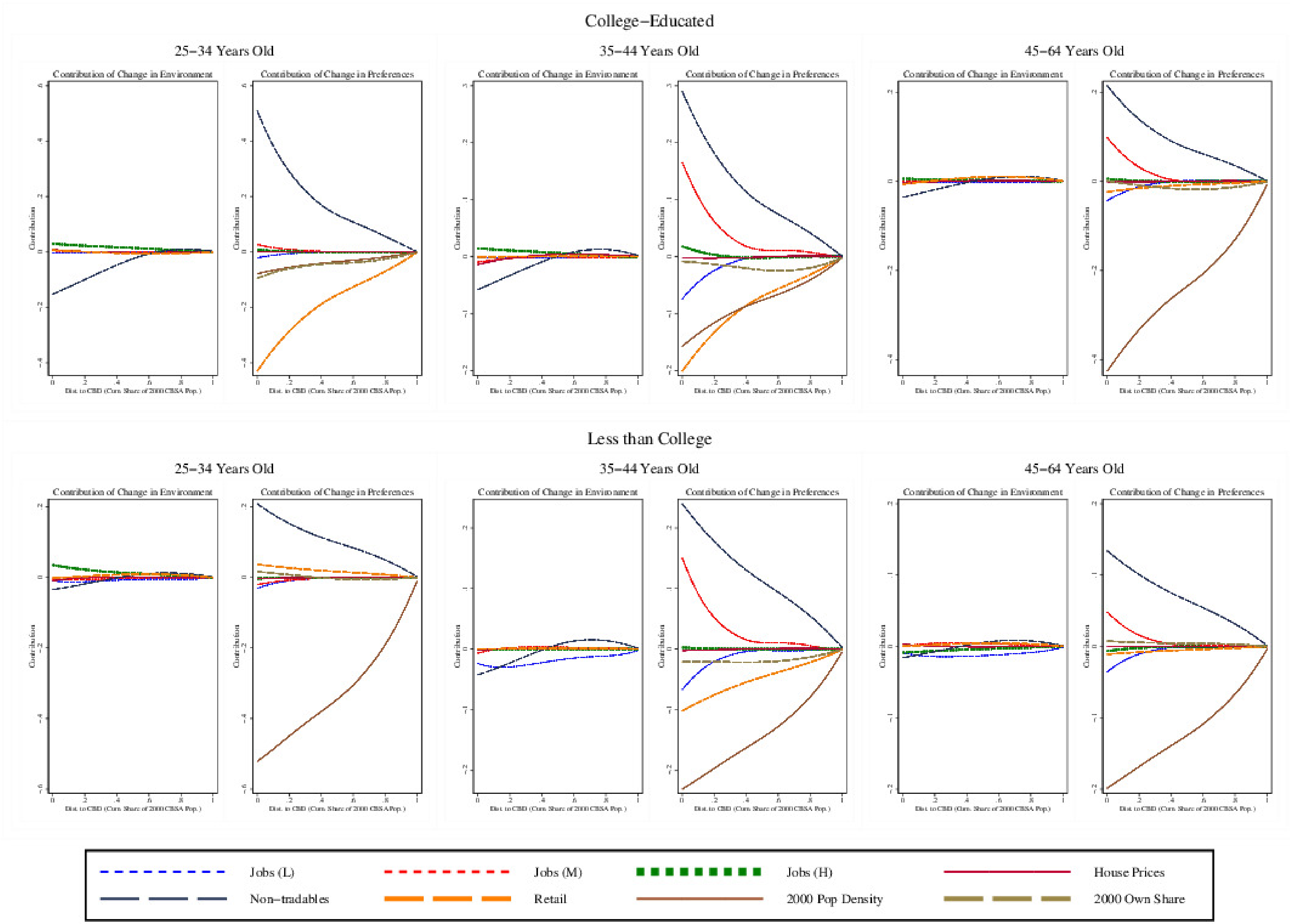
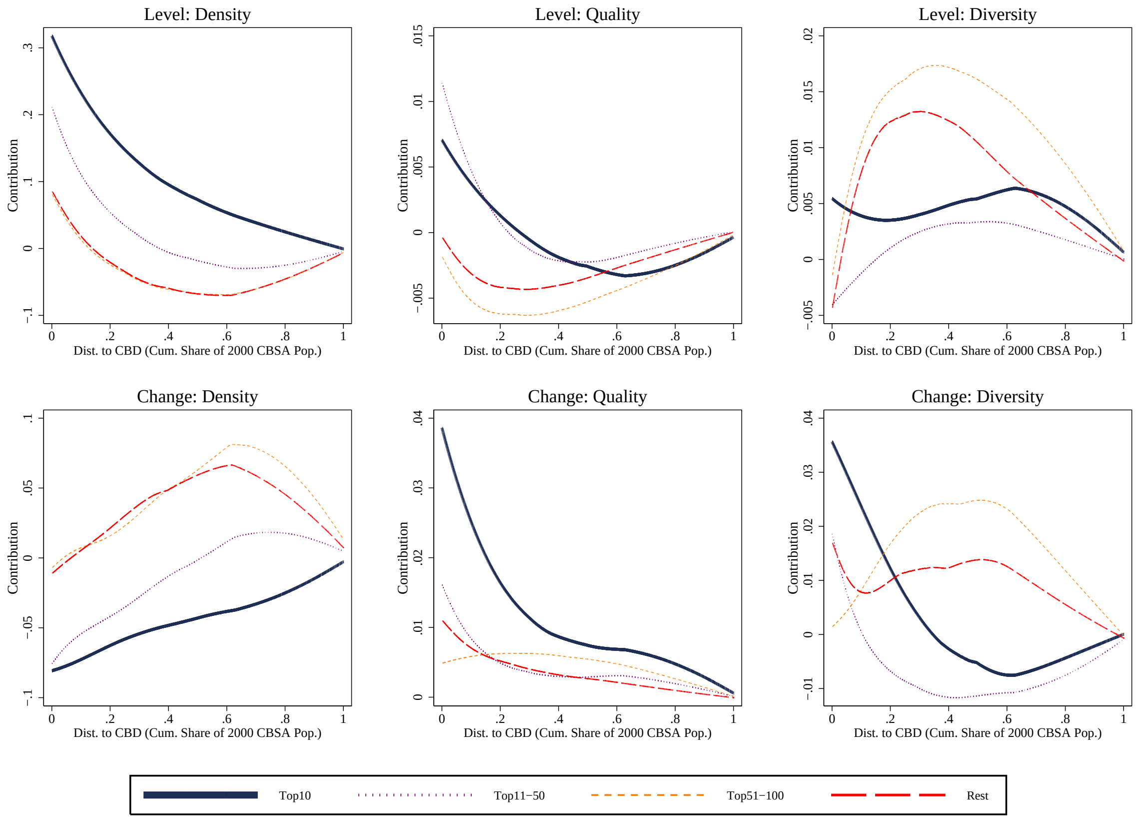


Table 4: Ranking of Restaurant and Non-Tradable Services Urbanizing Contribution Across Variables and Across Age-Education Groups

Specification	Contribution of 2010 Non-Tradable Variable(s) Rank	As a Portion of Urbanization Trend
Panel A: Base Specification (9 amenities)		
Base Specification (nested with all instruments)	1 (+)	82.56%
Nested-OLS (except for within CBSA share)	1 (+)	50.21%
Non-nested with CBSA Fixed Effects	1 (+)	75.47%
Panel B: Base Specification with Additional Controls		
Distance to CBD	1 (+)	72.69%
School, Crime and Transit	1 (+)	82.34%
Homophily Control	1 (+)	49.31%
Panel C: Base Specification with Alternative Housing Index		
Ferreira/Gyourko Hedonic Index	1 (+)	72.38%
Zillow 2 Bedroom Index	1 (+)	83.45%
Housing on the LHS	1 (+)	78.35%
Panel D: Specifications with Single Non-Tradable and Single Retail Variable		
Base Specification (2 Amenities)		
Restaurant Density	1 (+)	88.41%
Base Specification with Amenity Quality and Diversity:		
Restaurant Density	1 (+)	64.70%
Restaurant Quality+Diversity	5 (+)	2.31%

Figure 5: Restaurant Density, Quality, Diversity Contribution to Tract Composition Change Gradients by CBSAs Size



5.1 Crime, school and transit

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. Much 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 main regression specification adding controls for the change and levels in local school district rankings and per capita violent crime, and for 2014 levels of transit time of a five mile trip.[36] Panel B in Table 4 documents that the level of non-tradable services is still the most important determinant of urbanization for the young and college-educated in this specification. 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 crime is negative and significant for all groups - possibly because of reverse causality - *except* for the young and college-educated, who have a coefficient near 0 indicating little aversion to crime, especially relative to other groups. We find no evidence that young professionals move to areas with initially 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 Garr, 2010) 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 run the specification in Table 5 but using crime level 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.[37] Finally, as noted in Edlund et al. (2015), there is anecdotal evidence that central locations in large European cities are also

[36] Full regression results are reported in Table A.6. Per capita violent crime in a tract is the log of the total number of murder, rape, robbery, and aggravated assault incidents per capita from the UCR database as described in Appendix A. SchoolDigger.com compiles test scores and provides a ranking of each school district within each US 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. To control for transit performance, we use data on simulated transit trips from each tract centroid to a random sample of NETS establishments at various distances from that centroid, in various directions. We measure transit performance as the average time of a 5 mile transit trip starting from a tract's centroid, using the fitted value from a function of transit time on the distance from NETS establishments. Adding these variables reduces our sample size by two thirds.

[37] 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 downtowns smaller than central cities, motivated by the stylized facts in section 2. Second, our approach infers a general aversion to violent crime regardless of the area, rather than assigning specific aversion to "central city" crime to different groups.

experiencing rising demand from the young and college-educated, despite not having had the high rates and subsequent decline in crime that US central cities experienced. Combined, these pieces of evidence do not preclude an important role for crime decline in generating favorable conditions for urban revival, but they suggest that the root of this widespread, recent and youth driven phenomenon lies elsewhere.[38]

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 little attraction to highly ranked school districts, unlike the middle-aged and older college-educated.

We are wary of interpreting the coefficient on our post-period transit variable causally, but note that transit efficiency in 2014 does not correlate with a positive influx of the young and college-educated. Morever, transit users are disproportionately low income (LeRoy and Sonstelie 1983, Glaeser et al. 2008), which is inconsistent with urban transit explaining the urbanization of college-educated individuals.

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

	School Quality		Violent Crime Rate		Transit	Regression
	Change	Level	Change	Level	Time (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
Less than College Education:						
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. See Table A.6 in appendix for full regression results.

5.2 Alternative housing cost data

The house price index in our main specification 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 three different ways.

[38]One concern with the police department-level crime data is that it is too coarse to capture relevant within-CBSA variation in crime rates. To address this concern, we re-run the public amenity specification on the subsample of CBSAs in which no police department covers more than 30% of the population. The coefficients on violent crime change only marginally in response to this adjustment.

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 Zillow's rental index, which is available only for a small sample of tracts. 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. Panel C in Table 4 shows that non-tradable services levels remain the main driver of urban revival in all cases.

In an online Appendix G we investigate another explanation for urban revival related to housing markets, which is that limited mortgage credit availability following the housing crisis and recession of 2007-2009 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.

5.3 Homophily

Our regressions include the 2000 level of the own-group tract share to assess whether changes in 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 j. We also control for the change in population density computed in this way. Table 6 shows the results of this specification for the young-college group side-by-side with the base specification results for this group. Adding these controls indeed tempers the magnitude of almost all coefficients, as expected. However, our key result holds: Table 4 shows that the level of non-tradable services is still the most important factor drawing the young and college-educated downtown.

5.4 Commuting

The simultaneous determination of job and residential locations is a key identification concern in residential choice model estimation. This simultaneity problem is straightforward; young and 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,

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

	Base Specification		Homophily Controls	
Variable	Change [1]	Level [2]	Change [3]	Level [4]
Job Opportunities (LI)	-0.114***	-0.035***	-0.135***	-0.013**
Job Opportunities (MI)	-0.090***	0.036***	-0.050***	0.006
Job Opportunities (HI)	0.212***	0.012	0.170***	-0.004
House Price Index	0.019***	-0.009***	0.005	-0.006***
Restaurants	0.385***	0.394***	0.060***	0.056***
Food Stores	-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% significance level; ** – 5% significance level; ***–1% significance level. The change in house prices, level of local demographic share, change in amenity density, change in job opportunities, and change in the share of type d individuals within CBSA c who live in tract j are considered endogenous variables and instrumented in first stage regressions. Each regression is weighted by the share of type d in a CBSA c in year 2000.

which is often the stated 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, but the workplace fixed-effect costs us the ability to assess the relative the importance of jobs and amenities in explaining urban revival. So we use LODES data to compare estimates from our main residential choice specification with estimates from a workplace-residence choice model with workplace fixed effects. Crucially, this comparison reveals little bias in estimating a residential-choice model only, as we do elsewhere in the paper. This model and results are in appendix D.[39]

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-income people. We aggregate the LODES data into commute matrices (Figure 6)

[39]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. Here we show how commute data provides a sharper identification strategy to distinguish consumption from production in a within-city context.

showing the 2002-2011 percentage change in the number of workers living and working at various distance 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 eight 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 the 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 column, thus confirming 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) focuses exclusively on high-income workers, again in all CBSAs. Unlike the general working population, high income workers are not systematically decentralizing their workplaces and residences. We instead observe increases in the number of high-income 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-income 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-income workers are following jobs or whether jobs are following high-income worker. To distinguish these explanations, one 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-income 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-income people in large cities. The attractiveness of downtown residences to high-income workers must rise to explain why they incur larger commute costs than before to live there.

Though based on the location patterns of a different set of people than our main census results (LODES high-wage earners are a much larger group than the young and college-educated), these commuting results support our conclusion that residential amenities drive urban revival. We emphasize that in the US, the share of trips to consumption amenities is larger than the share of trips to work. One should not be surprised that such amenities are important determinants of location choices.

Figure 6: Commute Patterns

(a) **All Workers in All CBSAs**

Residence-CBD Dist. \ Workplace-CBD 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

(b) **All Workers in All CBSAs**

Residence-CBD Dist. \ Workplace-CBD 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

(c) **High-Income Workers in All CBSAs**

Residence-CBD Dist. \ Workplace-CBD 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

(d) **High-Income Workers in Largest 10 CBSAs**

Residence-CBD Dist. \ Workplace-CBD 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-income workers earn more than $3333/month in nominal dollars.

6 External Validity: NHTS Trip Shares and CEX Expenditure Shares

We now use NHTS trips and CEX expenditures to corroborate the rising attractiveness of non-tradable services to young professionals. We focus on the four amenity categories that have a reasonable counterpart in both the CEX and the NHTS: restaurants, bars, food stores, and apparel stores. For each amenity category 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.[40]

These results for restaurants and food stores are in Panels A and B of Figure 7 (the results for bars and apparel stores are in Appendix C). Each panel shows the shares in 2010 level on top and in 2000-2010 changes at the bottom, with expenditure on the left and travel on the right. The young and college-educated have the largest expenditures on and travel shares to restaurants; the lowest expenditure share on food; and the second lowest (after the middle-aged college group) trip share to buy food and other goods such as clothes and hardware. They also have either the most or second most positive change in expenditure shares and in travel shares to restaurants (the same holds for bars in Appendix C).[41] The opposite pattern holds for trips to buy goods (groceries/apparel/hardware) and food expenditures, for which the young and college-educated have the least proclivity. The absolute magnitude of the change is hard to interpret because non-tradable services are luxuries and the post-period overlaps with the Great Recession. Remarkably, changes in trip shares to restaurants (+4%) and, even more pronounced, changes in expenditure shares on bars (+33%) and changes in trip shares to go out (+8%) are positive for the young and college-educated. Changes in restaurant expenditures are negative for all groups. Of course, travel and expenditure shares may not capture preferences if their cost varies with proximity to amenities, and if the young and college-educated live closer to amenities in 2010. Using confidential geo-coded NHTS data, we verify that the travel patterns above hold controlling for the amenity density index of a traveler's residential tract. We do not have geo-coded data for the CEX.

[40]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 in Appendix A.

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

This increase in young professionals' 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 levels. That is, CEX and NHTS data lend credence to our key regression finding that young professionals experienced a positive change in their collective taste for non-tradable services that is larger than that for other age-education groups.

7 Explaining Changing Tastes

Our empirical analysis so far suggests that changing tastes for non-tradable service 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 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 indices in our main specification. We describe these indices briefly below and provide details in Appendix B.2.

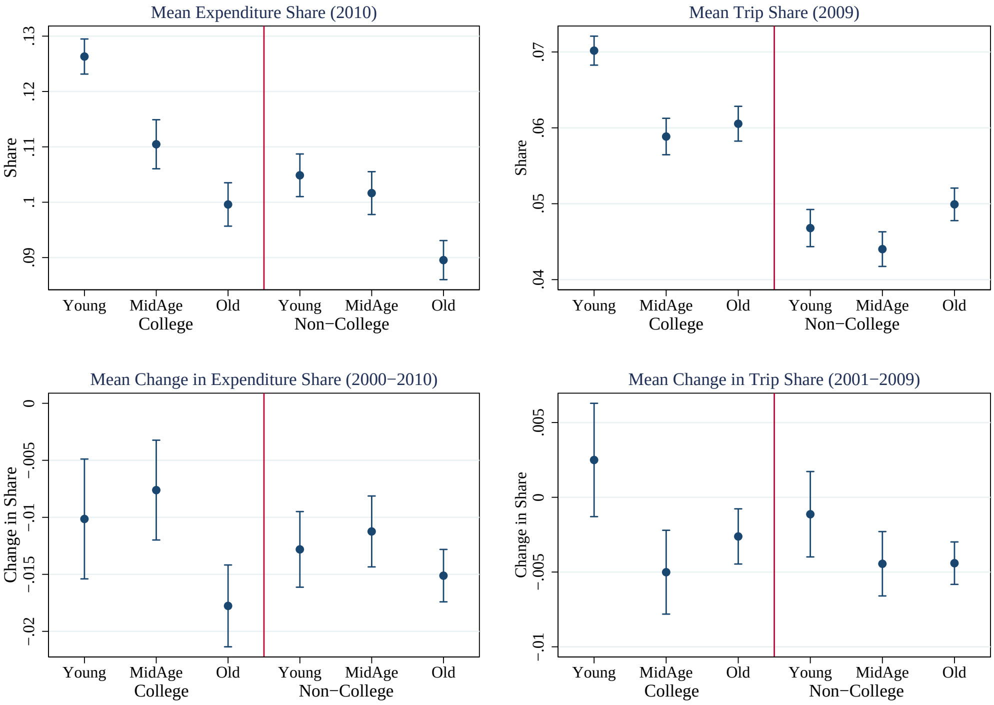
Our quality index uses data from ESRI business analyst, which divides each neighborhood in the U.S. into market segments, and assigns a "Market Potential Index" (MPI) to each chain in each segment. The MPI measures the propensity of each segment's inhabitants to visit a given chain relative to the average American. We use the MPIs in segments containing the largest share of young professionals to identify the restaurant and food chains that they prefer.[42] 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 largest 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 the young-professional MPI ratings near a tract, using the same transport costs weighting as for the amenity density indices. To alleviate concerns that changes in quality are driven by an influx of young professionals who report visiting chains near where they live, we instrument quality change 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 concentra-

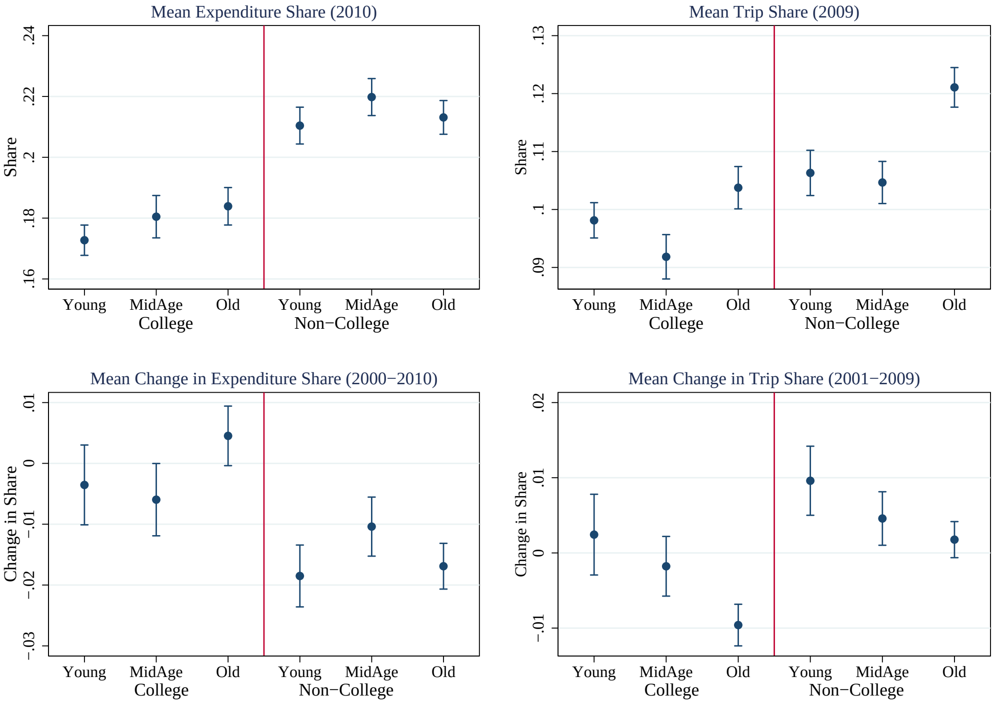
[42]We average MPIs across all segments within which individuals are both more than 50% college-educated and more than 50% aged 18 to 44. ESRI's MPIs come from the Survey of American Consumers, a proprietary dataset from GfK MRI.

Figure 7: Expenditure and Trip Shares on Tradable and Non-Tradable Services in the CEX and NHTS

Panel A: Restaurants



Panel B: Food Expenditures and Trips to Buy Goods



Notes: 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 goods includes food, apparel, and hardware. All confidence intervals are 95% intervals.

tion 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 college-educated are in Table 7. 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 that on food stores quality and diversity is near zero. This confirms the primacy of non-tradable services in attracting the young and college-educated. However, introducing these variables does not reduce the coefficient on restaurant 2000 level, which remains the main contributor to urban revival (Table 4).

Finally, Figure 2 shows that changes in restaurant quality and diversity contribute to the urbanization of young professionals *only* in large cities. These variables' ability to explain why urban revival is 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, but we do not find that these improvements explain the increasing taste of this group for non-tradable services.

Table 7: Nested-Logit Residential Location Choice Regression Results
Including AmenityQuality and Diversity

Variable	Quality and Diversity Change [1]	 Level [2]	Same Sample without Quality and Diversity Change [3]	 Level [4]
Job Opportunities (LI)	-0.094***	-0.029***	-0.137***	-0.041***
Job Opportunities (MI)	-0.075***	0.048***	-0.050***	0.062***
Job Opportunities (HI)	0.203***	0.006	0.210***	0.009
House Price Index	0.002	-0.025***	0.027***	-0.013***
Restaurants	0.380***	0.280***	0.272***	0.215***
Food Stores	-0.044***	-0.159***	-0.051***	-0.148***
Restaurant (Quality)	0.053***	0.018***		
Food Stores (Quality)	0.009	0.009		
Restaurants (Div)	0.174***	0.055***		
Food Stores (Div)	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% significance level; ** – 5% significance level; ***–1% significance level. The change in house prices, level of local demographic share, change in amenity density and quality, change in job opportunities, and change in the share of type d individuals within CBSA c who live in tract j are considered endogenous variables and instrumented in first stage regressions. Each regression is weighted by the share of type$d$ in a CBSA c in year 2000.

7.2 Homophily and non-tradable services

We now explore the role of homophily in the rising taste of young professionals for non-tradable services. We hypothesize that restaurants, bars, gyms and personal services establishment are increasingly attracting the young and college-educated because they provide opportunities to meet, network, and date other young professionals. 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 explain why the preferences of the young and college-educated have changed, but it hints at the mechanisms driving such changes.

To test this hypothesis, we interact the level and change in amenity densities with the 2000 share of the young and college-educated in a tract. Table 8 contains our main specification with these interactions in IV in column 1 and 2, and the same specification in OLS in column 3 and 4. Both specifications deliver similar results. The 2000 share of young and college-educated 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, indicating a more positive change in preferences for restaurant density in locations with large young-college shares. Importantly, introducing these interactions reduces by more than half the coefficient on the un-interacted restaurant variables in level. In other words, recent changes in the young college-educated preferences for non-tradable services depend on the presence of other young professionals nearby.

Table 8: Nested-Logit Residential Location Choice Regression Results
Homophily Interaction

	IV		OLS	
	Change	Level	Change	Level
Variable	[1]	[2]	[3]	[4]
Job Opportunities (LI)	-0.109***	-0.026**	-0.068***	-0.047***
Job Opportunities (MI)	-0.133***	0.006	0.047***	0.108***
Job Opportunities (HI)	0.242***	0.022**	0.033***	-0.053***
House Price Index	0.009	-0.001	0.062***	-0.001
Restaurants	0.347***	0.183***	0.010***	0.005**
Food Stores	-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% significance level; ** – 5% significance level; ***–1% significance level.

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 professionals spend more on and travel more to luxury and single amenities like bars and restaurants, then delayed family formation and income growth can explain their rising collective taste 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 in the mix of consumption amenities that they consume and visit. We then investigate the potential for changes in household and income type composition first to explain (i) the urban growth of young professionals in general, and (ii) the change in their preferences for non-tradable services that our model highlights.

Our analysis combines IPUMS data on the distribution of five household types and four income brackets across space and over time (Figure 8) with NHTS and CEX data on the types most likely to spend on or travel to consumption amenities (Figure 9). 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.[43] 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.[44] 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 8 shows the spatial distribution of different household and income types, with the 2000 share of each type within urban areas in blue and the 2000 share of each type within suburban areas in green. Panel B shows time trends, with the 2000 share of each type in light orange and the 2007-2011 share of each type in dark orange. Figure 9 reports expenditures and travel shares to non-tradable services - restaurants in Panel A and bars in Panel B - by household and income types in 2000. Appendix C provides a similar figure for food stores, apparel stores and trips to buy goods.

Solo households stand out in both sets of figures. In 2000, a plurality of young professionals

[43] 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% 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. We use 2007-2011 IPUMS data because it is the latest year to use 2000 PUMA definitions.

[44] 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 therefore classified as "others" in this data. Indeed, Figure 9 shows that expenditures and trip shares for "others" are similar to that for "solo".

live in solo households - more than for any other age-education group. Solos are by far the most urbanized type, accounting for 60% of urban but only 30% of suburban young professionals. Solos also have largest expenditure (CEX) and travel share (NHTS) of any household type for bars and restaurants. For instance, solos' expenditure share on restaurants is twice as large as that of families with young children, and that on bars is four times as large. Turning to income types, we observe more elaborate spatial sorting patterns, with both the poorest and the richest income types over-represented in urban areas. Non-tradable service amenities, however, are luxury goods, with richer households spending more on and traveling more to bars and restaurants. We note that poor and rich households possibly value downtowns for different reasons, with for instance non-tradable services drawing in the rich, and access to transit drawing 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 professionals. Figure 8 shows that young professionals 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. However, these changes only predict 17% of the actual urban-suburban young professional growth differential, and provide counterfactual predictions for other age-education groups. A similar shift share analysis by income type also pushes young professionals downtown, but this result is hard to interpret because of the late 2000s recession.

We now perform a similar analysis to explain changes in preferences for non-tradable service amenities. Figure 8 shows that household types with the highest propensity to spend on and travel to restaurants and bars ("solo" and "others") are growing while types with the lowest propensity (families with children) are shrinking.[45] A shift share analysis therefore suggests that changes in the family structure of young professionals increased their expenditures on and travel to both restaurants and bars. This push is small at less than 5 percent of initial 2000 levels.[46] A similar shift share analysis for income does not provide a positive push, because bars and restaurants are luxury goods and the recession reduces income measured in the post-period.

Finally, one may worry that solo and richer households patronize urbanized/single/luxury amenities 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, which is driven by something else. Using confidential geo-coded NHTS data, we find that the higher propensity of rich and solo

[45] 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 change in NHTS and CEX trip and expenditure shares, meaning that the increase that we document in section is not due to changes in household and income types.

[46] 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 household of type n in 2010 and $x_{n,00}$ is the expenditure (or travel) share for type n in 2000.

households to travel to non-tradable services relative to other types persists almost entirely after controlling for amenity density near a traveler's residence.

To summarize the results of this section, delayed family formation imply that young professionals are increasingly likely to live in households with high propensity for co-locating with, spending on and traveling to non-tradable services. Simple decompositions, however, show that although the differences across household types in propensity to travel to and spend on bars and restaurants are quite large, changes in family structure over the last decade are small enough to mute the overall impact of these trends. Income growth for the young and college-educated could have a similar impact because richer young professionals 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 innovation in mobile technology like mapping applications and establishment rating aggregators may complement urban amenities and disproportionately benefit digitally savvy young professionals. This hypothesis is hard to test directly. We look at the local share of NHTS establishments that are independent, because they stand to benefit more than chains from maps and review portals.[47] However, we find no evidence that this share of independent restaurant 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 gathers considerable media attention and interest from the general public. Using census data, we show that this revival is indeed happening in almost all large US 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 have flocked downtown.

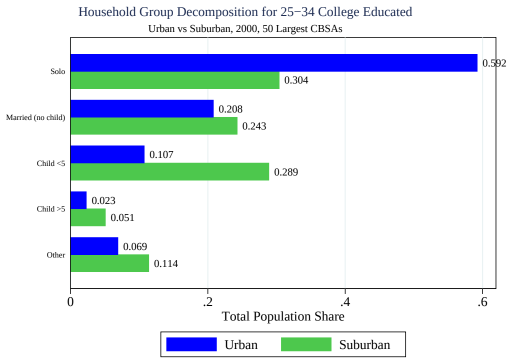
We evaluate the importance of various explanations for this trend. We find that diverging preferences for non-tradable services like restaurants, bars, gyms, and beauty salons explain the diverging location decisions of the young and college-educated relative to other groups.

[47] We define independent establishments in the NETS data as having at least 5 other establishments with the same name (see appendix section B.2 for details.) The NPD Group, a marketing agency, reports 53.8% of independent restaurants in the Spring of 2010. We find 49.6% in the NHTS.

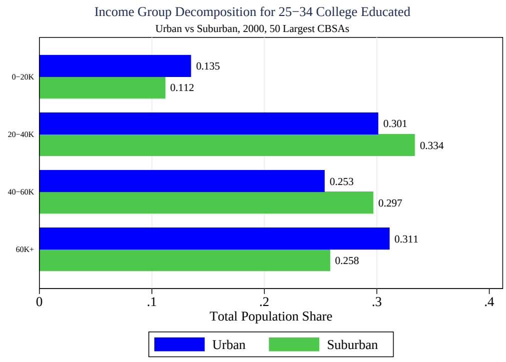
Figure 8: Share of 25-34 Year Old College-Educated Individuals by Household Type and Income Bracket.

Panel A: Share in Urban vs Suburban Area in 2000

By Household Types

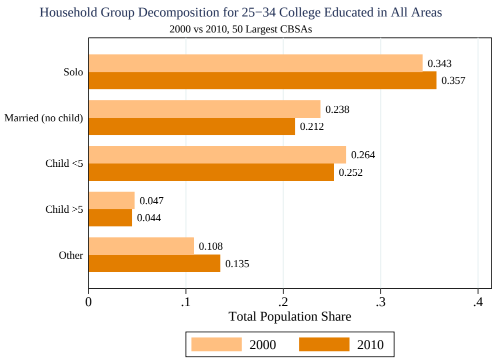


By Income Brackets

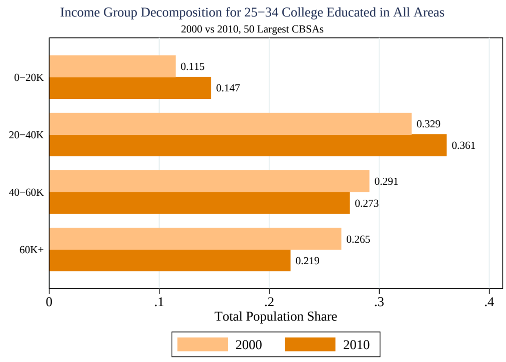


Panel B: Share in 2000 vs 2007-2011

By Household Types



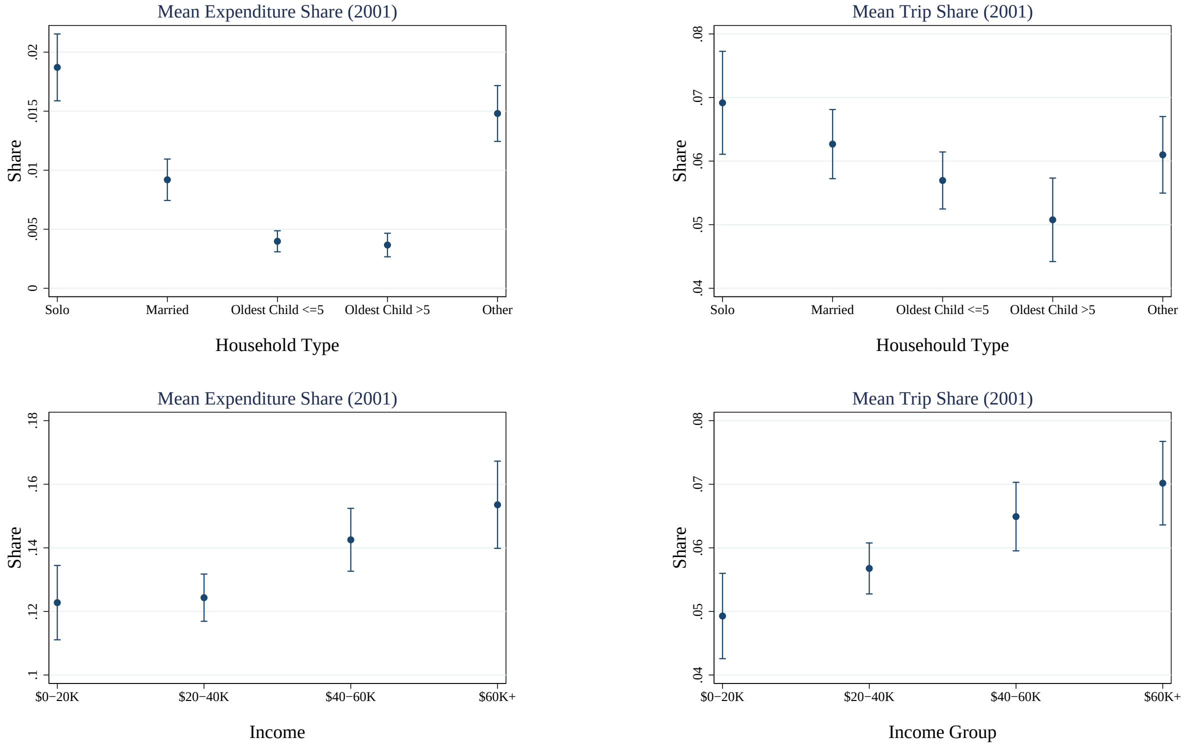
By Income Brackets



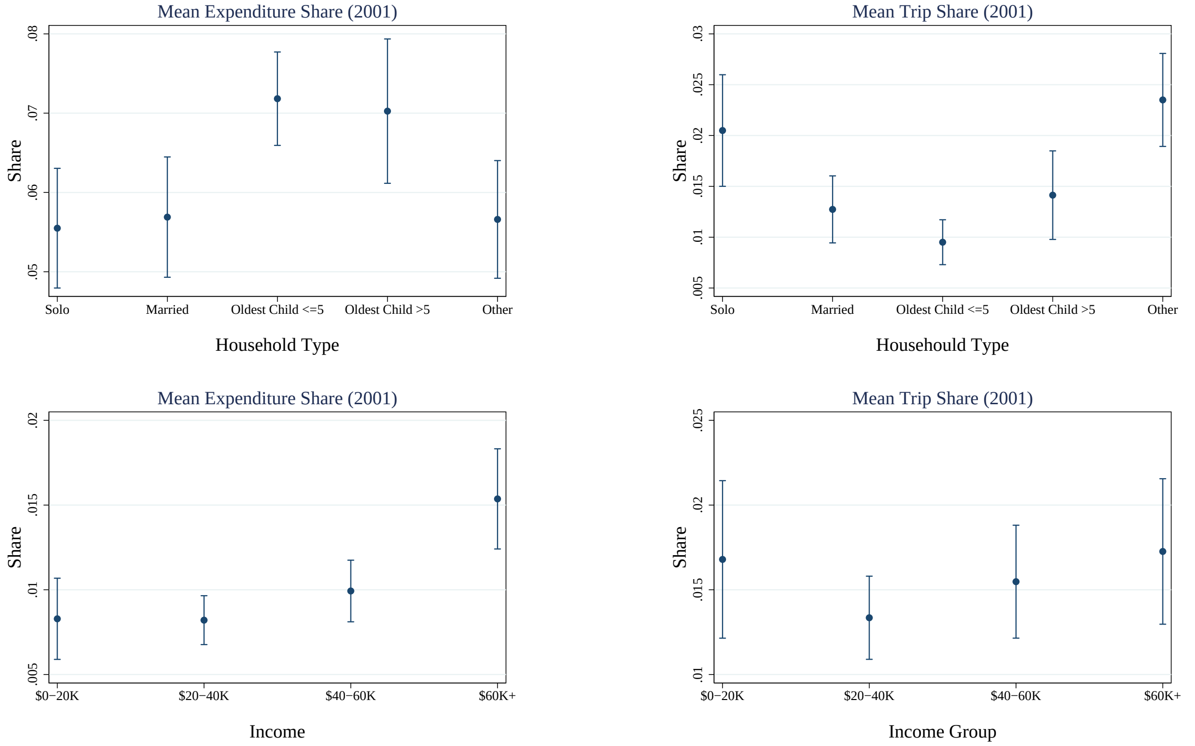
Notes: Shares computed out of all individuals 25-34 year old college-educated in the 50 largest CBSAs from the 5% sample of IPUMS in 2000 and an aggregate of five 1% samples from 2007 to 2011. 2007-2011 income is discounted to 1999 dollars by a factor 0.741 (recommmended in IPUMS data) and all income 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 CBD until it accounts for 10% of the population.

Figure 9: Expenditure and Trip Shares on Non-Tradable Services by Household Types and Income Groups for College Educated 25-34 Year Olds

Panel A: Restaurants



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



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 right and mean NHTS trip shares on the left. All confidence intervals are 95% intervals.

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 amenity quality and diversity drives young professionals downtown in large cities, but we could not establish that these improvements explain the changing collective taste that we estimate. In addition to change in the composition of amenities, changes in the composition of the young and college-educated themselves might explain their urbanization. For instance, the young and college-educated are increasingly likely to report living alone, and solos have much stronger demand for urban living and non-tradable service amenities, perhaps because they have more leisure time and disposable income, and higher demand for networking, socializing and dating opportunities. Indeed we find that the rising taste of the young and college-educated for non-tradable services depends in large part on the presence of other young professionals nearby (homophily). Other explanations, such as a complementarity between urban living and mobile technology that benefits digitally savvy young professionals, are harder to test and remain speculative.

It is striking that the classic factors used to explain residential location decisions (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 collective tastes mean that while high quality non-tradable services may compensate the young and college-educated gentrifiers 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 leave exploring these welfare implications to future work.

References

Albouy, David and Bert Lue, "Driving to opportunity: Local rents, wages, commuting, and sub-metropolitan quality of life," *Journal of Urban Economics*, September 2015, *89*, 74–92.

Bartik, Timothy J., "Who benefits from state and local economic development policies?," *Books from Upjohn Press*, 1991.

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

_ **and Daniel Hartley**, "Causes and Consequences of Central Neighborhood Change, 1970-2010," May 2016.

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

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

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

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

Carlino, Gerald A. and Albert Saiz, "Beautiful City: Leisure Amenities and Urban Growth," SSRN Scholarly Paper ID 1280157, Social Science Research Network, Rochester, NY December 2008.

Cortright, Joseph, "The Young and the Restless and the Nation's Cities," Technical Report, City Observatory October 2014.

Couture, Victor, "Valuing the Consumption Benefits of Urban Density," *University of California, Berkeley. Processed*, 2013.

Davidoff, Thomas, "Supply Constraints Are Not Valid Instrumental Variables for Home Prices Because They Are Correlated with Many Demand Factors," *Critical Finance Review*, May 2016, (ID 2400833).

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

Edlund, Lena, Cecilia Machado, and Michaela Sviatchi, "Bright Minds, Big Rent: Gentrification and the Rising Returns to Skill," Technical Report, NBER Working Paper w21729 November 2015.

Ellen, Ingrid Gould, Keren Mertens Horn, and Davin Reed, "Has Falling Crime Invited Gentrification?," Technical Report CES 17-27, Center for Economic Studies 2017.

Ferreira, Fernando and Joseph Gyourko, "Anatomy of the beginning of the housing boom: US neighborhoods and metropolitan areas, 1993-2009," Technical Report, National Bureau of Economic Research 2011.

Frey, William H., "The new urban revival in the United States," *Urban Studies*, 1993, *30*, 741–741.

Glaeser, Edward L., Albert Saiz, Gary Burtless, and William C. Strange, "The Rise of the Skilled City [with Comments]," *Brookings-Wharton Papers on Urban Affairs*, January 2004, pp. 47–105.

_ , Jed Kolko, and Albert Saiz, "Consumer city," *Journal of Economic Geography*, 2001, *1* (1), 27–50.

_ , Joseph Gyourko, and Raven E. Saks, "Urban growth and housing supply," *Journal of Economic Geography*, 2006, *6* (1), 71–89.

_ , Matthew E. Kahn, and Jordan Rappaport, "Why do the poor live in cities? The role of public transportation," *Journal of Urban Economics*, 2008, *63* (1), 1–24.

Graham, Matthew R., Mark J. Kutzbach, and Brian McKenzie, "Design Comparison of LODES and ACS Commuting Data," Technical Report, United States Census Bureau, Center for Economic Studies 2014.

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

Holian, Matthew J. and Matthew E. Kahn, "The Impact of Center City Economic and Cultural Vibrancy on Greenhouse Gas Emissions from Transportation," Technical Report, Citeseer 2012.

Igami, Mitsuru and Nathan Yang, "Unobserved Heterogeneity in Dynamic Games: Cannibalization and Preemptive Entry of Hamburger Chains in Canada," SSRN Scholarly Paper ID 2271803, Social Science Research Network, Rochester, NY September 2015.

Johnson, Kirk and Nick Wingfield, "As Amazon Stretches, Seattle's Downtown Is Reshaped," *The New York Times*, August 2013.

Kneebone, Elizabeth and Emily Garr, "The suburbanization of poverty: Trends in metropolitan America, 2000 to 2008," Technical Report, Brookings Institution January 2010.

Kotkin, Joel and Wendell Cox, "Cities and the Census," *City Journal*, April 2011.

Lee, Sanghoon and Jeffrey Lin, "Natural Amenities, Neighborhood Dynamics, and Persistence in the Spatial Distribution of Income by," Technical Report, Federal Reserve Bank of Philadelphia 2013.

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

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

McFadden, Daniel, "Conditional Logit Analaysis of Qualitative Choice Behavior," in "Frontiers in Econometrics," New York: Academic Press, 1973, pp. 105–142.

_ , "Modelling the Choice of Residential Location," in Anders Karlqvist, ed., *Spatial Interaction Theory and Planning Models*, New York: Elsevier North-Holland, 1978.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration and Local Employment Elasticities," Working Paper 21706, National Bureau of Economic Research November 2015. DOI: 10.3386/w21706.

Moretti, Enrico, *The New Geography of Jobs*, Houghton Mifflin Harcourt, 2012.

Neumark, David, Junfu Zhang, and Brandon Wall, "Employment dynamics and business relocation: New evidence from the National Establishment Time Series," *Research in Labor Economics*, 2007, *26*, 39–83.

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

_ , "Millennials, Baby Boomers, and Rebounding Multifamily Home Construction," SSRN Scholarly Paper ID 2637622, Federal Reserve Bank of Kansas City Working Paper, Rochester, NY June 2015.

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

Small, Kenneth and Erik Verhoef, *The Economics of Urban Transportation*, 2 edition ed., New York: Routledge, November 2007.

Stock, Kyle and Vanessa Wong, *Chipotle: The Definitive Oral History* February 2015.

Teulings, Coen N., Ioulia V. Ossokina, and Henri L. F. de Groot, "Welfare Benefits of Agglomoration and Worker Heterogeneity," Technical Report 289, CPB Netherlands Bureau for Economic Policy Analysis December 2014.

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

Waddell, Paul, Chandra Bhat, Naveen Eluru, Liming Wang, and Ram M. Pendyala, "Modeling interdependence in household residence and workplace choices," *Transportation Research Record: Journal of the Transportation Research Board*, 2007, *2003* (1), 84–92.

Appendices

A Data Appendix

A.1 Census data and ACS data

Census Tract Data and Definitions In every new census, some tracts are split or consolidated and their boundaries change to reflect population change over the last decade. The LTBD provides a crosswalk allowing to transform any tract level variable from 1970 to 2000 censuses into 2010 tract geography. This reweighting relies on census blocks population and area, and census blocks are small enough to ensure a high degree of accuracy. [48]

For our stylized facts on recent urban growth, we assemble a database of constant 2010 geography census tracts using the LTDB and data from the NHGIS for the 1980-2000 censuses and the 2008-2012 ACS (which already uses 2010 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 having a high degree of economic and social integration with that core. We assign 2010 census tracts to CBSAs based on 2013 CBSA definitions, and our model estimation sample consists of all 335 metropolitan area CBSAs.

A.2 LODES data

The LODES data comes from the Longitudinal Employer-Household Dynamics 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 opportunity indexes and wage group-specific Bartik instruments. We use the OD data for the residence-workplace model of subsection 5.4. 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 income groups ($1,250/month or less, $1,251/month to

[48] One source of error is that census blocks are sometimes split into different portions and assined to different 2010 tracts.

\$3,333/month, and greater than \$3,333/month). For all our analysis, 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). 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. There are two aspects to confidentiality protection in the LODES data.[49] First, the residential location of workers is synthesized. That is, the residential census block of a workers 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% of residences are coarsened above the census tract level, so synthesis has no impact on the vast majority of our sample 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 counts, and both tract-level aggregation and weighted regressions ensure a minimal impact of these procedures on our estimates.[50]

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 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% of establishments. In the remaining cases we assign an establishment to its zipcode centroid.

Neumark et al. (2007) assess the NETS reliability by comparing it to other establishment datasets (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, but warns against noisy year-on-year changes. They also report that NETS has better coverage than other data sources for very small establishments (1-4 persons), which is often the case of urban service amenities.

[49] Another source of measurement error comes from the LEHD source data, in which 40% 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.

[50] 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.

We assess the precision of the NETS by considering aggregate growth of chain establishments. 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 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 amenity category, as well as the SIC codes used to define these categories.[51]

A.4 Zillow house price indexes

Our main house price index comes from Zillow.com.[52] The Zillow House Value Index (ZHVI) for all homes (i.e. single family, condominium, and cooperative) 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 zipcodes in 2000 and 8,941 zipcodes in 2010 and the Zillow Rent Index (ZRI) for all homes, which is calculated each month for 13,875 zipcodes in 2010. For each zip code, we compute a yearly index by averaging over all months of the year. We map zipcodes to tracts with a crosswalk from the U.S. Department of Housing and Urban Development. 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 zipcodes, we normalize the residential share in zipcodes with available data to 1. If a tract does not fall into a zipcode 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 (UCR) data from 1990, 2000 and 2010. The data is broken down into violent crimes (murder, rape, robbery, and aggravated assault), property crimes (burglary, larceny, and motor vehicle theft) and arson. We only keep violent crime. UCR relies on each city's police district to self-report their crime statistics to

[51] 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% of the US population, suggesting close to 550,000 restaurants nationally. The NETS, at 416 807 restaurants nationally in 2010, appears to miss some establishments in other datasets.

[52] 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

Columns 2 and 3 show the total number establishments in each category in the US in 2000 and 2010 respectively.

the FBI. Thus, we lack coverage if any city did not report. All CBSAs have many cities. In 1990, there were 9222 cities reporting, which increased to 11,044 in 2010, partially because new cities were incorporated. To impute city-level data to census tracts, we use a GIS software to map every 2010 census tract into the corresponding city or cities that it overlaps with. We then assign the crime total for each city to the tracts that overlap with it (population-weighted overlap) assuming that population and crime are uniformly distributed within tracts and within cities. 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 cities.

A.6 Consumption Expenditure Survey (CEX) data

The Consumer Expenditure Survey (CEX) is conducted by the Census Bureau for the 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, for each respondent, all expenditures including 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 categories as follow:

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 7166 individuals in 1998-2002 and 7111 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 categories 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 (final person weight) to compute population estimates of mean trip shares.[53] Our sample size for the 24-35 year old college-educated (smallest group) is 6228 individuals in 2001 and 7309 individuals in 2009.

[53]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 5, but we know the age-range of the youngest child. If any child in a household does not fill the survey and we only know that the youngest child in the household is less than five, then we assume that the child who did not fill is less than five.

B Variable Definitions

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

B.1 Dependent Variable: Share of residents of type d living in tract j

The dependent variable 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_{jct}^{d} be the number of individuals of group d in tract j in CBSA c. Then the share of all type d residents who live in tract j in CBSA c at time t is:

$$s_{jct}^{d} = \frac{n_{jct}^{d}}{\sum_c \sum_j n_{jct}^{d}}.$$

B.2 Amenity Variables

Consumption Amenity Density Indexes We measure the level and change in the availability of different categories 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 category 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) for restaurants. The higher this elasticity, the lower the weight on establishments far away from an individual, and the more localized the amenity index. The price of a visit to an establishment is a constant derived from the CEX for that category, plus a transport cost by foot from the tract centroid.[54] So for each category a the density index in tract j is:

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

where *p* is the average price of visit to an establishment in amenity category 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 NHTS establishments in category 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

[54] Using CEX expenditures that most closely match our amenity category, we set a price of \$34.8 for museums, \$36.7 for golf/parks 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% of the average US wage as suggested in Small and Verhoef 2007).

to an establishment i.[55] 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 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.

Amenity quality, diversity and and independent establishment index

The restaurant and food store quality indices of section 7.1 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 category a in tract j is:

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

The amenity diversity indexes of section 7.1 are inverse Herfindahl indexes, which capture the diversity of the 70 restaurant SIC8 types and 66 food store SIC8 types:

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

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 5 members, and 1 to all other "independent" establishments.

Predicted Establishment Entry and Exit for Amenity Instrument Table A.2 displays regression results from equation 4 aggregated over all 1078 SIC8 codes for establishments included in our nine amenity indices. The Table provides the percentage of SIC8 codes for which a control variable has a positive (column 1) or negative (column 2) and significant coefficient at the 10% 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.)[56]

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

[56] 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 control variable in the entry and exit regression in Equation 4 at the SIC8 level, and provides the percentages of significant variable at the 10% level out of all 1078 SIC8 codes within our 9 amenity categories.

B.3 Jobs Opportunity 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 opportunity index for a tract j' for wage group g is:

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

where $n^g_{j'jt}$ is the number of persons 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'.

Table A.3: Tract-level Predicted Establishment 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 control variable in the entry and exit regression in Equation 4 at the chain level, and provides the percentages of significant variable at the 10% level out of the 61 restaurant chains for which we have MPI data.

C Additional Tables and Results

C.1 Urbanization Contribution Rankings

Table A.4: Ranking of Restaurant and Non-Tradable Services Urbanizing Contribution Across Variables and Across Age-Education Groups

	IV				OLS (except for within-CBSA share)			
	Rank of Variable Contribution		Rank of Young College Coeff (/6)		Rank of Variable Contribution		Rank of Young College Coeff (/6)	
Specification	Level	Change	Level	Change	Level	Change	Level	Change
Panel A: Base Specification (2 amenities - reporting contribution of restaurant variable only)								
Base Specification (2 Amenities)	1 (+)	13 (-)	1 (+)	1 (+)	3 (+)	8 (-)	1 (+)	1 (+)
Non-nested with CBSA Fixed Effects	1 (+)	14 (-)	1 (+)	1 (+)	6 (+)	10 (-)	1 (+)	1 (+)
Panel B: Base Specification with Additional Controls (reporting contribution of restaurant variable only)								
Distance to CBD	1 (+)	14 (-)	1 (+)	1 (+)	3 (+)	9 (-)	1 (+)	1 (+)
School, Crime and Transit	1 (+)	18 (-)	1 (+)	2 (+)	3 (+)	10 (-)	1 (+)	2 (+)
Homophily Control	1 (+)	14 (-)	2 (+)	3 (+)	5 (+)	10 (-)	1 (+)	1 (+)
Panel C: Base Specification with Alternative Housing Index (reporting contribution of restaurant variable only)								
Ferreira/Gyourko Hedonic Index	1 (+)	13 (-)	1 (+)	1 (+)	3 (+)	7 (-)	1 (+)	2 (+)
Zillow 2 Bedroom Index	1 (+)	13 (-)	1 (+)	1 (+)	3 (+)	8 (-)	1 (+)	1 (+)
Housing on the LHS	1 (+)	10 (-)	4 (+)	4 (+)	3 (+)	7 (-)	1 (+)	6 (+)
Panel D: Specifications with Additional Amenity Controls								
Base Specification with Amenity Quality and Diversity:								
Restaurant Density	1 (+)	20 (-)	1 (+)	1 (+)	3 (+)	17 (-)	1 (+)	1 (+)
Restaurant Quality	7 (+)	4 (+)	2 (+)	2 (+)	6 (+)	5 (+)	1 (+)	2 (+)
Restaurant Diversity	14 (+)	3 (+)	4 (+)	4 (+)	13 (+)	10 (+)	2 (+)	2 (+)
Panel E: Contributions of All Non-Tradable Service Variables for Base Specification								
Base Specification with 9 Amenities								
Restaurants	1 (+)	24 (-)	3 (+)	4 (+)	4 (+)	19 (-)	1 (+)	1 (+)
Bars	3 (+)	11 (+)	2 (+)	2 (+)	5 (+)	10 (+)	1 (+)	3 (+)
Personal Services	2 (+)	22 (-)	5 (+)	5 (+)	7 (+)	20 (-)	6 (+)	6 (+)
Gyms	4 (+)	14 (-)	2 (+)	2 (+)	2 (+)	11 (-)	2 (+)	3 (+)

C.2 First-Stage Result for Main Specification in Table 2

Column 2 of Figure A.5 reports the reduced-form first-stage statistics, column 2 reports the first-stage SW conditional F-statistic from Sanderson and Windmeijer (2016), and column 3 reports an under-identification test, also from Sanderson and Windmeijer (2016).

Table A.5: First Stage

Census IV regression, 25-34 College

Endogeneous Variables	Reduced-Form F Stat	Conditional SW F Stat	Under-ID SW Chi-2
Change in House Price	873.95	82.59	2480.91
Change in Food Store Density	350.17	24.90	748.06
Change in Restaurant Density	231.34	9.17	275.57
Change in Within CBSA share	53.23	22.70	681.89
Change in Job Opportunities (L)	1823.31	68.40	2054.64
Change in Job Opportunities (M)	1888.26	69.60	2090.80
Change in Job Opportunities (H)	1641.77	79.14	2377.45
Share of Same Type	635.09	18.49	555.53
Population Density	2853.82	145.46	4369.67

Notes: Sanderson and Windmeijer (2016) do not report critical values for their F-statistic and recommend the use of Cragg-Donald critical values from **?**, which are unavailable for regressions with more than 2 endogenous variables. The standard rule of thumb is that an F-stat smaller than 10 is weak, in the sense that either that the bias of the IV estimator is larger than 10% of the bias of the OLS estimator at the 5% confidence level or else that a 5% Wald test rejects hypotheses at more than the 10% level **?**.

C.3 Full Robustness Tables

Table A.6 shows the full regression results for the specification with school, crime and transit of Table 5.

C.4 Additional NHTS and CEX results

Figure A.1 and A.2 are the same as Figure 7 and 9, but for food stores and apparel stores instead of restaurants and bars. Figure A.1 shows expenditure and trip shares in level and change by age-education group, while Figure A.2 shows 2000 levels for the young and college-educated by household types and income groups.

Table A.6: Nested-Logit Residential Location Choice Regression Results Including School, Crime and Transit

Panel A: College Educated

Variable	25-34 Year Olds Change [1]	25-34 Year Olds Level [2]	35-44 Year Olds Change [3]	35-44 Year Olds Level [4]	45-65 Year Olds Change [5]	45-65 Year Olds Level [6]
Job Opportunities (LI)	-0.137***	-0.143***	-0.054***	-0.219***	0.066***	-0.106***
Job Opportunities (MI)	0.288***	0.177***	0.347***	0.234***	0.176***	0.165***
Job Opportunities (HI)	0.191***	0.109***	0.123***	0.121***	-0.009	0.046***
House Price Index	0.080***	-0.037***	0.021***	-0.008**	-0.010	0.013**
Restaurants	0.304***	0.258***	0.023*	0.011	0.027	-0.014
Food Stores	0.044**	-0.052*	0.052***	0.068***	0.091***	0.070***
2014 Transit Time		0.035***		0.005*		-0.013**
School Quality	0.001	-0.004	0.012***	-0.004*	0.023***	0.036***
Violent Crime	0.002	0.034***	-0.011***	0.023***	-0.007**	0.022***
Population Density		-0.136***		-0.106***		-0.192***
Share of Same Type		-0.065***		0.004		-0.093***
Within-CBSA share	0.689***		0.887***		0.594***	
Observations	10,452		10,353		10,648	

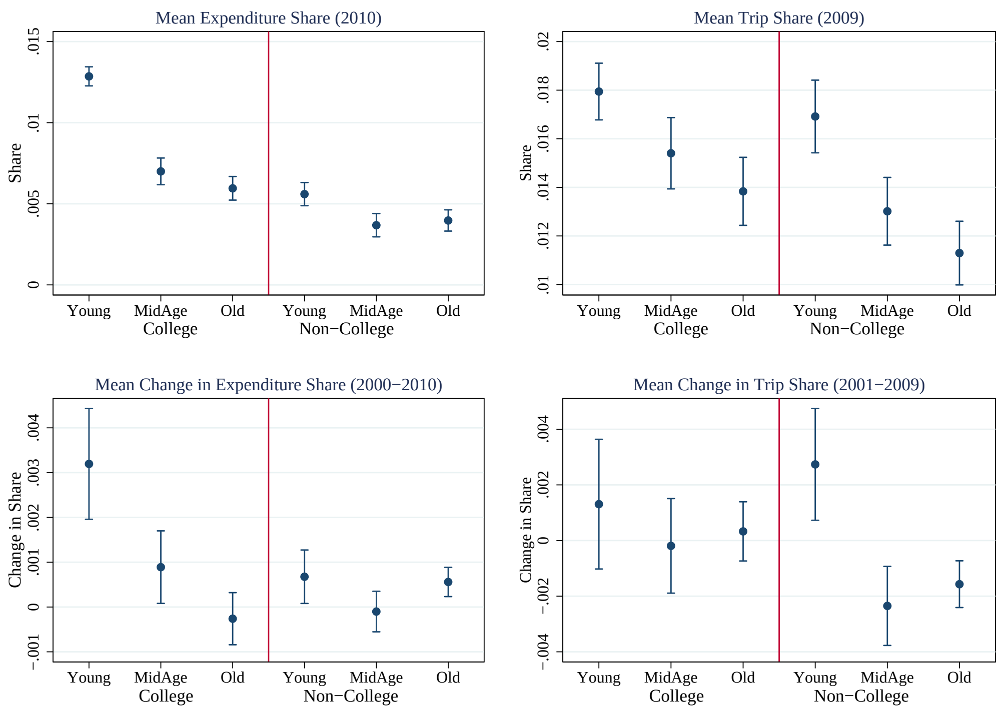
Panel B: Non-College Educated

Variable	25-34 Year Olds Change [1]	25-34 Year Olds Level [2]	35-44 Year Olds Change [3]	35-44 Year Olds Level [4]	45-65 Year Olds Change [5]	45-65 Year Olds Level [6]
Job Opportunities (LI)	-0.341***	-0.211***	-0.421***	-0.199***	-0.299***	-0.155***
Job Opportunities (MI)	0.537***	0.016	0.925***	0.157***	0.754***	0.035
Job Opportunities (HI)	0.360***	0.228***	0.028	0.191***	0.048	0.253***
House Price Index	0.016	-0.087***	0.074***	-0.035***	0.093***	-0.093***
Restaurants	0.430***	0.314***	0.186***	0.095***	0.321***	0.209***
Food Stores	0.141***	-0.054	0.165***	0.091***	0.165***	0.055
2014 Transit Time		0.061***		0.052***		0.050***
School Quality	-0.002	-0.014	0.003	-0.038***	0.005	-0.033***
Violent Crime	-0.023***	0.012	-0.032***	0.017**	-0.018***	-0.005
Population Density		-0.322***		-0.186***		-0.280***
Share of Same Type		0.006		0.061***		-0.010
Within-CBSA share	0.349***		0.650***		0.662***	
Observations	10,807		10,812		10,833	

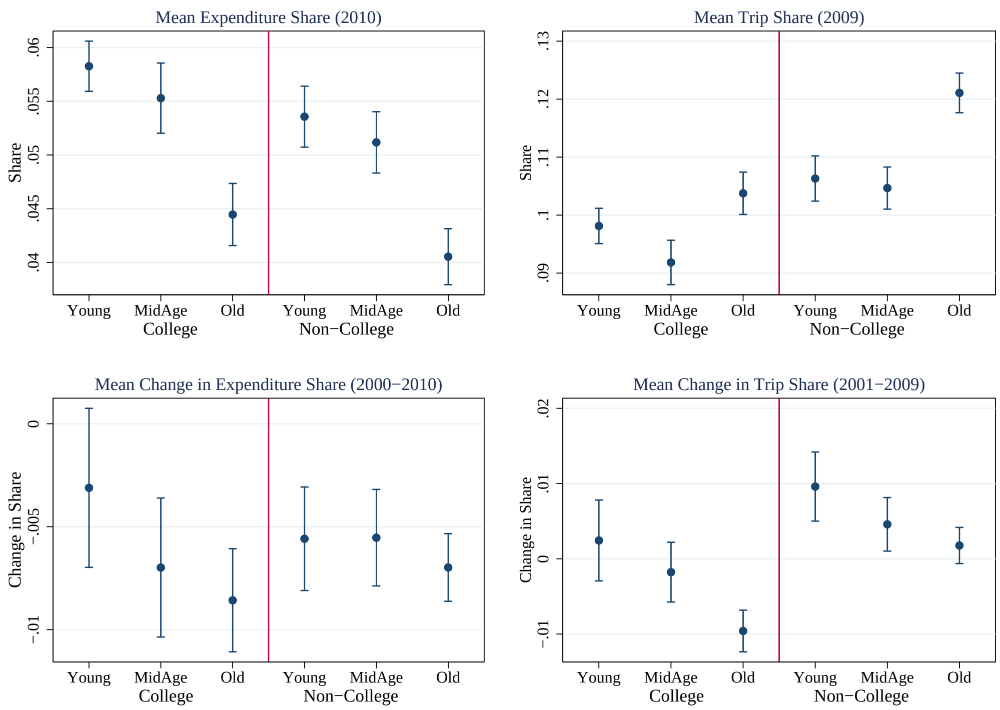
Notes: * – 10% significance level; ** – 5% significance level; ***–1% significance level. The change in house prices, level of local demographic share, change in amenity density, change in job opportunities, and change in the share of type d individuals within CBSA c who live in tract j are considered endogenous variables and instrumented in first stage regressions. Each regression is weighted by the share of type d in tract j in year 2000.

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

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



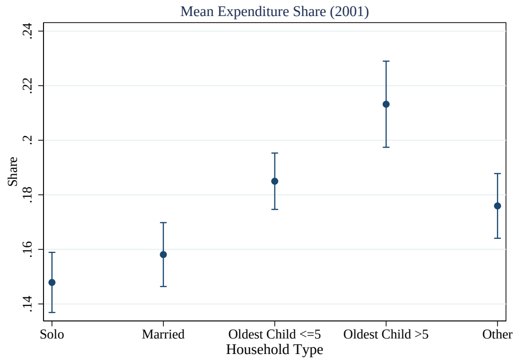
Panel B: Apparel Expenditures and Trips to Buy Goods

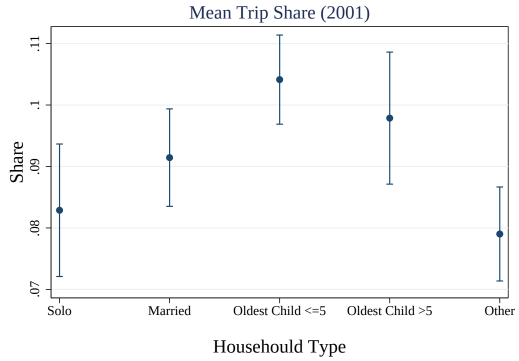


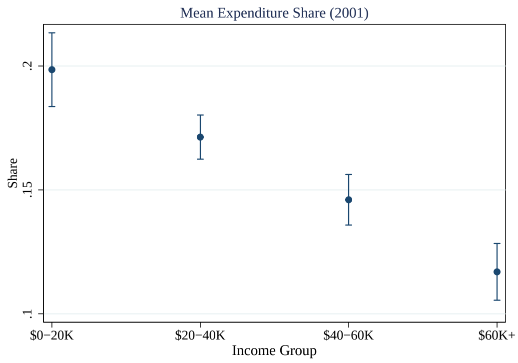
Notes: Data from the Consumer Expenditure Survey (CEX) and the National Household Travel Survey (NHTS). The left hand column of each panel shows mean CEX expenditure shares for each age-education group and the right hand column shows mean NHTS trip shares. Trip share to buy goods includes food, clothes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

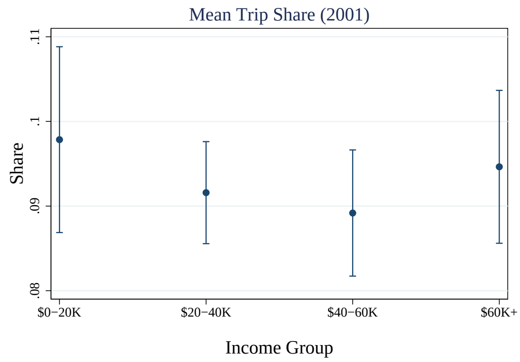
Figure A.2: Expenditure and Trip Shares on Goods by Household Types and Income Bracket for 25-34 Year Old College-Educated Individuals.

Panel A: Food Expenditures and Trip Shares to Buy Goods

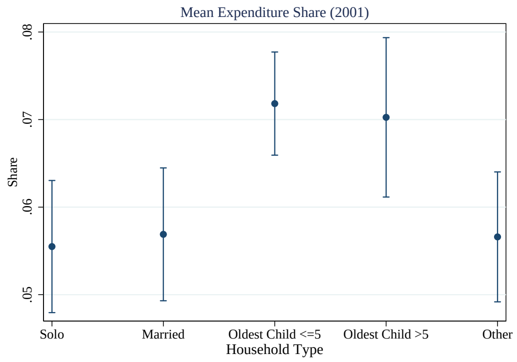


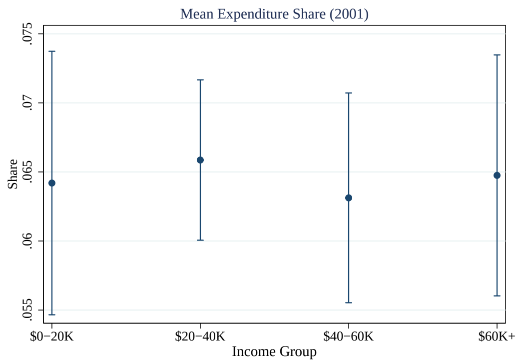






Panel B: Apparel Expenditures





Notes: 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 right and mean NHTS trip shares on the left. Trip share to buy goods includes food, clothes and hardware. All confidence intervals are 95% intervals.

D Commute Model

D.1 Specification and Identification

The model is as 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}:

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

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. ξ^d_{jct} and χ^d_{kct} are unobserved group-specific, time-varying quality of each residential and workplace location and θ^d_{ct} is an unobserved time-varying quality of CBSA c for individuals in group d.[57] We solve the model as in section 3 and the resulting estimating equation is:

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

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 jobs location. To understand this assumption, note that as in Monte et al. (2015), individuals select both their place of work and 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 also relies on first-differencing and a rich set of controls, but it exploits workplace fixed-effects instead of instrumental variables to estimate the 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 resi-

[57]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.

dential choice model estimated with the LODES data. To do so, we collapse the LODES data at the residential tract level and starting from equation A.5 we obtain the following estimating equation:

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

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 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 time

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 opportunities 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. The full set of nine consumption amenities highlight the impact of the workplace fixed effects on the residential characteristics coefficients. Column 1 provides, for comparison, OLS results for the young college-educated from our main specification in Table 1. Column 2 and 3 show LODES results for high-income individuals, with the residential choice model of equation A.6 in column 2 and the workplace fixed effect model of equation A.5 in column 3. High-income individuals 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 opportunities to near 0, which shows that once we control for where they work, high-income 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, personal services, food and apparel stores is positive and 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 models.[58]

Non-tradable service levels also make an important contribution to urbanizing high-income workers - second most important - even though most workers earning more than $3333 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 high-income individuals' aversion to long commute 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-income workers. Second, Figure 6 - and direct computations - 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 commute is older workers (>54 years old), which is consistent with Census results in Table 2, in which middle-aged and old college-educated have the most positive change in taste for proximity to high wage jobs (this

[58] 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

	25-34, College-Educated		High Wage (Residential)		High Wage(R-W Pair)	
	Change	Level	Change	Level	Change	Level
Variable	[1]	[2]	[3]	[4]	[5]	[6]
Job Opportunities (LI)	-0.069***	-0.045***	0.121***	0.019	0.042***	0.000
Job Opportunities (MI)	0.045***	0.106***	-0.319***	0.116***	-0.114***	-0.007*
Job Opportunities (HI)	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***
Museums and Libraries	-0.002***	-0.008***	0.000	0.013***	-0.001***	0.007***
Golf and parks	-0.001**	0.003***	0.000	-0.001	0.002***	0.005***
Gym and sports	0.011***	0.018***	0.017***	-0.004	0.019***	0.007***
Restaurants	0.006***	0.006***	0.009***	-0.035***	0.008***	-0.022***
Bars	0.001*	0.004***	0.008***	0.023***	0.005***	0.018***
Personal Services	0.011***	0.003	0.027***	0.002	0.027***	0.007***
Merchandise Stores	-0.004***	-0.013***	0.007***	0.054***	-0.002***	0.024***
Food Stores	0.001	-0.007**	0.019***	0.009*	0.015***	-0.001
Apparel Stores	0.005***	0.002	0.005***	-0.020***	0.003***	-0.025***
Share of Same Type		-0.004**		-0.175***		-0.144***
Tract Distance						-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 j in year 2000.

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 commute - through higher value of time - drives downtown gentrification. Rising value of time for high income individuals is consistent with the rising appeal of both jobs and amenities, and could be the topic of promising empirical work.