Urban Revival in America, 2000 to 2010*

Victor Couture
University of California, Berkeley

Jessie Handbury
University of Pennsylvania and NBER

July 2016

Preliminary Draft

Abstract

This paper documents and explains the striking reversal of fortune of urban America from 2000 to 2010. We show that almost all large American cities have experienced rising numbers in young professionals near their Central Business Districts over the last decade. We assemble a rich database at a fine spatial scale to test a number of competing hypotheses explaining this recent trend. We first estimate a residential choice model to assess the relative roles of amenities, job locations, and housing prices in drawing the young and college-educated downtown. We find that initial conditions of consumption amenities, especially service establishments, explain the diverging location decisions of the young and college-educated relative to their non-college-educated peers and their older college-educated counterparts. The coefficients on these initial conditions suggest that preferences for these amenities are changing over time. We investigate this hypothesis using complementary datasets, where we find that non-tradable service amenities are also playing an increasingly dominant role in the expenditure and travel decisions of the young and college-educated relative to other groups. Finally, we show that these new trends are partially explained by the changing income composition and family structure of the young and college-educated.

[a]Prottoy Aman Akbar, Yue Cao, Yizhen Gu, Jeffrey Jacobs, Hae Nim Lee, Ellen Lin, Daniel Means, and Jungsoo Yoo provided us with outstanding research assistance. We thank David Albouy, Nate Baum-Snow, Don Davis, Jorge De la Roca, Gilles Duranton, Ben Faber, Fernando Ferreira, Joe Gyourko, 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 would like to thank the Research Sponsors' Program of the Wharton Zell-Lurie Real Estate Center for generous financial support. Victor Couture would like to thank the Fisher Center for Real Estate and Urban Econonics for generous financial support.

1 Introduction

Mounting anecdotal evidence indicates that urban areas in American cities have experienced a reversal of fortunes since 2000, but a clear characterization of this trend has until recently proven largely elusive. In this paper, we document the extent of urban revival in the U.S. from 2000 to 2010 and seek to explain it. We first show that urban revival affects almost all large CBSAs in the United States. It is a highly localized phenomenon, characterized by large increases in young professionals near the Central Business District (CBD) of each CBSA. We then assemble a rich database at a fine spatial scale to test a number of hypotheses explaining the urbanization of young professionals. We devote particular attention to recent trends in the location of jobs and consumption amenities, and changes in the preference of different socio-economic groups for living in proximity to jobs and amenities. We find that the initial levels of highly urbanized consumption amenities like bars and restaurants have more power to explain the recent urbanization of young professionals than changes in the relative availability of jobs and amenities over the last decade, which rarely favored downtowns. We offer a structural interpretation of these results as capturing the importance of changing tastes for proximity to these amenities, and use complementary data on travel and expenditure shares to support this interpretation. Given the importance of the distribution of the college-educated for spatial success *across* cities (Glaeser et al. 2004; Moretti 2012; Diamond 2012), these stark *within*-city trends have important implications for the future of urban America.

We document the scope and size of urban revival by presenting a set of stylized facts, many of them new. We first confirm that, as in previous decades, the aggregate population is growing faster in the suburbs, relative to downtowns. However, in the 50 largest CBSAs, the population of 25-to-44 year old college-educated Americans is growing three times faster in downtown areas than in the suburbs.[1] The downtown areas experiencing urban revival are small in size, but the aggregate effects are large. For instance, in large CBSAs, downtown areas representing 5% of the population account for nearly 25% of the total growth in college-educated 25-35 year olds. Young professionals are urbanizing so fast in the largest 50 CBSAs that, despite the suburbanization of older age cohorts, a majority of these cities have seen their total college-educated population grow faster in downtowns relative to suburbs.[2] This last result stands in stark contrast with the poor relative performance of almost all downtowns from 1970 to 2000. In many ways, our work on urban revival complements the existing literature documenting and explaining the suburbanization era of the last century (Glaeser and Kahn (2004); Baum-Snow (2007); Boustan (2010) and others).

A number of competing hypotheses have the potential to explain our stylized facts on recent changes in the location choices of college-educated Americans. We are able to test many of these hypotheses by estimating a nested-logit residential choice model at the tract level. In this model, individuals first choose a CBSA to live in, and then choose a residential tract within that CBSA, based on tract characteristics like jobs, house prices and amenities. We allow individual preference parameters to vary both across age-education groups and over time. Our estimated model explains the differential growth of various age-education groups across tracts and, crucially, distinguishes the impact of recent changes in tract characteristics from that of recent changes in group-specific preferences for these characteristics. These exercises require the assembly of a rich dataset of geographically-consistent tracts in 2000 and 2010. To obtain our stylized facts and data on residential choices by demographic groups, we use Census and American Community Survey (ACS) tables. Our main database for the location of consumption amenities is the NETS data, which contains the universe of US establishments in 2000 and 2010. To measure job location, we use the LODES database, which contains data on the universe of tract-to-tract commute

[1] This trend is much more uneven in smaller cities.

[2] 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.

flows for different demographic and socio-economic groups.[3] We complement these primary datasets with data on natural amenities from Lee and Lin (2013), house prices from Zillow, school quality from SchoolDigger.com, crime from the Uniform Crime Reporting database, access to transit from Google Maps, expenditure shares from the Consumer Expenditure Survey (CEX), travel behavior from the National Household Transportation Survey (NHTS), and establishment quality from ESRI.

Our empirical framework contrasts with existing work in two important ways. First, we provide new measures of proximity to consumption amenities in the retail and service sectors, which we define precisely in both product (e.g., food vs apparel stores) and geographical space. To measure the role of these amenities in household location decisions we must tackle reverse causality issues. To do so, we construct novel instruments for the changes in consumption amenities in a given category, based on the interaction of national growth in an establishment type (at the chain or SIC8 level) with an estimate of the attractiveness of the pre-existing business environment for that establishment type. This instrument combines insight from the IO literature on cannibalization, preemption, competition, and agglomeration (e.g., Igami and Yang 2015) with the standard Bartik instrument in labor and urban economics.[4] Our ability to measure consumption amenities also 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 as an important driver of central city gentrification. Edlund et al. (2015), however, focus on the taste for shorter commutes of high-skilled workers.

Second, we estimate a two-period model using data for all CBSAs, instead of using a cross-section of data from a small survey sample, as is standard when estimating residential choice or residential-workplace choice models at fine spatial scale.[5] This first-difference specification allows us to control for some omitted variables that are constant in each location and reveals important inter-temporal variation in the factors that drive location choices of different demographic groups.

Our preference parameter estimates successfully explain the urbanization of the young and college-educated, especially in larger cities, and the suburbanization of the old and non-college educated everywhere. Of course, city size and proximity to downtowns are not themselves parameters of our estimated model. Instead, we show that the downtowns of large cities have special characteristics that attract the young and college-educated. In particular, we find that while a slight urbanization of high-income jobs over the last decade contributes to explaining urban revival, the parameters that signal changing tastes for urban consumption amenities play a more important role in explaining why the young and college-educated are disproportionately moving downtown in big cities. These results suggest that consumption amenities are a key factor explaining the residential location decisions of young professionals *within* CBSAs, possibly more important than jobs or the standard non-consumption amenities like crime, school, transit and group homophily that others have shown to be relevant in similar models. Perhaps unsurprisingly, we find that non-college educated individuals, who are not urbanizing, have stronger preferences than the college-educated for urban transit access. Similarly, only older cohorts have a positive preference for

[3] This commute data also allows us to extend the residential choice model, estimated using census data on residences, to a residential-workplace choice model. The results from this model will be used to demonstrate the robustness of our estimates – for the preferences for consumption amenities, in particular – to precise controls for workplace location (unobserved in the Census data) with the inclusion of workplace fixed effects. The intuition for this empirical strategy is similar to that in Glaeser et al. (2001), who suggest that an increase in reverse commuting (from central cities to the suburbs) signals the importance of urban amenities.

[4] We draw from the existing location choice literature to construct tract-level versions of standard instruments for house prices and job growth.

[5] Baum-Snow and Hartley (2016) also estimate a tract-level residential choice mode across CBSAs, but not a residential-workplace choice model. Albouy and Lue (2015) estimates a within-city residential-workplace model using data for all CBSAs using one year of data in 2000, and using large geographical unit of analysis (PUMAs). They find that the variation in quality of life is as important within metropolitan areas as across them motivating 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.

school quality, and although violent crime is indeed declining faster in urban relative to suburban areas, safety is a less important factor in the location decision of the young and college-educated than for all other groups which are not urbanizing. We caution against over-interpreting results on crime, school and transit access given measurement and identification issues that we discuss later in the paper, but overall the data delivers intuitive results that explain why these factors are unlikely to be the key drivers of the recent urbanization of the young and college-educated.

We conduct a series of complementary analyses to assess the plausibility of our main result on the importance of changing tastes for amenities, and to explain what may drive these changes in tastes. We begin by providing evidence from CEX expenditure shares and NETS trip shares for various consumption amenity categories. We show that differences across age & education groups in the share of expenditures and trips to an amenity and in the changes in such shares often correspond to differences in the model's estimated preferences and changes in preferences across groups for that amenity. For instance, our model parameter estimates suggest that the young and college-educated have a stronger taste for living in proximity to bars than other groups, and that they have experienced the most positive change in this taste over the last decade. Our expenditure and travel data confirm that young professionals spend the most on drinks away from home, travel the most to "go out", and have experienced the fastest increase in such expenditures and trips over the last decade.[6] We then investigate three hypotheses with the potential to explain the changing preferences of the young and college-educated for living in proximity to service amenities. First, what we interpret as a changing taste for high amenity density may in fact be due to changing amenity composition in high density areas, towards establishments that cater to the specific tastes of young professionals. We find support for this hypothesis after adding measures of restaurant composition to our residential choice model, using marketing studies from ESRI to identify restaurant chains favored by young professionals. Second, changes in the income distribution and family structure of the young and college-educated may shift their consumption towards non-tradable service establishments. We find support for this hypothesis by using complementary data on expenditure and trip shares for different amenities, which include information on individual income and family structure.[7] Third, recent innovation in mobile technology may complement urbanized amenities, which benefits digitally savvy young professionals. This hypothesis is speculative and hard to test directly. One indirect test within our residential-choice model is to control for the share of independent restaurants, which presumably stand to benefit more than chains from the convenience of mapping applications and the availability of review portals. However, we find no evidence that a high share of independent restaurants significantly impact the location decisions of young professional. We are investigating these hypotheses further in complementary work.

Finally, we investigate one last prominent explanation for urban revival that cannot be tested in our tract-level residential choice framework: the change in mortgage credit availability between 2000 and 2010. This hypothesis is that mortgage lending practices increased the demand of younger individuals for rental housing - which is highly urbanized - by restricting credit availability to new homeowners in the aftermath of the 2007-2009 housing crisis. The main flaw in this hypothesis is the timing of the housing crisis. The 2000s include more years of historically easy mortgage credit than of restricted credit. In fact, we find that homeownership rates among young professionals have *increased* from 2000 to 2010, in both urban and suburban areas. Further supporting the view that the housing crisis did not drive urban revival, we use the earliest ACS data available (2005-2009) and find patterns of urban revival that are very similar to those observed by comparing 2000 to later ACS averages. More than half of the 2005-2009 time period comes *before* the housing crisis, which again challenges the notion that reduced access to homeownership drives urban revival.

[6] Bars feature large variation in expenditure shares and trip shares, and in changes in such shares across groups, making such comparison easier. For other amenities, changes in trip and expenditure shares have large confidence intervals and such comparisons are harder.

[7] 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).

It is important to emphasize that while this paper focuses on documenting and explaining the recent urbanization of young professionals, what we call "urban revival" may have complex welfare consequences across the socio-economic spectrum. 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 follow. We describe the data in section 2. Section 3 presents the stylized facts on urban revival. Section 4 present the residential choice model, estimation, and robustness checks. Section 5 tests hypotheses on the drivers of the changes in preferences for amenities that we estimate in the previous section. Section 6 presents additional analysis on the housing tenure hypothesis and section 7 concludes.

2 Data

To establish the stylized facts on recent urban growth that motivate our empirical analysis, we assemble a database describing the residential and workplace locations of U.S. individuals at a decennial frequency. Tract-level population counts by age and education levels are from the decennial censuses of 1970 to 2000 and the American Community Survey (ACS) 2008-2012 aggregates, downloaded from the National Historical Geographic Information System (NHGIS). We complement the Census and ACS data on the residential locations of individuals with the LEHD Origin-Destination Employment Statistics (LODES) for 2002 and 2011. The LODES data provides counts of people in different age and income groups who live and work in a given census block pair.[8]

In all of our analysis, the main geographical unit is a census tract within a Core-Based Statistical Area (CBSA). We construct CBSAs from census tracts, using constant 2010 tract and CBSA boundaries from the Longitudinal Tract Data Base (LTDB). Our urban definition is based on proximity to the Central Business District (CBD) of the principal city of each CBSA using the CBDs defined in the 1982 Census of Retail Trade.[9] We interchangeably refer to CBSAs as cities and downtown areas as urban areas.

To explain these stylized facts, we build datasets describing the density and quality of amenities, the density of workplace locations, and house prices in the vicinity of each census tract. We use the LODES data to characterize accessibility to different types of job opportunities, as well as current places of employment. To compute amenities indices describing the accessibility of each census tract to various types of consumption opportunities such as restaurants and apparel stores, we use two datasets: (i) a geo-coded census of establishments in 2000 and 2010 from the National Establishment Time-Series (NETS); (ii) travel times between these establishments and census tract centroids by foot from Google Maps. We also use data on expenditure shares from the CEX and data on trip shares to different amenities from the National Household Transportation Survey (NETS) to provide external validity for the relative and changing preferences for amenities that we estimate in our location-choice model. We refine these indices to reflect establishment quality using ESRI's Market Potential Index (MPI), which measures the propensity of different socio-economic or demographic groups to shop in a given chain store. Our house price index for 2000 and 2010 is the Zillow "All Home Index" at the zip code level, that we match to 2010 tract geography using a USPS zip code-Census tract crosswalk file downloaded from HUD.gov. We expand the dataset beyond those tracts that the Zillow data covers by approximately 30% by spreading the house price indexes that we observe in 2000 and 2010 across all tracts within a tract-group, a set of three to four neighboring tracts defined in Ferreira and Gyourko (2011).[10]

[8] In early versions of the LODES data, census block pairs with very few individuals were simply censored. In the more recent version of the LODES that we use, confidentiality issues are addressed by making the data partially synthetic, in a way that preserves some key aggregate statistics from the data. We describe the procedure through which the synthetic data is generated in appendix A. Note that we do not use the census block data directly, but we instead aggregate the data at the census tract level, as recommended in the LODES documentation.

[9] CBD coordinates are sourced from Holian and Kahn (2012).

[10] Ferreira and Gyourko (2011) do similarly estimating hedonic price indexes at the tract-group level.

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 defined as the average travel time of a 5 mile trip to a random set of NETS establishments, starting from a each tract's population-weighted centroid. We obtain city-level data on violent crime (murder, rape, robbery, and aggravated assault) from the Uniform Crime Reporting (UCR) in 2000 and 2010. The data comes from each city's police district reports to the FBI. For each year, we transform city-level crime data into 2010 tract level data using Census shapefiles. In 2010, this mapping projects 11,044 cities or county subdivisions into 57,095 census tracts (out of 73,057 tracts in the entire United States).[11] . We obtain data on within-state ranking of school districts in 2004 and 2010 from SchoolDigger.com, that we match to 2010 tract boundaries using Census shapefiles.[12]Data on natural amenities, like precipitation, hilliness or coastal proximity for each census tract are from Lee and Lin (2013).[13]

To test the housing tenure hypothesis and to provide additional evidence on recent trends in household formation and income growth that can explain the changing preferences of young professionals, we require counts of individuals by home ownership status, household type, and income within each age-education group. To obtain these counts, we aggregate micro-data from the 5% Integrated Public Use Micro-data Series (IPUMS) sample of the 2000 census and the 5% IPUMS sample from 2006-2010 ACS surveys.

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

3 Stylized Facts

In this section, we establish a number of stylized facts about urban revival in US cities from 2000 to 2010, most of them new. These facts motivate the rest of our empirical analysis. We find that the average American is still suburbanizing, but uncover strong localized evidence of urban revival. We characterize this urban revival phenomenon as large increases in the young and college-educated population residing near the CBD of almost all large CBSAs. To conduct this analysis, we assemble a dataset of geographically-consistent census tracts with decennial census data from 1970 to 2000, and the 2008-2010 ACS aggregates for 2010. We define urban areas by sequentially adding census tracts closest to the CBD until the total urban population reaches no more than 5% of total CBSA population. Such areas are best thought of as downtowns.[14]

We then refine these stylized facts using LODES data on the number of workers in the universe of tract-to-tract residential-workplace pairs. This commute data allows us to describe changes in both workplace and residential location by age and income groups. Disentangling the role of job location from that of residential location characteristics in explaining the urbanization of higher income or more educated people is an objective motivating much of our empirical analysis. The commute data shows that both residences and jobs are decentralizing, within the set of all CBSAs. Within the 10 largest CBSAs, however, the reverse is true and both residences and jobs are centralizing. This centralization trend is *entirely* driven by high-income people, in the upper third of the income distribution. Most importantly, the data clearly shows that high-income people working at any distance from the CBD are more likely to be living near the CBD in 2011 than in 2002. This centralization of residential location

[11]CBSAs always consist of many cities, but there are never separate data for downtowns as small as those that we define, and in most CBSAs we have only one data point for the principal city. In some cases like Houston and Altlanta, the part of the principal city laying in different counties report different numbers.

[12]There are typically multiple census tracts within each school district. While we believe that SchoolDigger.com is the most comprehensive database available, we have school ranking data for less than half of our CBSAs sample of tracts.

[13]As we explain in section 4, we use a first difference framework which differences out the constant characteristics of each residential tract. This implies that we cannot put natural amenities directly in the regression as controls, but, as we will show, they can be used as instruments for other endogenous variables.

[14]We can replicate all of our main stylized facts with alternative downtown definitions (e.g., 5%, 10% or 15% of population, or keeping all tracts with centroids within 2, 3, 4 or 5 miles of the CBD) as long as the urban area is small enough.

holding job location fixed demonstrates that job centralization alone cannot explain our urban revival stylized facts.

3.1 Urban revival

Claims of urban revival are not new. The 1960s and 1970s were times of rapid decline for urban areas in America, with many central cities losing a significant share of their 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 American cities have become commonplace, and widely relayed by the the popular press. Census tables, however, tell an unequivocal story of continued suburbanization (Kotkin and Cox, 2011).

Figure 1a provides one way to visualize the continuing suburbanization of American cities since 1970. There is a plot for each decade between 1970 and 2010. Each plot shows the number of CBSAs in which either suburban or urban growth has been faster, for the 100 largest CBSAs. These CBSAs are ranked by 2010 population on the x-axis, and results are aggregated by groups of 10 CBSAs. The blue bar represents the number of CBSAs in which downtown population has been growing faster than suburban population within a group of 10 CBSAs. The green bar represents the number of CBSAs in which suburban population has been growing faster, with the green and blue bars always adding up to 10. For instance, the plot on the lower right shows that within the 10 largest CBSAs in the United States, only one experienced faster growth downtown relative to its suburbs. Looking instead at CBSAs with size ranking from 10 to 20, we see that none of them have urbanized during this period. Indeed, only 2 of the 100 largest CBSAs have experienced faster urban growth between 2000 and 2010. This pattern of faster suburban growth applies as well to previous decades for which we have data, and it is robust to different definitions of "urban," such as using central cities (not shown).[15]

Slower urban population growth does not necessarily preclude urban revival. Downtowns are often already built up and subject to heavy housing regulations (Glaeser et al., 2006), so their desirability is more likely to manifest itself through higher house prices and a demographic shift towards wealthier individuals than simply through faster population growth. In fact, many authors argue that in recent decades, the best indicator of spatial success is the share of an areas inhabitants that received a college education (Glaeser et al. (2004), Moretti (2012)). It is therefore natural to define urban revival as the urbanization of a city's college-educated population.

Figure 1b replicates Figure 1a, but considering growth in college-educated population alone. The results undercover a new, previously undocumented trend: between 2000 and 2010, a majority of the 50 largest CBSAs have experienced faster urban than suburban growth in college-educated individuals. This is not the case for any decade between 1970 and 2000. In the 1990s for instance, the college-educated population urbanized in only 11 out of the 50 largest CBSAs. The recent urbanization of college-educated individuals occurs mostly in the largest cities, and a sizable majority of CBSAs ranked 50 to 100 have not experienced faster urban than suburban growth in college-educated individuals from 2000-2010. It is important to highlight that the fastest growth of the college-educated population is in downtown areas relatively close to the CBD, so this localized trend is not apparent when defining urban areas as central cities; this probably explains why the urbanization of college-educated Americans has not been documented before.[16]

We now further refine our investigation, and break down the growth in college-educated population between 2000 and 2010 by age groups.[17] Americans become considerably less mobile as they age, and we do not expect new locational trends to predominantly affect older cohorts. We also expect the residential and workplace preferences

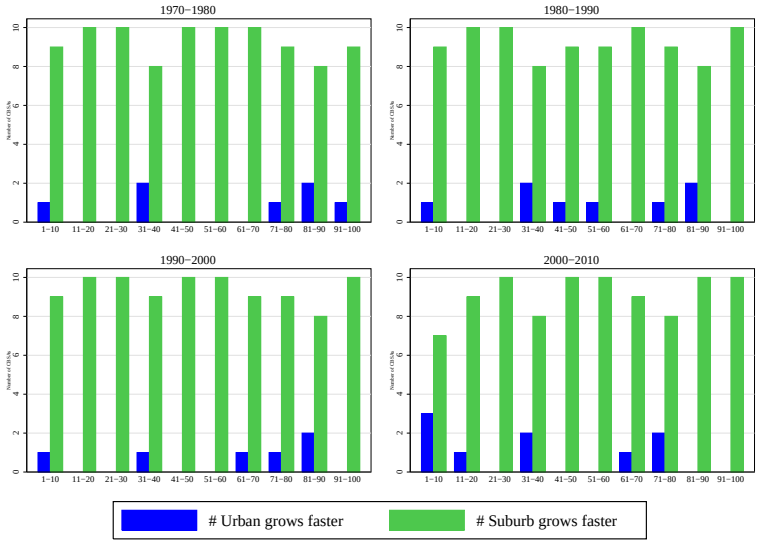
[15] We use the term 'suburb' for simplicity, to describe everything outside the downtowns that we define. Clearly, some of these areas within our suburbs are quite urban by most standard definitions. When we define downtowns as central cities, our definition of suburb becomes the usual one.

[16] Fee and Hartley (2012) document localized changes in aggregate population at various distances from CBDs and find that cities with increasing near-CBD population density have higher per capita income growth at the MSA level.

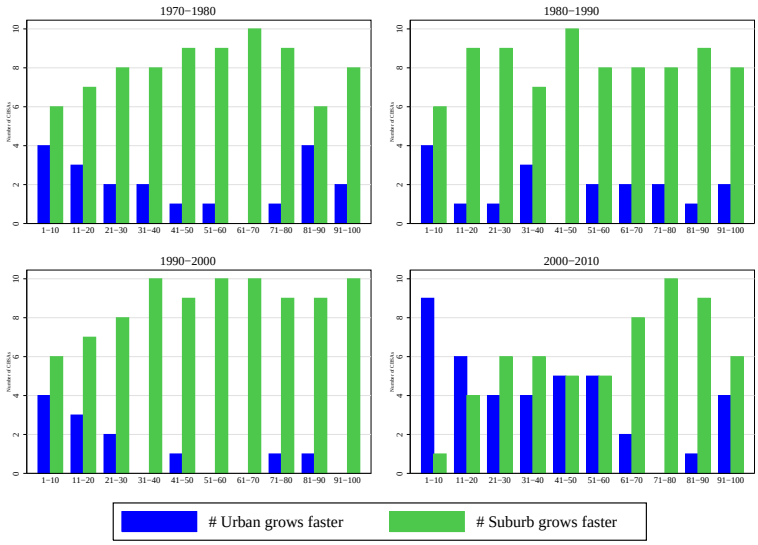
[17] Data on education by age group is only available at the census tract level starting in 2000, so this investigation is only possible for the last decade.

Figure 1: Downtown vs. Suburban Growth in the Largest 100 U.S. CBSAs, 1970-2010

(a) Total Population

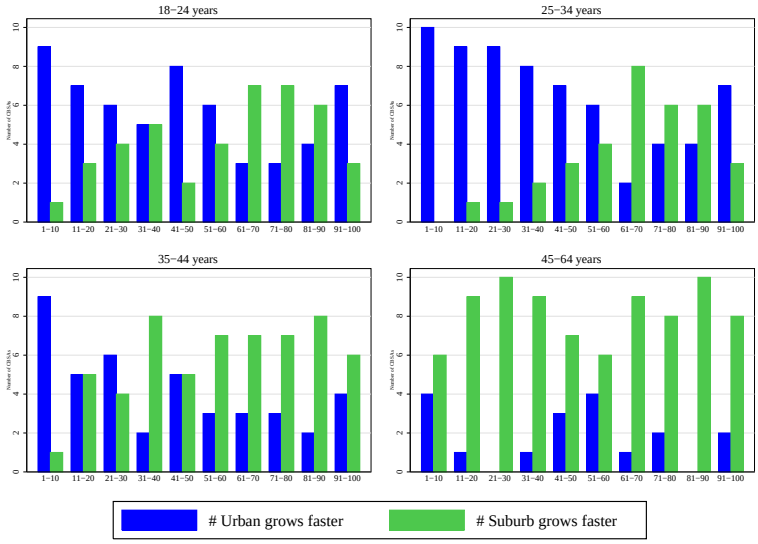


(b) College-Educated



Notes: Data from decennial censuses 1970–2000 and ACS 2008–2012. Each of the figures' four plots presents data for a different decade, starting from 1970–1980 in the upper left-hand plot to 2000–2010 in the lower right-hand plot. The x-axis ranks the 100 largests CBSAs by 2010 population, in groups of 10. For each CBSA, downtown is defined as all census tracts nearest to the CBD and totaling at most 5% of a CBSA population. The suburb contains the rest of a CBSA. In Figure 1a the blue bar represents the number of CBSAs in which downtown population has been growing faster than suburban population within a group of 10 CBSAs. The green bar represents the number of CBSAs in which suburban population has been growing faster. Figure 1b is similar but restricts attention to the college-educated (4 year degree or more) population only.

Figure 2: Downtown vs. Suburban Growth in the Largest 100 U.S. CBSAs, 2000-2010
College Educated



Notes: Data from decennial census 2000 and ACS 2008–2012. All plots are for 2000–2010. Each of the figure's four plots presents data for a different age group within the college-educated (at least 4 year degree) population, starting from 18-24 year old college-educated in the upper left-hand plot to 45-64 year old college-educate in the lower right-hand plot. The x-axis ranks the 100 largests CBSAs by 2010 population, in groups of 10. For each CBSA, downtown is defined as all census tracts nearest to the CBD and totaling at most 5% of a CBSA population. The suburb contains the rest of a CBSA. The blue bar represents the number of CBSAs in which downtown college-educated population in a given age group has been growing faster than suburban college-educated population of that age group within a group of 10 CBSAs. The green bar represents the number of CBSAs in which suburban college-educated population of a given age group has been growing faster.

of the younger generation to differ from that of older Americans.[18] Moreover, the popular press emphasizes the urbanization of both young people and retiring baby-boomers. A recent report by CEO for Cities (Cortright, 2014) - and covered extensively by the New York Times (Miller, 2014) - also uses 2000 census data and 2008-2012 ACS data, and shows that the 25-34 college-educated population are growing faster downtown than in the suburbs in the majority of the 51 largest MSAs.[19] We confirm and expand this narrative to the older 35-44 college-educated group, but interestingly we find that the popular press gets it wrong for educated baby-boomers, whose relative growth is faster in the suburbs of almost all large cities.

Figure 2 shows the number of CBSAs in which urban growth is faster than suburban growth between 2000 and 2010, for four age groups. A sizable majority of the 50 largest CBSAs register faster urban growth for college-educated 18-24 year olds, 25-34 year olds and 35-44 year olds. However, the 45-64 age group is still suburbanizing, as is the 65+ group (not shown), contrary to the claim that retiring baby-boomers are increasingly likely to choose urban locations.[20] Strikingly, we find that the college-educated 25-34 age group grows faster in the urban area of 23 of the 25 largest CBSAs. The exceptions are Riverside, whose downtown is on a much smaller scaler than that of other large cities, and Detroit. In the smaller CBSAs, however, young professionals are much less likely to be urbanizing.

Figure 2 is suggestive, but it is important to emphasize the magnitudes of these recent trends in the locational preferences of young professionals for the downtowns of large cities. To do so, we compute the aggregate growth of each age-education group, in both urban and suburban areas. We consider the 50 largest CBSAs, in which about 150 million people live. We find that the 25-34 year old college-educated population grew 3.2 times faster downtown, with 44% growth downtown versus 14% growth in the suburbs. Similarly, the 35-44 year old college educated group grew 3 times faster downtown, with 30% growth downtown versus 10% growth in suburban areas. Though our urban areas are small, the aggregate impact of the urban revival patterns we document is not negligible. To show this, we compute the percentage of total young professionals growth that occurs within the urban areas of the 50 largest CBSAs. We find that although urban areas account for 5% of the population (by construction), they account for 24% of the total increase in the college-educated 25-34 year old population and 11.5% of the total increase in the college-educated 35-44 year old population between 2000 and 2010.

Finally, we note that these results are robust to using income instead of education groups, and to using alternative datasets. For instance, the LODES data that we use to estimate our residential-workplace choice model contains data on workers by age group and income group (but not their interactions). Using the same downtown definition, we show that high-income workers in the LODES data are also urbanizing in a sizable majority of large CBSAs between 2002 and 2011. Other work on central city gentrification using census data deliver results consistent with the trends documented here, and for instance Baum-Snow and Hartley (2016) show that downtowns are becoming, richer, more educated and more white, while Edlund show that they becoming more educated and with higher house prices. In the next two subsections, we further characterize urban revival from 2000-2010. We then devote the rest of the paper to explaining these patterns.

[18] For instance, Ihrke et al. (2011) use the 2009 ACS to show that 15.4% of American changed residence over the previous year. This percentage drops to 7% for Americans older than 45.

[19] In an earlier report for CEO for Cities, Cortright (2005) documents the growth in college-educated Americans in "close-in" neighborhoods, defined as all tract within 3 miles from the CBD, from 1990 to 2000. These close-in neighborhoods are relatively similar in size to our downtowns, and our results are robust to using this definition. However, Cortright (2005) uncovers only very uneven trends, because, as we show, urban revival really picks up only in the 2000s.

[20] Recent work by Rappaport (2015) suggests that the aging baby-boomer generation will continue to support strong demand for multi-family units, but posits that these downsizing households will select to remain close to their original locations. This is consistent with our finding that baby-boomers do not contribute to urban revival.

3.2 Urban revival and changing neighborhood composition

In Appendix B, we perform a detailed growth decomposition, to assess the relative importance of changing population density versus changing composition as drivers of urban revival. We decompose the difference between urban and suburban college-educated growth into four components: the change in the college-educated population shares in urban and suburban tracts, reflecting their changing composition, and the change in the urban and suburban populations in urban and suburban tracts, reflecting changing population density.

We find that urban revival in the 50 largest cities is accounted for almost entirely by the rising share of college-educated individuals in urban areas. Strikingly, the urban population in these large cities is stable on average from 2000-2010, while the suburban population is growing. A key feature of the last decade of urban revival is therefore that the change in urban composition is dramatic enough to generate faster young professional growth in urban relative to suburban areas, *despite* stagnant urban and rising suburban populations.

Figure 3 provides a more detailed picture of evolving tract composition at different distances from the city center, and highlights how localized these new trends are. Each plot represents a different age-education group. Panel A shows a kernel density plot of the share of tracts' population accounted for by an age-education group in 2000 at various distances from the city center, across all CBSAs in the sample. Distance from the CBD is in cumulative share of the *total* 2000 CBSA population, and for instance all tracts to the left of the 5% vertical line account for 5 percent of the cumulative CBSAs population (recall that our urban definition is 5th percentile within each CBSA). For instance, Panel A for the 25-34 year old college-educated group shows that they account for a higher share of total tract population closer to the CBD, reflecting their urbanization in 2000 relative to the rest of the population. Panel B shows the changes in tract composition at different distance from the city center from 2000 to 2010. As expected, urban tracts are seeing their composition shift towards higher shares of college-educated 25-44 year olds, and lower shares of non-college educated 25-44 year old. The older cohorts represent a higher share of tracts population at all distances from the CBD, but are growing relatively faster in the suburbs. For the college-educated 25-44 year olds, the fastest composition change is happening very close to the CBD, but increase in their relative shares can be seen beyond the 5% population area that we have defined as being urban in our analysis above.

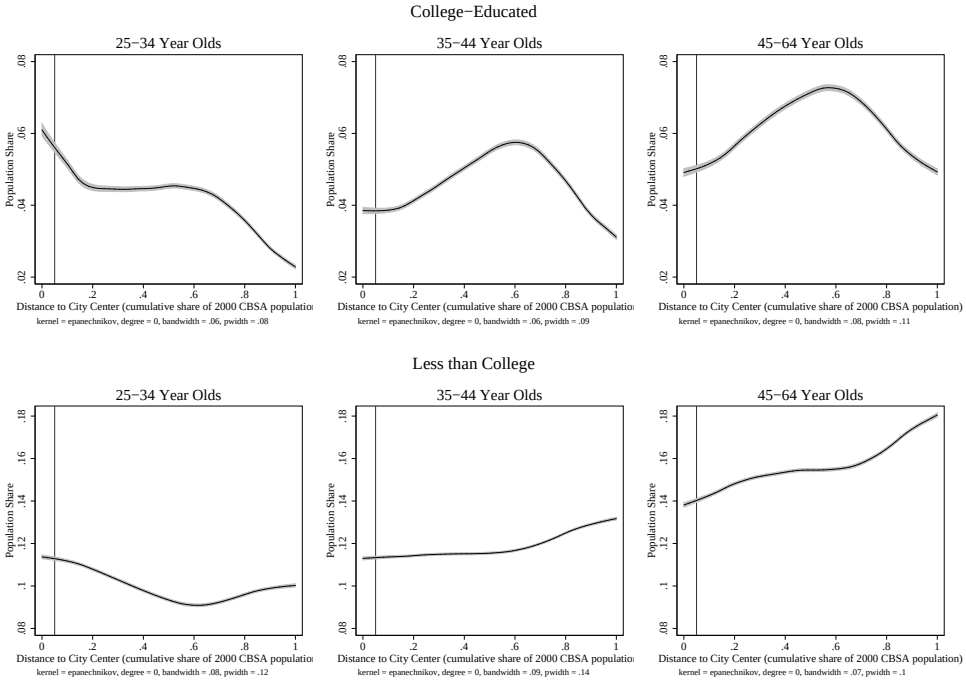
Figure 4 offers a detailed tract mapping in Philadelphia and confirms the trends highlighted in Figure 3. Panels A and B show the population share of 25-34 year old college-educated individuals in all tracts located within the Philadelphia CBSA in 2000 and 2010, where urban tracts are delienated by the area outlined in black. Panel C shows the growth in this share between 2000 and 2010. The left-hand plots show the young-college shares in all tracts in the CBSA, while the right-hand plots show these shares in tracts closer to the CBD. The right-hand plot in Panel A shows that the Center City area of Philadelphia had a relatively high young-college share in 2000 but was surrounded by a ring of tracts with very low shares of young professionals. The right-hand plot in Panel C indicates that the young-college share grew in almost all urban tracts in Philadelphia between 2000 and 2010 and, resulting in the spreading of the downtown area with young-college shares above 10% beyond the center city area in 2010, as can be seen in Panel B. Outside the urban area, we see a mix of tracts where this young-college share has increased and decreased.

3.3 Urban revival and changing commute patterns

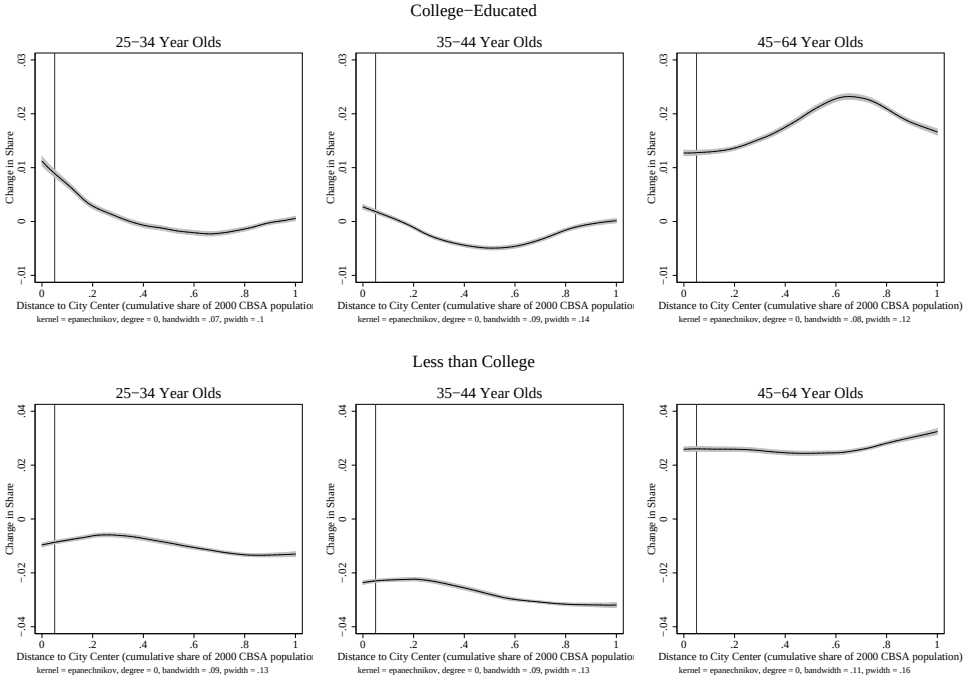
The previous sections document strong centralization trends in the residential choice of younger, college-educated and high-income Americans living in large cities. We now show that while both residences and workplaces are decentralizing in the general population, the reverse is true for high-income people in large cities, whose residences and workplaces are simultaneously centralizing. Crucially, our use of commute data allows us to show that *holding*

Figure 3: Tract Composition Gradients

Panel A: Share of Tract Population in 2000



Panel B: Change in Share of Tract Population (2000 to 2010)

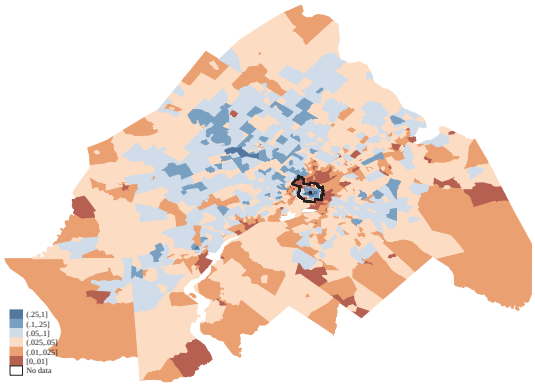


Notes: Panel A shows a kernel of the share of a tract population that is accounted for by each demographic group plotted against the distance of the tract from the city center. Panel B shows a kernel of the change in this share plotted against the distance of the tract from the city center. The distance from city center is measured as the cumulative share of the 2000 CBSA population represented by a given tract as well as all of the other tracts whose centroids are of equal or smaller distance to the CBD. Data from 2000 dicennial census and the 2008-2012 ACS using 2010 consistent-boundary census tracts in all U.S. CBSAs. City centers are from the 1982 Census of Retail Trade Holian and Kahn (2012). Each kernel is estimated weighting tracts by 2000 population.

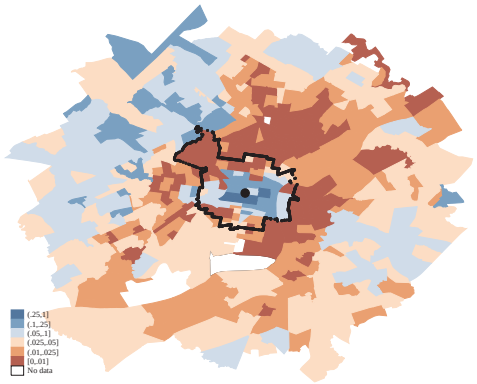
Figure 4: Tract-Level Population Share of 25-34 Year Old College-Educated Individuals//Philadelphia-Camden-Wilmington, PA-NJ-DE-MD CBSA

Panel A: 2000 Level

All Tracts

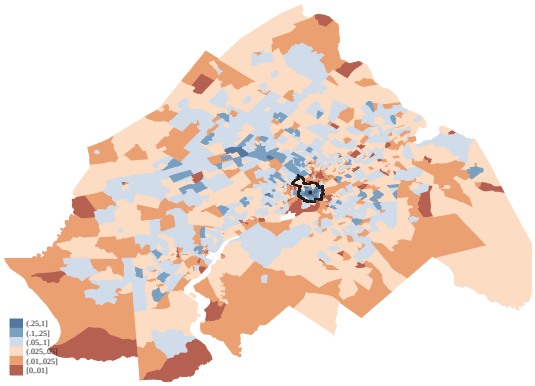


Closer-In Tracts

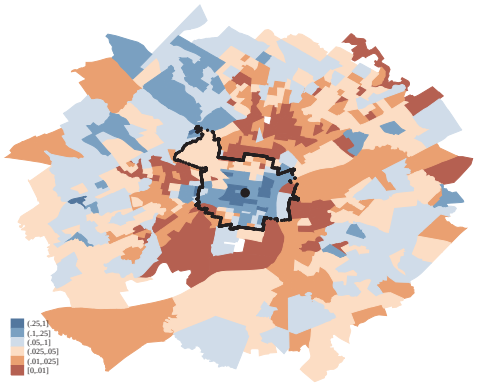


Panel B: 2010 Level

All Tracts

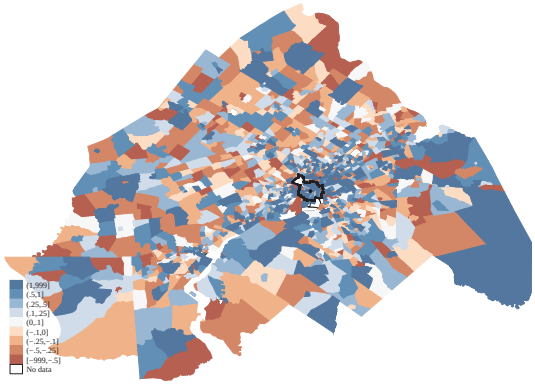


Closer-In Tracts

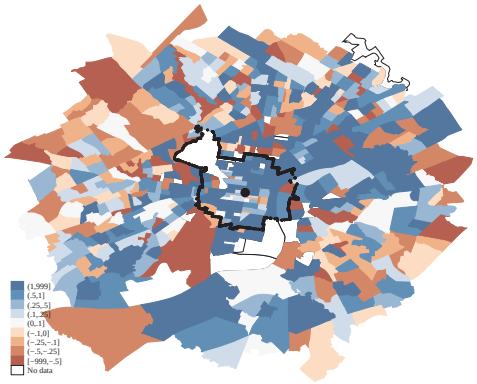


Panel C: 2000-2010 Growth

All Tracts



Closer-In Tracts



Notes: The maps in Panels A and B above reflect the population of 25-34 year old college-educated individuals as a share of the total population in tracts in the Philadelphia-Camden-Wilmington, PA-NJ-DE-MD CBSA. The maps in Panel C show the log change in this population share from 2000 to 2010. The maps on the left-hand side of each panel show the population shares and growth in all tracts in the CBSA, while the plots on the right-hand side of each panel show the population shares and growth for tracts whose centroids are 15km or less from the CBD. Data from 2000 decennial census and the 2008-2012 ACS using 2010 consistent-boundary census tracts. City center is from the 1982 Census of Retail Trade (Holian and Kahn, 2012).

workplace distance from the CBD fixed, high-income workers in large cities live closer to the CBD in 2011 than in 2000, which implies that they incur longer commute costs than before to live near these CBDs. This result suggests that factors other than job centralization also drive residential centralization.

The LODES commute data contain counts of workers by workplace-residential census block pairs, for three income terciles (high, middle, and low-income) and three age groups (29 or younger, 30-54, and 55 or older). We aggregate these data at the census tract level in all of our empirical work. To visualize the data, we aggregate worker counts into commute matrices whose cells are defined by the distance from the CBD of the centroid of the tracts in they live and work. The rows of the commute matrix shows the number of workers *living* within different distance bins from the CBD. The eight bins are: between 0-1 mile, between 1-2 mile, 2-4 mile and so on until 16-32 mile, with an extra cell for individuals living more than 32 miles from the CBD. The column of the commute matrix represent the number of workers *working* at different distance from the CBD, using the same distance bins.

Figure 5 shows the percentage change from 2002 to 2011, for workers within each cell for such commute matrices in Panels a) to c). The color of the cell varies from dark blue indicating the most negative change and dark red indicating the most positive change. The matrix in Panel a) displays data for all workers living and working in each of the 333 CBSAs covered by the LODES data. Looking down each column of Panel a) provides a particularly stark representation of the national suburbanization trends. The workplace distance from the CBD is constant within each column, so a matrix with blue at the top and red at the bottom shows that the American working population is increasingly living in locations further out from the CBD than their workplaces. Residential locations farthest from the CBD (32+ miles) have experienced the largest percentage increase in population, amongst workers working anywhere within 32 miles of the CBD. Looking right to left at each row of Panel a) shows that workers living within 8 miles of the CBD are working farther from the CBD in 2011 than in 2002, but there is no job decentralization trend for residents who have already suburbanized.

The stylized facts in the previous subsections indicate that certain locations and populations have been bucking this national suburbanization trend in the past 10 years. Panel b) focuses exclusively on high-income workers. For this set of workers, we do not see the systematic decentralization of both workplaces and residences that we observed for the entire population, but we do not observe systematic centralization of both workplaces and residences either. We instead observe increases in the number of high-income households either commuting from the suburbs to jobs downtown or from downtown to jobs in the suburbs. Finally, focusing on high-income workers in the 10 largest CBSAs, Panel c) displays commute patterns consistent with our stylized facts. High-income workers are living and working closer to CBDs in large CBSAs. This correlation brings the question of whether high-income workers are following jobs or whether jobs are following them, which is an important identification issue when attempting to explain the urbanization of the young and college-educated. To make progress on this question, it is useful to look at commute patterns within each column, holding workplace location fixed. Recall that, the red cells below the diagonal in panel a) indicated relatively high growth in the population of *all* workers living further from the CBD than where they worked. In panel b), we saw groups of red cells both below *and* above the diagonal. This indicated that high-income workers are increasingly commuting from the suburbs to downtown, with the rest of the population, but also increasingly commuting from downtown to jobs in the suburbs. That is, high-income workers are moving to the downtowns in spite of the commute to jobs in the suburbs. In panel c), where we focus on high-income workers in the largest 10 CBSAs, we see that this increase in reverse commuting is, in fact, stronger than the increase in standard commuting. This demonstrates that factors other than job locations are driving the urbanization of high-income people, particularly in large cities. The attractiveness of large cities' downtowns as residential locations for high-income workers must be increasing, to explain why they are willing to incur larger commute costs than before to live closer to the these CBDs.

Figure 5: 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) **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

(c) **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. CBDs are from the 1982 Census of Retail Trade Holian and Kahn (2012). The top three matrices present national commuting patterns for young workers($\leq$29), middle-age workers (30-54), and old workers ($\geq$ 55); the bottom three matrices present national commuting patterns for low-income workers($\leq$\$1250/month), mid-income workers (\$1250/month-\$3333/month), and high-income workers (>\$3333/month). Given a row, the distance between workplace tracts and CBDs increases from left to right; in each column, the distance between residence tracts and CBDs increases from top to bottom. Each cell represents the percentage change from 2002 to 2011 of the number of certain type of people working and living at given distances from CBDs. Red cells indicate increase in the number of people working and living at given distances from CBDs whereas blue cells indicates small changes, even decrease, in the number of people working and living at given distances. The darker the cell colors are, the more dramatic changes are. Top ten CBSAs are New York-Newark-Jersey City, Chicago-Naperville-Elgin, Dallas-Fort Worth-Arlington, Dallas-Fort Worth-Arlington, Houston-The Woodlands-Sugar Land, Washington-Arlington-Alexandria, Miami-Fort Lauderdale-West Palm Beach, Atlanta-Sandy Springs-Roswell, San Francisco-Oakland-Hayward, and Detroit-Warren-Dearborn.

4 Residential Choice Model

We now specify and estimate a discrete-choice model of residential location. In appendix D we augment this model to study the joint workplace-residential location decision. The model delivers an estimating equation capturing the effect of both changes in the environment (jobs, amenities, and house prices) from 2000 to 2010, and of the initial 2000 levels in these variables on changes in the share of an age-education group living in a given tract. The utility specification is a variant of the random utility model developed by (McFadden, 1972; McFadden and others, 1978))

Each person i of type d chooses its residential location tract j in CBSA c in year t to maximize its indirect utility function V_{jct}^{id}:

$$\max_j V_{jct}^{id} = \alpha_t^d \mathbf{A_{jct}} + \beta_t^d \mathbf{T_{jct}^d} + \gamma_t^d p_{jct} + u_{jc}^d + \xi_{jct}^d + \theta_{ct}^d + \psi_{ct}^{id}(\sigma^d) + (1-\sigma^d)\varepsilon_{jct}^{id} \tag{1}$$

where $\mathbf{A_{jt}}$ is a vector of observable time-varying amenities in tract j, such as accessibility to services and shopping, location within a good school district, or proximity to other young college-educated individuals. $\mathbf{T_{jt}^d}$ is a vector of observable time-varying characteristics of accessibility to jobs in tract j, in particular proximity to job opportunities and average commute length of residents in tract j. $\mathbf{p}_{jt}$ is a house price index in tract j at time t. Preference parameters α_t^d, β_t^d and γ_t^d are time- and group-specific, to capture different preferences across age and education groups, and changes in such preferences. The coefficient on house prices γ_t^d deserves further discussion. The term captures a taste for housing, for instance due to a large family. While we do not derive the indirect utility function in equation 1 from the primitive of a general equilibrium model, the non-stochastic part of this indirect utility can easily be derived - for instance using the monocentric city model in Brueckner et al. (1999) and Duranton and Puga (2015) - from a utility specification that is quasi-linear in housing consumption, in which γ governs the demand for housing. u_{jc}^d represents the unobserved time-invariant quality of each residential location for an individual of type d while ξ_{jt}^d represents a time-varying quality. θ_{ct}^d represents an unobserved time-varying quality of CBSA c for all individuals of type d.

We assume a nested-logit error structure, where $\psi_{ct}^{id}(\sigma^d)$ and ε_{jct}^{id} are random individual- and time-specific taste shocks for CBSA c and residential tract j, respectively. The CBSA taste shocks, $\psi_{ct}^{id}(\sigma^d)$, are independent draws from a random distribution that goes to zero as σ^d goes to zero. The residential tract taste shocks, ε_{jct}^{id}, are independent draws from the extreme value distribution. 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. We derive a linear regression for this nested logit specification as in Berry (1994), modified to use first-differenced data.

We can write the share of type d individuals living in residential location j in year t as the product of the within-CBSA share of individuals living in location j in year t and the CBSA share of individuals in year t:

$$s_{jct}^d = s_{j|ct}^d s_{ct}^d,$$

where

$$s_{j|ct}^d = \frac{\exp\left(V_{jct}^d/(1-\sigma^d)\right)}{D_{ct}^d},$$

and

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

where J_c denotes the set of residential locations in CBSA c, C denotes the universe of CBSAs, $D_{ct}^d = \sum_{j\in J_c}\exp\left(V_{jct}^d/(1-\sigma^d)\right)$ and $V_{jct}^d = \left(\alpha_t^d \mathbf{A_{jct}} + \beta_t^d \mathbf{T_{jct}} - \gamma_t^d p_{jct} + \theta_{ct}^d + u_{jc}^d + \xi_{jct}^d\right)$ denotes the mean utility for an individual of type d

from residential location j in year t. Following Berry (1994), this collapses to:

$$s^d_{jct} = \frac{\exp\left(V^d_{jct}/(1-\sigma^d)\right)}{\left(D^d_{ct}\right)^{\sigma^d} \sum_{c\in C}\left(D^d_{ct}\right)^{1-\sigma^d}}. \tag{2}$$

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

$$\ln s^d_{jct} - \ln s^d_{\bar{j}\bar{c}t} = \frac{V^d_{jct} - V^d_{\bar{j}\bar{c}t}}{1-\sigma^d} - \sigma^d\left(\ln D^d_{ct} - \ln D^d_{\bar{c}t}\right). \tag{3}$$

Substituting $D^d_{ct} = \sum_{j\in J_c}\exp\left(V^d_{jct}/(1-\sigma^d)\right)$ and $\ln s^d_{jct} = \ln s^d_{ct} + \ln s^d_{j|ct}$ into (3) and rearranging terms we have that:

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

Substituting in for the relative mean utility from location j relative to location $\bar{j}$ we have that:

$$\widetilde{\ln s^d_{jc}} = \alpha^d_t \tilde{\mathbf{A}}_{\mathbf{jct}} + \beta^d_t \tilde{\mathbf{T}}_{\mathbf{jct}} - \tilde{\gamma}^d_t p_{jct} + \tilde{\theta}^d_{ct} + \mu^d_{jc} + \tilde{\xi}^d_{jct} - \sigma^d \widetilde{\ln s^d_{j|c}}, \tag{4}$$

where $\tilde{Y}_l = Y_l - Y_{\bar{l}}$ for each variable Y and we normalize $\mu_{\bar{j}c}$ to equal zero.

We estimate the parameters governing these choices using data from 2000 and 2010. Differencing from 2010 to 2000, we obtain our estimating equation:

$$\widetilde{\Delta\ln s^d_{jc}} = \alpha^d_{2010}\Delta\tilde{\mathbf{A}}_{\mathbf{jc}} + \Delta\alpha^d\tilde{\mathbf{A}}_{\mathbf{jc,2000}} + \beta^d_{2010}\Delta\tilde{\mathbf{T}}_{\mathbf{jc}} + \Delta\beta^d\tilde{\mathbf{T}}_{\mathbf{jc,2000}} + \gamma^d_{2010}\Delta\tilde{p}_{jc} + \Delta\gamma^d\tilde{p}_{jc,2000} + \Delta\tilde{\theta}^d_c + \Delta\tilde{\xi}^d_j + \sigma^d\widetilde{\Delta\ln s^d_{j|c}} + \epsilon^d_{jct}, \tag{5}$$

where $\Delta X = X_{2010} - X_{2000}$ for both variables and coefficients.[21]

Note that unobserved time-invariant tract characteristics (e.g., nice weather or architecture) cancel out in first-difference.The error term of this regression is therefore $\Delta\tilde{\xi}^d_j + \epsilon^d_{jkt}$, i.e., the sum of any unobserved changes in the perceived quality of a residential location and an additional term ϵ^d_{jkt} capturing any remaining measurement error. $\Delta\tilde{\theta}^d_c$ is simply a CBSA fixed-effect, to be estimated. Equation 5, which we derived from a discrete-choice model, also delivers an intuitive structural interpretation of regression coefficients that we will use to interpret our results. According to this interpretation, coefficients on changes in characteristics from 2000 to 2010 (e.g. $\Delta\tilde{\mathbf{A}}_{\mathbf{j}}$) capture level of preferences (i.e., α^d_{2010}), while coefficient on initial levels of characteristics (e.g. $\tilde{\mathbf{A}}_{\mathbf{j,2000}}$) capture change in preferences from 2000 to 2010 (i.e., $\Delta\alpha^d$).

4.1 Variable Definitions

In this subsection, we provide details on the computation of our dependent variable, as well as of our measures of amenities, job availability, and house prices. We discuss instruments and identification in the next subsection.

4.1.1 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. We study six different demographic groups indexed by d, consisting of the interaction of three age groups (25-34 year

[21] Note that $\alpha_{2010}X_{2010} - \alpha_{2000}X_{2000} = \alpha_{2010}\left(X_{2010} - X_{2000}\right) + \left(\alpha_{2010} - \alpha_{2000}\right)X_{2000} = \alpha_{2010}\Delta X + \Delta\alpha X_{2000}$

olds, 35-44 year olds and 45-64 year olds) and two education groups (individuals with and without a 4-year college degree). For instance one group is '25-34 year olds with a college degree.' Let n^d_{jct} be the number of individuals of type 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:

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

4.1.2 Amenity Variables

We now describe the variables that we use to measure the amenities available in each tract, $\mathbf{A_{jct}}$, in our main specification. These variables include consumption amenity indexes and local population shares that are included to control for within-group homophily and changes in other endogenous amenities. In robustness exercises, we also include measures of school quality and crime rates, which we observe for only a subset of CBSAs and have no instrument for, as well as access to transit, which we observe only in 2014, and controls for amenity quality, which we can measure only for a very small number of amenity categories. These variables are described in the relevant sections below.

Consumption Amenity Indexes We create variables measuring both the level and change in consumption amenities in each tract. These indexes are based on the price index methodology developed in Couture (2013) and measure the availability of 9 different types of service and retail establishments around the centroid of each census tract. We provide a brief description of our methodology below, and a detailed presentation in Appendix E . Each amenity index is high if there are many establishments of a given type within a short travel time of a centroid. These indexes account for exact establishment location and tract-specific travel times, and are considerably more precise than controls for amenities used in existing studies such as proximity to Central Business District or density of establishments over a large area.

These indexes require the travel times from the centroid of each census tract to the universe of establishments that households might feasibly visit. We employ the NETS database, which contains the exact locations of the universe of U.S. establishments in 2000 and 2010, as well as each establishment's SIC8 industry and name. We compute travel times using results from Google maps searches.

The amenity index for a given category is the inverse of a CES price index, in which the price of visiting an establishment *includes transport cost*. The price of a visit to an establishment is equal to a constant expenditure derived from the Consumer Expenditure Survey, plus a cost of transportation from the tract centroid that assumes a value of time equal to \$12 dollars per hour.[22]We compute indices using travel time by foot, by car and by transit, but only present indices using travel time by foot, which better capture the localization of amenities near a tract and allows to compute stronger instruments, and whose importance is highlighted in the popularity of the Walk Score.

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. We estimate indexes for the following 9 categories of establishments, selected using their SIC8 codes:

- 'Museums, galleries and libraries' (museums, art galleries, libraries, etc.)
- 'Golf and parks' (golf courses and amusement parks)

[22]Some amenity categories like restaurants have a low price per visit (\$10.20) while other categories like apparel stores have a high price per visit (\$60.40). When the price per visit is high, transport costs become a relatively less important factor in the travel decision, and therefore the amenity index puts more weight on establishments far away. As a result, amenity indexes for cheap amenities are more localized.

- 'Gym and sports' (gyms, tennis courts, etc.)
- 'Restaurants' (full-service, fast food, etc.)
- 'Bars' (bar, clubs, lounge, etc.)
- 'Personal Services' (nails, hair, beauty, etc.)
- 'Food Stores' (food stores small and large)
- 'Apparel Stores' (apparel stores)
- 'General Merchandise Stores'

While geographically precise, these amenity indices may hide heterogeneity in amenity quality. In Subsection 5.1 we use marketing data from ESRI to develop a measure of restaurant quality that captures the presence of establishments that young college-educated individuals prefer.

Population Density and Local Demographic Shares in 2000 Recent work demonstrates the relevance of endogenous consumption amenities - amenities that correlate with the share of college-educated residents - in explaining cross-CBSA location choices. We suspect that similar mechanisms are at work within CBSAs, which draw households towards (or away) from neighborhoods depending on their demographic or socio-economic composition. Although we expect our consumption amenity indexes will account for some of these endogenous amenities, we control for any residual effects with control variables for tract population density in 2000 as well as the shares of household in teh same demographic groups in 2000 (homophily).

4.1.3 Jobs Variables

Here we describe the variables that we use to measure the proximity of each tract to employment opportunities, $\mathbf{T_{jct}}$. Being close to work obviously implies a shorter commute, but may also alleviate the location choice problem of dual-career households, and proximity to other job opportunities may become relevant depending on future career events. We measure this in two ways:

Job opportunity index First, we use the LODES data to compute a distance-weighted average of the number of jobs in tracts surrounding each residential tract. We compute this index for three types of jobs: high-income jobs paying more than \$3333 per month, middle-income job paying between \$1000-\$3333 per month, and low-income jobs paying less than \$1000 per month. These three groups correspond roughly to income terciles in 2002. The job opportunity index for a tract j' for income group g is:[23]

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

where $n^g_{j'jt}$ is the number of persons who work in tract j, but do not live in tract $j^{'}$.

Average distance to work Second, we use the LODES data to compute the average commute distance of workers living in tract j, weighted by job numbers in surrounding tracts:

$$\text{avg dist to work}_{jt} = \sum_{j'} \tilde{w}_{jj't} d_{jj'} \text{ where } \tilde{w}_{jj't} = \frac{n_{j'jt}}{\sum_{j'} n_{j'jt}}$$

[23] Note that our job opportunity index is defined over three income groups g, while our dependent variable in our main specification is defined over age-education group d. We do not have measures of jobs by age and education group at the tract level.

Where the job opportunity indexes reflect proximity to jobs in general, the average distance to work variable reflects proximity to the actual employment locations of the households residing in each tract.

4.1.4 Housing Costs

We measure the level and change in housing prices using zipcode level data from Zillow.com, that we match to census tracts. We use the Zillow House Value Index (ZHVI) from 2000 and 2010, which measures the median value of all (non-distressed) properties.[24]

4.2 Identification

There are various challenges to identifying the effect of the variables above on residential location decisions. The first-difference regression allows us to control for time-invariant tract characteristics that could be correlated with our regressors. The vast array of controls alleviates - but does not eliminate - omitted variable bias. Clearly, however, neither first-differencing nor adding controls can resolve reverse causality, which affects variables appearing in changes (Δ). For instance, a change in the share of young professionals in a tract may have a direct effect in attracting consumption amenities and jobs. We therefore instrument for changes in amenity indices, job opportunities, house prices, and the nested-logit within-CBSA share. We describe these instruments briefly below, devoting more time to the most novel of these, for our amenity indices, in Appendix E. We do not have an instrument for changes in crime or school quality, each of which are only available for a limited set of tracts in our sample, or for access to public transit, which we only observe in 2014. We therefore only consider crime, school quality, and transit as controls in robustness checks of the model.

Our regressors for 2000 amenity, jobs, and house price levels do not suffer from reverse causality, and the first-difference controls for correlation with any time-invariant omitted variables. However, omitted variable bias on 2000 levels can remain if some time-varying factors are missing from our regression. This could happen, for instance, if households and businesses move into areas in anticipation of future changes in unobservable factors. To alleviate this concern we instrument for 2000 levels of house prices and local demographic shares.

As a robustness exercise, we also implement another identification strategy using data on worker commutes. The commute data allows to hold constant the workplace location of individuals when estimating the discrete-choice model, thereby convincingly isolating the effect of changes in residential characteristics from changes in job location.

Instruments for Consumption Amenity Indexes Changes in the density and type of local establishments are likely correlated with changing neighborhood demographics. To design an instrument for the change in the consumption amenity indexes, we seek factors explaining variation in amenity location from 2000 to 2010 but exogenous to changes in neighborhood demographics conditional on our set of control variables. Specifically, we exploit variation across firms in their national business expansion strategy in conjunction with spatial variation in the attractiveness of the pre-existing business landscape for establishment entry. This strategy draws both from the Bartik (1991) instrument methodology familiar in labor and urban economics, and from recent evidence from the industrial organization literature on the importance of cannibalization and preemption concerns in determining firm entry (Igami and Yang, 2015; Toivanen and Waterson, 2005).

We first regress SIC8-level (e.g. Korean Restaurant) establishment entry from 2000 to 2010 in a tract on variables capturing moments of the pre-existing commercial environment in 2000. These variables capture the number of establishments in either product space (same SIC8) or related product space (e.g. same SIC6 but

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

different SIC8) at different distance from a tract. When we compute measures of restaurant quality in Section 5.1 we predict entry at the chain-level instead. We obtain predictions for establishment entry at the tract level from the fitted value of these regressions and use these predictions to compute the predicted change in each amenity index. Detailed explanations of this procedure and results from our entry and exit regressions are included in Appendix E. Here, however, it is worth noting that these regression results highlight the local nature of entry and exit decisions and indicate that cannibalization and competition concerns are the key supply-side predictors of entry. We also find evidence on the importance of agglomeration forces in related but not exactly similar product space, and the tendency of establishments, especially within a given chain, to locate in market areas that they have already penetrated, but not in the immediate vicinity of an existing establishment. The best existing business environment for entry appears to be one with other establishments doing closely related business nearby, but none that are in exactly the same market segment.

As a first step to assess the instruments' ability to predict actual changes in our amenity indices, we regress each instrument on actual change the corresponding index. The correlation between the instrument and the actual change in the amenity index is positive and significant at the 5% level in all cases.[25]

A valid instrument for the change in an amenity index must be relevant i.e. correlated with changes in the actual amenity index conditional on all other regressors in equation 5. We report the range of standard first-stage F-statistics alongside our main nested-logit estimation results to demonstrate that this is the case. The instrument must also be exogenous, i.e., uncorrelated with the error term in equation 5 conditional on all other regressors. This exclusion rrestriction upon which this strategy relies deserves discussion. The key feature of this instrument is its exogeneity to changes in local preferences for amenities: all the predictions are obtained using coefficients estimated at the national level, unaffected by any single tract. More generally, the exclusion restriction for a Bartik-type instrument is that the initial distribution of establishments in 2000 is exogenous to changes in the number of establishments in tract from 2000-2010. To the extent that the expansion strategy of different SIC8 categories or chains is driven by supply factor (e.g. sharing fish suppliers for Sushi restaurants) this exclusion restriction is likely to hold. So a first challenge to the exclusion restriction is that changes in demand, for instance changes in national tastes for restaurants, lead some input to the instrument (i.e., the 2000 proximity to different types of restaurants) to affect the dependent variable (changes in residential shares) directly. This concern is alleviated through the vast array of controls for amenity levels already in the regression. Indeed, the conditional exclusion restriction only requires that the instrument for changes in the amenity index be exogenous conditional on the initial density in 2000 around a tract, and similarly that the instrument for restaurant quality be exogenous conditional on initial quality around a tract. A second, related challenge to the exclusion restriction is that changes in demand for features of the 2000 business environment that enter our instrument but are not controlled for in the second-stage regression - for instance changes in tastes for a particular type of restaurant, e.g., Korean - generate a correlation between our instrument and changes in residential share for a given demographic group. Here we note that our instrument for each amenity category relies on aggregating a large number of SIC8 entry and exit predictions, and the coefficients on our prediction variables capture plausibly unbiased variation in the cannibalization, competition, agglomeration and market penetration strategy of different SIC8 or chains. For instance, our instruments predicts faster growth in markets in which existing restaurants have low cannibalization concerns, and the stronger cannibalization concern of Korean relative to French restaurants is plausibly exogenous to changes in relative demand for these restaurants. That being said, one cannot entirely rule out that the strategic behavior of different SIC8 or chain correlates with relative changes in national demand for that SIC8 or chain.

[25] We experimented with creating indices for 2 additional categories: theaters/music venues and movie theater/bowling, but perhaps unsurprisingly we could not predict entry in these types of establishments and removed them from consideration in the paper. So overall the instrument worked in 9 out of 11 amenity categories and for all chain level regressions that we attempted to measure amenity quality..

Instruments for Housing Prices and Local Demographic Shares Changes in housing prices suffer from reverse causality, because they depend on our dependent variable, i.e., the change in the share of each age & education group . House price levels and local demographic shares in 2000 capture many tract characteristics, and are therefore at risk of being correlated with time-varying unobservables. To overcome the endogeneity of house price changes and levels, we exploit the correlation between housing prices, the spatial income distribution, and plausibly exogenous fixed natural amenities identified by Lee and Lin (2013). The idea is that natural amenities (oceans, lakes, mountains, etc.) act like anchors to impose supply constraints on land, whereby driving up relative house price levels, as described in Gyourko et al. (2013). These supply constraints also plausibly amplify the reaction of house prices to demand shocks, so we also use these natural amenities as instruments for changes in house prices.

Such instruments have been criticized in the context of cross-CBSA regressions, for instance by Davidoff (n.d.), who argues that such constraints are correlated with demand factors and shows that constrained cities like New York and San Francisco also have more productive workers. Our within-CBSA instrument is less vulnerable to this criticism, and in this setting a key concern is rather that changing tastes for these natural amenities explain the changing location choices of the demographic groups we study. To deal with this concern we follow Bayer et al. (2007) and instrument for house price changes in a given census tract using the 2000 population-weighted average natural features of census tracts whose centroids are between one and three miles of the centroid of the tract in question, controlling for the average natural features for tracts whose centroids are within one mile as a control. The idea here is that the natural amenities in nearby neighborhoods act as supply constraints in these neighborhoods.. This limits the supply of close substitutes to the houses in the neighborhood in question, whereby driving up prices there. The key exclusion restriction is that the natural features further than one mile away from a census tract does not impact demand for that census tract, conditional on the natural features within one mile of the census tract.

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 and 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.

As an additional instrument for housing prices and an instrument for local demographic shares, we include historical tract-level 1970 population shares for each of the age-education demographic group represented in our analysis. The exclusion restriction here is that, conditional on all controls in the regression, historical shares affect changes in demographic shares only via the mean reversion/within-group agglomeration effects. The key is that historical shares are not correlated with unobserved changes in the characteristics of a residential tract that attract (or detract) households of a given demographic group to (from) that tract between 2000 and 2010 (i.e., $\Delta\tilde{\xi}^{d}_{jc}$).

Instruments for Job-Related Variables We also instrument for changes in the job opportunity (avg num job opp$^{g}_{jt}$) and average distance to work (avg dist to work$^{g}_{jt}$) indexes. We use the same LODES data to obtain Bartik-type predictions for the change in the number of workers in each income group (i.e., high-income, mid-income, and low-income). These predictions depend on the industrial composition of each tract, and on the national growth of jobs in different income group for each industry across 20 NAICS sectors. We index each income group by g and industry by i. The predicted change in the number of group g workers in tract j between 2002 and 2011 is:

$$\widehat{\Delta n^{g}_{j}} = \sum_{i} \left(\frac{n^{g}_{ij2002}}{\sum_{i} n^{g}_{ij2002}} \right) \Delta n^{g}_{i}$$

where n^g_{ij2002} denotes the number of group g workers working in industry i in tract j in 2002, and$\Delta n^g_i = \sum_j (n^g_{ij2011} - n^g_{ij2002})$ denotes the nationwide growth in group g workers in industry i between 2002 and 2011. We can use these predictions to compute instruments for the change in proximity to job opportunities in income group g in tract j':

$$\text{instr}(\Delta\text{avg num job opp}^g_{j'}) = \sum_j w(d_{j'j})\widehat{\Delta n^d_j} \text{ where } w(d_{j'j}) = \frac{1/(d_{j'j}+1)}{\sum_j 1/(d_{j'j}+1)}$$

To calculate the instrument for average distance to work in tract $j^{'}$, we use a weighted average of distance between tracts where the weights are determined by the change in the share of group g workers commuting between those tracts, as predicted above,

$$\text{instr}(\Delta\text{avg dist to work}^g_{j'}) = \sum_j d_{jj'}\Delta\tilde{w}^g_{jj'} \text{ where } \Delta\tilde{w}^g_{j'j} = \left(\frac{n^g_{jj'2002}}{\sum_j n^g_{jj'2002}}\right)\widehat{\Delta n^g_j}$$

Instrument for change in the share of type d individual within CBSA c who live in tract j We now derive instruments for $\Delta s^d_{j|c}$, the change in the share of type d individual within CBSA c who live in tract. We calculate a set of instruments that capture various exogenous factors that affect the attractiveness of tract j relative to all other tracts in a CBSA. For each instrument described above, we compute an instrument for the within-CBSA share, $instr(\Delta s^d_{j|c})$, as the average value of an instrument $instr$ (e.g., instrument for change in job opportunities) in tract j relative to all other tracts k in the CBSA c in which tract j is located. So we compute each instrument as:

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

where $N_{c_j}$is the number of tracts in CBSA c.

4.3 Regression Results

Table 1 presents regression results for the nested-logit model in equation 5, with the full set of instruments described in section 4.2. The model excludes school, crime, and transit for which we have more limited data and therefore relegate to our robustness analysis of in subsection 4.6. In Appendix A.3 we also present the OLS results.

Panel A of Table 1 presents coefficients for the college-educated age groups, while Panel B presents coefficients for the non-college educated age groups. Columns 1 and 2 of Panel A displays parameter estimates for a regression on changes in the share of college-educated 25-34 year olds living in a tract. Coefficients for variables in first-difference (i.e., the change in that variable from 2000 to 2010) are in column 1, while coefficients for variables in 2000 levels are in column 2. Columns 3 and 4 contain results for the same regression for college-educated 35-44 year olds, and column 5 and 6 present results for college-educated 45-64 year olds. The columns of Panel B match that of Panel A. Most coefficients in the table are significant at the 1% level.

Quantitatively interpreting the coefficients on the indexes we have calculated is not straightforward so we present only standardized coefficients. As an example of how to interpret these parameter estimates, the 0.09 coefficient on change in bars for the 24-35 college-educated group means that a one standard deviation increase in the bar index *increases* the share of 24-35 year-old college educated living in this tract by 9%.. The structural interpretation of the coefficient on a variable in first-difference is that of a preference parameter in 2010, while the coefficient on a variable in 2000 level has an interpretation as a change in preference from 2000 to 2010. We often adopt this interpretation in our discussion of the results. Therefore, the positive coefficient on changes in the bar index for the college-educated 24-35 year olds captures a preference for living near this amenity, and the

Table 1: Nested-Logit Residential Location Choice Regression Results

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]
House Price Index	0.04***	-0.01***	-0.05***	0.03***	-0.05***	0.02***
Job Opportunities – Low Inc.	-0.13***	-0.05***	-0.07***	-0.11***	0.00	-0.06***
Job Opportunities – Mid Inc.	-0.03**	0.04***	-0.02	0.12***	-0.02	0.12***
Job Opportunities – High Inc.	0.17***	0.03***	0.15***	0.05***	0.10***	0.02**
Avg. Travel Distance	0.06***	0.03***	-0.02*	-0.01	-0.05***	-0.03***
Museums, galleries and libraries	-0.02***	-0.02***	0.06***	0.04***	0.04***	0.02**
Golf and parks	-0.01**	-0.01**	-0.05***	-0.02***	-0.06***	-0.03***
Gym and sports	0.00	-0.01	-0.02	-0.03*	0.19***	0.22***
Restaurants	0.10***	0.09***	-0.05***	-0.12***	-0.05**	-0.17***
Bars	0.09***	0.09***	0.09***	0.08***	0.02	0.00
Personal Services	0.08***	0.06***	0.17***	0.25***	0.21***	0.26***
General Merchadise Stores	0.00	0.00	0.03***	0.02	0.10***	0.11***
Food Stores	-0.07***	-0.16***	-0.08***	-0.21***	-0.02	-0.15***
Apparel Stores	0.04***	0.05***	0.08***	0.07***	-0.09***	-0.16***
Population Density	-	-0.04***	-	-0.08***	-	-0.20***
Share of Same Type	-	-0.08***	-	-0.07***	-	-0.10***
Within-CBSA share	0.77***	-	0.77***	-	0.62***	-
R-squared	0.90		0.87		0.74	
Observations	36,103		36,122		37,025	

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]
House Price Index	-0.08***	0.00	-0.02**	0.04***	0.00	0.05***
Job Opportunities – Low Inc.	-0.37***	-0.12***	-0.41***	-0.30***	-0.26***	-0.17***
Job Opportunities – Mid Inc.	-0.17***	-0.03	0.43***	0.40***	0.09**	0.11***
Job Opportunities – High Inc.	0.47***	0.09***	0.06*	0.03**	0.10***	0.04***
Avg. Travel Distance	0.31***	0.15***	0.21***	0.11***	0.31***	0.16***
Museums, galleries and libraries	-0.11***	-0.12***	-0.02	-0.06***	0.02	-0.02
Golf and parks	0.04	0.02	-0.22***	-0.11***	-0.13***	-0.06***
Gym and sports	-0.11***	-0.12***	-0.11***	-0.11***	-0.27***	-0.33***
Restaurants	0.10**	0.05	0.27***	0.28***	0.32***	0.29***
Bars	0.03	0.02	-0.04	-0.05***	0.02	0.03
Personal Services	0.26***	0.36***	0.02	-0.03	0.24***	0.27***
General Merchadise Stores	-0.04	-0.05	0.09***	0.11***	0.06***	0.04*
Food Stores	-0.02	-0.07	0.05*	-0.01	-0.01	-0.16***
Apparel Stores	0.09**	0.14***	0.02	0.03	0.12***	0.13***
Population Density	-	-0.20***	-	-0.13***	-	-0.22***
Share of Same Type	-	0.10***	-	0.10***	-	0.02*
Within-CBSA share	0.10**	-	0.51***	-	0.56***	-
R-squared	0.00		0.53		0.32	
Observations	37,391		37,478		37,565	

Notes: * – 10% significance level; ** – 5% significance level; ***–1% significance level. The change in house prices, level of local demographic share, change in consumption, change in job opportunities, change in average distance to work, 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. The F-statistics for these regressions are above 100 for all endogenous regressors, with the exception of the nested-logit within-share, which has a first-stage F-statistics between 19 for 25-34 non-college and 81 for 45-64 college and may be weakly instrumented. Critical values for weak identification tests (Stock and Yogo (2005)) are not readily available for models with more than 2 endogenous regressors.

positive coefficient on the level of the bars index suggests that this taste has been getting stronger over the last decade.The small and generally non-significant coefficient on bars in level and changes for all groups aside from the young and middle-age college-educated indicates indifference or slight aversion to locating near bars for older or non-college-educated groups, that is relatively stable through time. [26]

Overall, most coefficients in Table 1 have the expected sign. Considering coefficient on changes, all groups are attracted to high-income jobs and none want to co-locate with golf courses and amusement parks. The coefficient on change in house prices is negative and significant for 4 out 6 age-education groups and only positive and significant for the 24-35 college-educated group. While a positive coefficient is inconsistent with a structural interpretation of this coefficient as (the negative of) a demand for housing, given how endogenous house prices are to increases in demand, and particularly, college-educated demand, for living in a tract we take this success rate in estimating negative coefficients as supporting the specification.

It is worthwhile noting that the coefficients on the jobs variables - both levels and changes - take some of the largest estimated values for most demographic groups. Conditional on the approximate distance to their own place of employment (represented by the average travel distance variable), all groups tend to want to live in proximity to high-wage jobs and all groups but one significantly want to locate away from low-wage jobs. This common taste is perhaps less surprising if you recall that the "high-wage" category refers to jobs with monthly wages of only $3333 and above, so these job opportunities may be relevant for high-school as well as college-educated individuals. Nevertheless, proximity to jobs - both current and potential - is an important determinant in the residential location decisions of households. This does not, however, imply that jobs are the most important factor explaining the redistribution of household types between downtown and the suburbs. We will see below that amenitiestend to play a more important role in explaining the whether households choose to live near to, or far from, the CBD.

Finally, we note the difficulty of assessing whether the signs of coefficients in 2000 levels are reasonable or not. In subsection 4.6.4, we evaluate the plausibility of our interpretation of these coefficients as capturing changes in preferences using data on changes in expenditures and trip shares from 2000 to 2010 for a small number of amenity categories. For instance, the strong tendency of coefficients in levels and changes to have a similar sign also emerges, albeit to a lesser extent, from our analysis of travel and expenditure data.[27]

4.4 Does the model explain urban revival?

In this section, we establish the model's ability to generate our stylized facts from Section 3 on the fast growth of young-college educated individuals near the CBD of large cities. First, we simply attempt to use the model's fitted values to replicate the pattern of urban vs. suburban growth by city size for each age-education group documented in Figure 2. Second, we perform a more detailed examination of the model's performance and assess its ability the replicate the change in tract composition at different distances from the CBD from Figure 3. After establishing the model's ability to fit these two sets of moments, we assess the importance of each factor (amenities, house prices, jobs, etc.) in explaining urban revival by measuring the contribution of each variable towards this fit.

It is worth emphasizing that we do not perform out-of-sample predictions, and that we fit our model with the same data that we use to estimate it. However, fitting the model is interesting because our regressions do not include any controls for either distance from the CBD or for city size. The goal of this exercise is therefore to ask

[26] Note that we do not expect every group to have a preference for living near every amenity category. Built amenities that one rarely visits are probably dis-amenities, and indeed most neighborhoods have zoning regulations preventing commercial use in the vicinity of residential areas.

[27] One worry with this IV specification is that the high correlation between coefficients in levels and changes stems from the high correlation between the level and change for a given variable in the data. This could happen if levels are an important exogenous predictor of changes in the first stage regression. We note that the instrument is almost always a more important predictor of the change in amenities than the initial level, generally by a factor two or three. Moreover, OLS results should be less affected by this correlation, and they also highlight the importance of levels of bars in restaurants in pushing young professionals downtown.

whether the variables included in the model capture the special characteristics of the downtown of large cities that made them ripe for urban revival over the last decade.

Replicating urban vs. suburban growth across CBSAs The first step in replicating our stylized facts is to derive urban and suburban growth from the fitted model. We start from the fitted value of the regression in equation 5, to obtain the fitted change in the share of group d who lives in tract j from 2000 to 2010, relative to a base tract. We always exclude the term for change in within-CBSA share (i.e., the 'nested-logit' term) from this fitted value, because it has explanatory power by construction. Starting from this fitted value, we easily obtain the fitted 2010 share of group d in tract j, by differencing out the actual share change in base tract, and the actual initial share in 2000. We then recover the fitted *population* of group d living in tract j in 2010 by multiplying this fitted share by the total population of group d. With fitted 2010 tract population of group d in hand, we can compute urban and suburban growth since 2000 within each CBSA exactly as in Section 3.

Figure 6 compares model-generated urban vs suburban growth obtained using fitted 2010 population with actual urban vs. suburban growth obtained using actual 2010 population. We use fitted values from the nested-logit model with instrumental variables shown in Table 1. The bar charts in Figure 6 are similar to those of Figure 2 from the stylized facts, with the blue bars representing the number of CBSAs in which urban growth is faster than suburban growth, and the green bars representing the number of CBSAs in which suburban growth is faster than urban growth.[28] We again group CBSAs by population size bins, but this time we define five bin sizes: very large (>3M), large (1.5-3M), medium-sized (0.5-1.5M), small (0.3-0.5M) and very small (<0.3M) CBSAs. Panel A displays model-generated growth on the left and actual growth on the right, for the 24-35 year-old college-educated group. The model clearly captures the urbanization of young professionals, and generates faster urban than suburban growth in large and very large cities. The model also successfully explains our finding that very large and large cities almost all experience urban revival, while outcomes are much more uneven in smaller cities.

In Panel B we produce the same histogram, but for 25-34 non-college-educated individuals, a group that is not urbanizing. In this case, the model correctly generates faster suburban growth in cities of all sizes. Interestingly, it also captures the better performance of smaller cities' downtowns for this group. The model performs equally well for older cohorts (not shown).

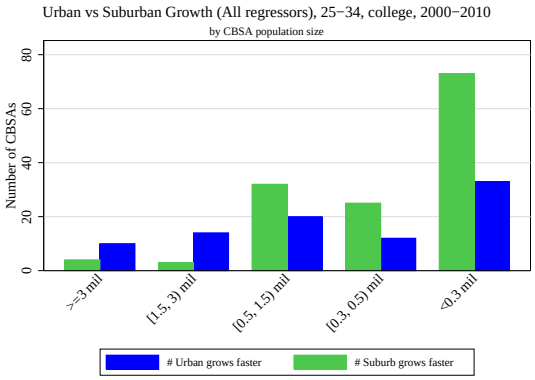
Figure 7 documents the model's ability to explain *which* CBSAs experience urban revival, as opposed to explaining outcomes within city size bins. The results demonstrate that our regressions capture the factors explaining the urbanization of the young and college-educated in large cities, but perform unevenly for smaller cities. For each CBSA and each demographic group, we compute the difference in their urban and suburban growth using both the model's fitted 2010 population counts and the the actual 2010 population counts. In Figure 7, we plot the fitted values against the actual values separately for CBSAs with more than 1.5 million people and for those with less than 1.5 million people. We find the strongest correlation between fitted and actual values for 25-34 year olds (both college and non-college-educated) in CBSA with more than 1.5 million people. Our model performs reasonably well for older cohorts, but less well for the middle-aged - for whom changes in patterns of location change have been less dramatic over the last decade - and in CBSAs smaller than 1.5 million. As shown in the bar charts in Figure 6, the model roughly matches the relative number of smaller CBSAs in which the urban population is growing faster than the suburban population, but the low correlations in Panel B of Figure 7 reflects the model's failure to capture *which* small CBSAs are seeing faster urban or suburban growth for each demographic group.

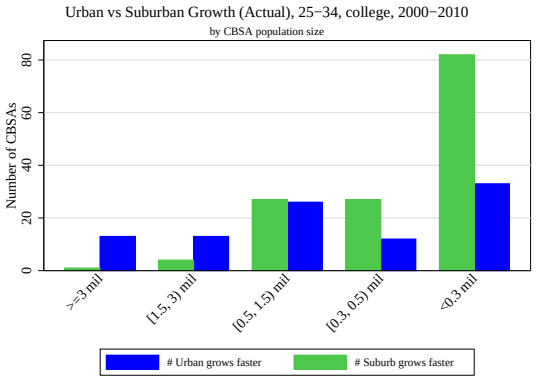
Replicating changing tract composition gradients We now evaluate the model's performance in capturing changes in tract composition at different distances from the CBD. Figure 8 provides, side-by-side, the kernels of

[28] Note that the set of CBSAs is Figure 6 is not the same as that in the stylized facts of section 3, because some CBSAs drop out of our sample because of data issue, mostly due to lack of house price data.

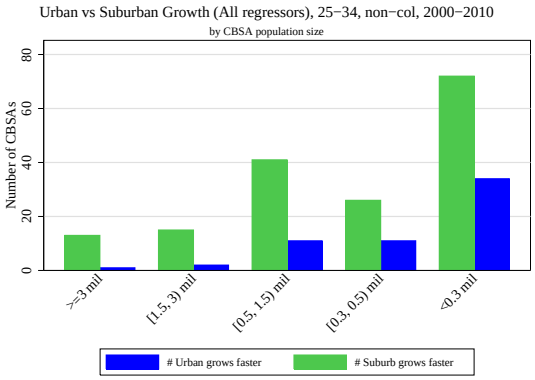
Figure 6: Predicted vs Actual Urban-Suburban Growth: 25-34 year olds

Panel A: College-Educated 25-34 Year Olds





Panel B: Non-College-Educated 25-34 Year Olds



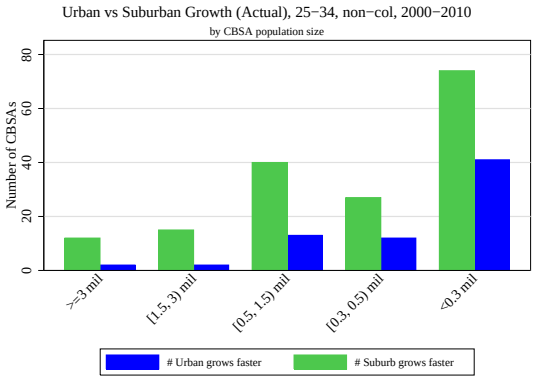
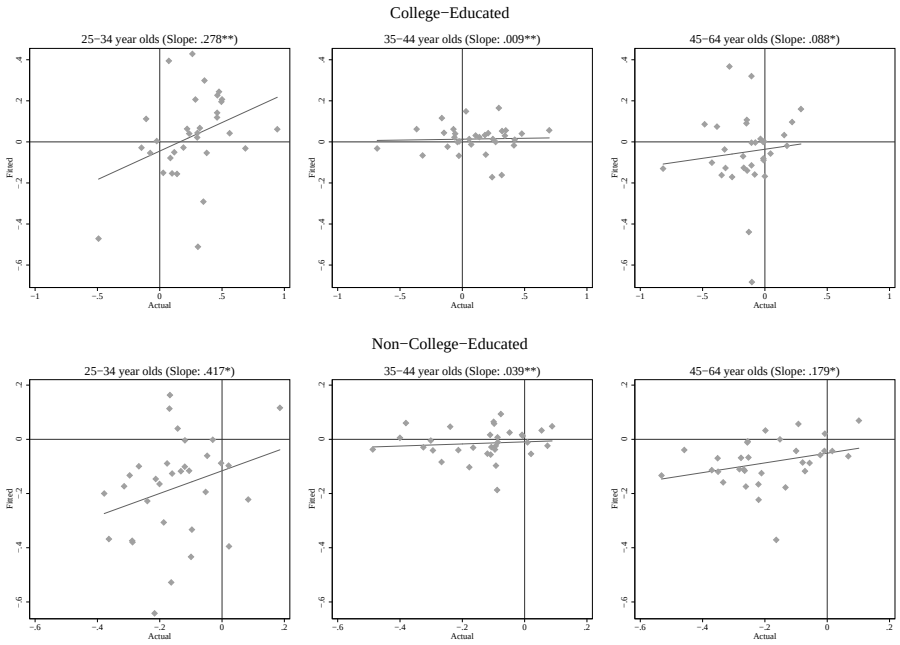
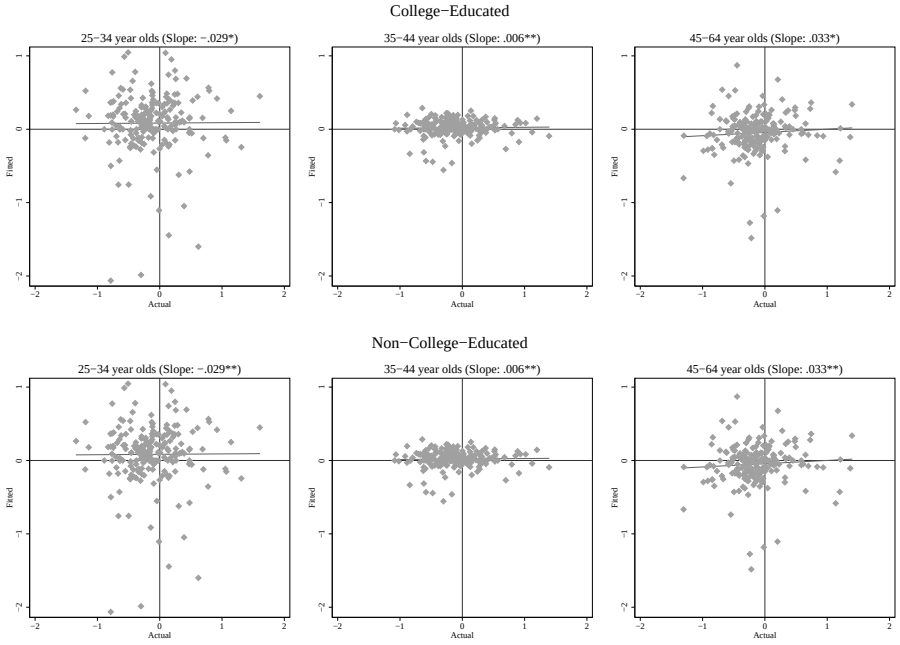


Figure 7: Predicted vs. Actual Log Relative Urban-Suburban Population Growth

Panel A: CBSAs with Population Over 1.5 million



Panel B: CBSAs with Population Under 1.5 million



Notes: CBSAs with actual log growth greater than 2 are dropped.

the change in tract-level shares of each group at different distances from the CBD obtained from the model, and the actual kernels already presented in Figure 3. This figure implicitly weights large CBSAs more heavily, so given results on large cities from Figures 6 and 7 above, it is not surprising that we match the general trends for the younger cohorts. Interestingly, our model also does quite well at fitting the U- and inverted-U shaped patterns observed for the older cohorts of the college educated.

The fit, however, is not perfect: our model counter-factually suggests that the share of 45-64 year old non-college-educated individuals is growing faster downtown than in the suburbs. Overall though, our regressions pick up the general qualitative trends in the location decisions of the young and college-educated relative to other cohorts.

.

4.5 Which variables explain urban revival?

In this section, we investigate the ability of the the coefficients in Table 1 to inform our understanding of the factors driving urban revival. We start by providing a simple breakdown of the variables, highlighting those with the most ability to explain the urbanization of college-educated 25-34 year olds (Figure 9). We then consider all age-education groups, and provide a precise assessment of the ability of different groups of variables to predict the relative composition of tract at various distances from the city center highlighted in section 3.2.

For a variable to explain the urbanization of the young and college-educated, it must take a larger (smaller) value in urban relative to suburban areas, and its coefficient must be positive (negative) for young college-educated individuals. For instance, urban areas are denser, and therefore usually have higher 2000 levels of amenity indexes for non-tradable services like restaurants, bars and personal services, and the coefficients on these levels in Table 1 is positive for young professionals. Column 1 in Figure 9 contains *non-standardized* coefficients for the 25-34 year-old college-educated group for all variables other than the within-CBSA share which, by construction, is correlated with the observed changes in urbanization rates. Columns 2 and 3 contain the standardized mean value of different variables in urban (downtowns) and suburban areas. Combining this information is useful because variables explaining the urbanization of young professionals must have both a large coefficients and a large difference in mean value between urban and suburban areas. Column 4 contains the product of column 1 and the difference between column 2 and 3. This product allows us to rank the variables from those making the largest contribution to explaining the urbanization of young professionals (on top, in green), to those delivering the strongest push in the opposite direction (at the bottom, in red).[29] Three of the five variables with the strongest push are 2000 levels of highly urbanized non-tradable service amenities i.e. bars, restaurants and personal services. Apparel stores are also in the top five. The level and change in high-wage jobs are in position six and seven, respectively, because, although young college-educated are attracted to high-wage jobs, the relative increase in high-wage jobs in urban areas has been relatively modest over the last decade and their attractiveness (the coefficient on the level) has been relatively stable. Also apparent from Panel A is that many amenity variables in first-difference, including bars, restaurants and personal services, work against the urbanization of the college-educated, because such amenities have been growing (very slightly) faster in suburban areas over the last decade. We also note that a few variables like the level of bars explain not only the urbanization of young professionals, but also their urbanization relative to all other groups which have had stable or declining preferences for it.

An even sharper variable classification could look at variables that not only explain the urbanization of the young and college-educated relative to other groups, but also the relative salience of this trend in large relative to small cities. Unfortunately, the model does not deliver any such variables. Amenities, for example, are relatively

[29] Given that all variables are logged, the interpretation of the coefficient in column 4 is the difference in the share of young college-educated living in a tract - relative to a base tract - due to the urban-suburban difference in that variable.

Figure 8: Predicted vs. Actual Urban-Suburban Tract Composition Change Gradients

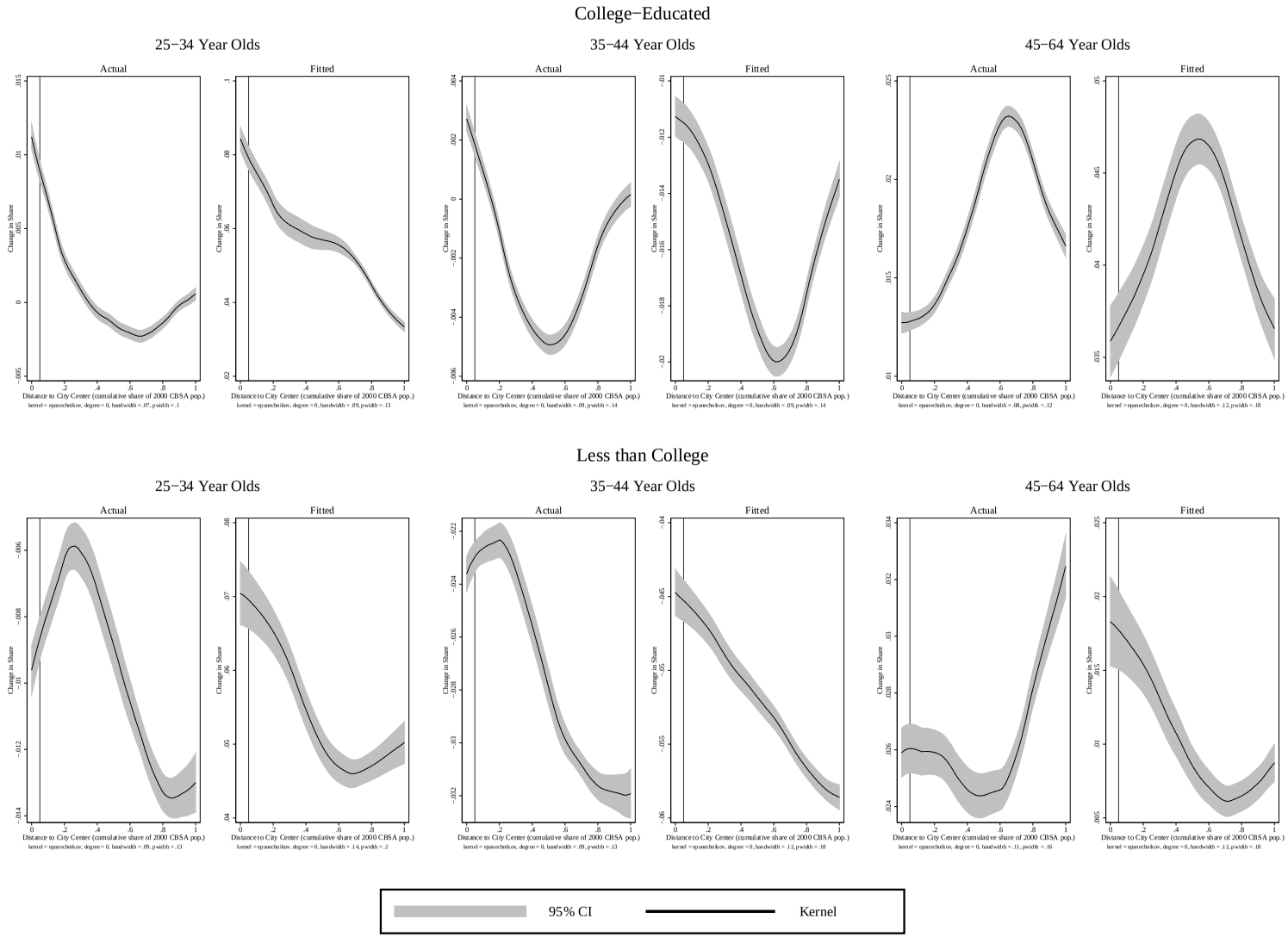


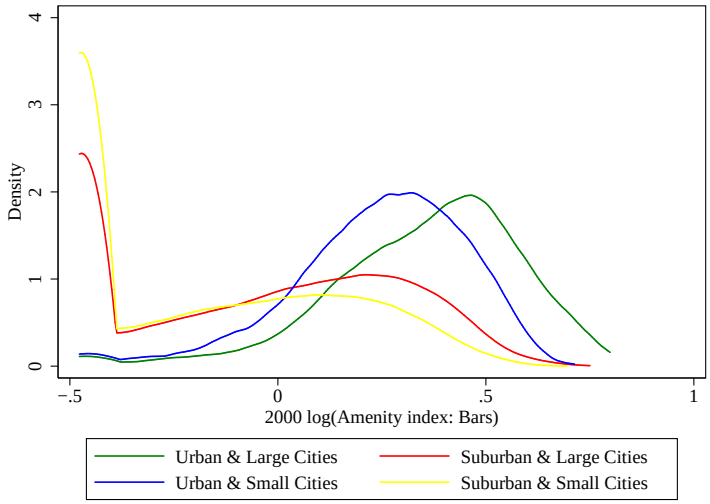
Figure 9: What variables explain the urbanization of the young and college-educated?

		Variable Mean		
	Coefficient	Suburban	Urban	Contribution
Variable	[1]	[2]	[3]	[4] = [1]*([3]-[2])
Bars	0.22***	-0.10	0.31	0.09
Restaurants	0.2***	-0.19	0.16	0.07
Personal Services	0.17***	-0.26	0.01	0.05
Apparel Stores	0.17***	-0.05	0.22	0.05
Job Opportunities -- Mid Inc.	0.09***	0.19	0.44	0.02
Job Opportunities -- High Inc.	0.05***	0.08	0.36	0.01
Δ Job Opportunities - High Inc.	0.37***	-0.11	-0.08	0.01
Δ Food Stores	0.7***	-0.03	-0.01	0.01
Δ House Price Index	0.13***	0.04	0.10	0.01
Δ Job Opportunities - Low-Inc.	-0.39***	-0.08	-0.09	0.01
House Price Index	-0.02***	-1.00	-1.23	0.00
General Merchadise Stores	-0.01	-0.01	-0.21	0.00
Δ Job Opportunities -Mid Inc.	-0.07**	-0.16	-0.18	0.00
Δ Gym and sports	0.02	-0.03	-0.02	0.00
Δ General Merchadise Stores	-0.04	0.08	0.09	0.00
Δ Golf and parks	0.14**	0.00	0.00	0.00
Golf and parks	0.06**	0.00	-0.02	0.00
Δ Museums, galleries and libraries	0.16***	-0.07	-0.08	0.00
Share of Same Type	-0.07***	0.33	0.37	0.00
Δ Apparel Stores	-0.38***	-0.07	-0.06	0.00
Avg. Travel Distance	0.04***	0.43	0.32	0.00
Δ Avg. Travel Distance	0.05***	-0.82	-0.91	0.00
Δ Bars	-0.41***	-0.04	-0.03	-0.01
Gym and sports	-0.03	-0.04	0.16	-0.01
Δ Restaurants	0.58***	0.10	0.08	-0.01
Δ Personal Services	0.54***	0.22	0.19	-0.01
Job Opportunities -- Low Inc.	-0.11***	0.07	0.26	-0.02
Population Density	-0.02***	-0.90	0.54	-0.03
Museums, galleries and libraries	-0.1***	-0.19	0.11	-0.03
Food Stores	-0.54***	-0.17	0.10	-0.14

more urbanized in small cities. Figure 10 illustrates this for the 2000 level of the bars amenity index. The plot shows the distribution of this index within four sets of tracts: urban areas of large cities, urban areas of small cities, suburban areas of large cities, and suburban areas of small cities. In 2000, this index is higher in urban relative to suburban areas but not relatively moreso in large cities. These patterns explain the uneven performance of our model for predicting the urbanization rates of the young and college-educated in small cities, as shown in the previous section.

.

Figure 10: 2000 Log(Bars Amenity Index)



We now turn to studying how each factor in our residential location choice framework contributes to the curvature of the moments that we fit in Figure 8. We first outline how these fitted moments are calculated. We start with the fitted value from equation 5 for the expected change in the share of the total national population in a given demographic group d that resides in tract j relative to the share of that demographic group that resides in our base tract l:

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

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

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

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

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

We divide this fitted population for demographic group d by the observed total population of tract j in 2010 to arrive at the share of tract j's population in demographic group d in 2010. We then difference this 2010 fitted level

from the observed value of this share in 2000 to get a fitted prediction of the change in population share that is represented in the composition plots:

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

Putting this together and rearranging terms we have that:

$$\widehat{\Delta s_{d|jc,2010}} = \left(\frac{pop^{d}_{jc,2000}}{pop_{jc,2000}}\right)\left(\underbrace{\left[\prod_{k}\exp\left(\widehat{\delta^{d}_{k}}X_{jc,k}\right)\right]}_{\text{Estimated Scaling Factors}}\underbrace{\exp\left(\widehat{\Delta\ln s^{d}_{lc}}\right)\left(\frac{pop^{d}_{2010}}{pop^{d}_{2000}}\right)\left(\frac{pop^{all}_{jc,2010}}{pop^{all}_{jc,2000}}\right)^{-1}}_{\text{Observed Changes in Population Shares and Levels}} - 1\right) \quad (6)$$

Equation 6 shows that the contribution of any single regression factor, X_k, to the spatial distribution of demographic group d across tracts j (and, therefore, to the change in each demographic group's population share in each tract) depends on a scaling factor $\exp\left(\widehat{\delta^{d}_{k}}X_{jc,k}\right)$, where $\widehat{\delta^{d}_{k}}$ is the estimated non-standardized regression coefficient on tract characteristic $X_{jc,k}$. The change in demographic group d's population share of tract j from 2000 to 2010 is determined by the product of each of these scaling factors, the change in demographic group d's population share of the base tract l from 2000 to 2010, and the change in the aggregate population of demographic group d from 2000 to 2010, divided by the change in the tract j population from 2000 to 2010.

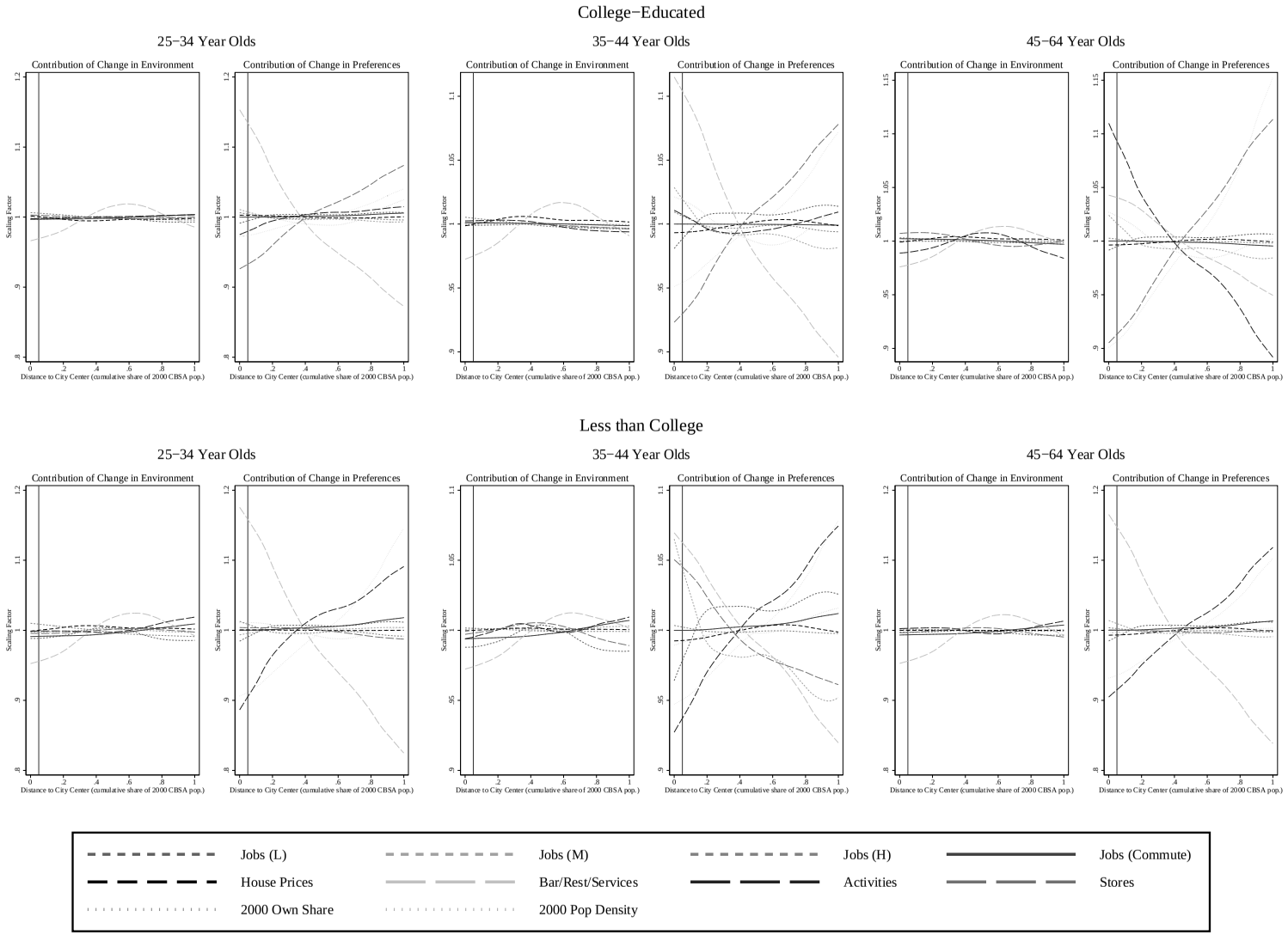
Figure 11 shows estimated kernels of these scaling factors for each age-education group. The plots on the left-hand side for each age-education group depict the kernel of the scaling factors associated with the change in environment ($\widehat{\alpha^{d}_{2010}}\Delta\tilde{\mathbf{A}}_{\mathbf{jc}}$, $\widehat{\beta^{d}_{2010}}\Delta\tilde{\mathbf{T}}_{\mathbf{jc}}$, and $\gamma^{d}_{2010}\Delta\tilde{p}_{jc}$), while those on the right-hand side depict the scaling factors associated with the change in preferences ($\widehat{\Delta\alpha^{d}}\tilde{\mathbf{A}}_{\mathbf{jc,2000}}$., $\widehat{\Delta\beta^{d}}\tilde{\mathbf{T}}_{\mathbf{jc,2000}}$., and $\Delta\gamma^{d}\tilde{p}_{jc,2000}$). The plots all share the same scale, so it is easy to see the relative importance of the change in preference variables in driving the changing composition of census tracts near and far from U.S. CBDs. We have grouped amenity variables into three curves: non-tradable services (including bars, restaurants, and personal services), activities (including gym and sports, golf and parks, and museums, galleries, and libraries), and retail stores (including apparel stores, general merchandise stores, and food stores).

Overall, the results for the college-educated 25-34 year olds here reflect the sum of the patterns identified in the urban/suburban analysis in Table 9. The increasing tastes for non-tradable services (bars, restaurants, and personal services) are the biggest contributing factor in the increase in the young-college shares of downtown tracts. In fact, these amenities are the strongest urbanizing factor for all demographic groups except for college-educated 45-64 year olds, for whom activities play a more important role. The distaste for retail stores is also common across all household types, but the varying degree of this distaste is a key factor that generates the different curvatures of the fitted moments in Figure 8. The relative dominance of the changing tastes for these two groups of amenities is the most pronounced for the college-educated 25-34 year olds. Other groups display relatively more pronounced changes in their preferences for other environmental factors such as proximity to jobs and population density, which plays an important suburbanizing force for all but the young and college-educated.

4.6 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.

Figure 11: Factors Contributing to Predicted Urban-Suburban Tract Composition Change Gradients



4.6.1 Crime and School Quality

School quality and crime rates are presumably important determinants of location choices in the United States. The well-documented decline in central city violent crime since 1990 (e.g., Levitt 2004) is a potential explanation for the reurbanization of young college-educated Americans, and much anecdotal evidence suggests that school quality is a key determinant of the suburban location choice of families with children, though there is little evidence of a relative improvement in urban relative to suburban schools and notable work by Bayer et al. (2007) indicates that household willingness-to-pay for school quality has been over-estimated in the past.

In Table 2 we report coefficients for an IV nested-logit regression adding controls for the change and levels in local school district rankings and per capita violent crime to our base specification presented in Table 1.[30][31] We note that including school and crime variables reduces our tract sample size by more than two thirds and that we cannot instrument the endogenous change in these variables.

The signs of the coefficients on crime and school rankings largely match the expectation that crime is a disamenity and that school quality (measured as the inverse school district ranking within-state) is an amenity to all but 25-34 year olds, who are less likely to have school-aged children. The standardized coefficients on changes in crime, denoting an aversion to crime, are all relatively small and negative. The young college-educated group stands out with a coefficient that is much smaller and not significant, denoting little aversion to crime for that group. Using our usual interpretation of the coefficient on levels as capturing a change in preferences, we conclude that aversion to crime is relatively stable from 2000 to 2010, and again the young college-educated group stands out as the only group experiencing a significant reduction in aversion to crime.

Turning our attention to schools, we estimate positive coefficients on the change in school quality for both 35-44 and 45-64 college- and non-college-educated groups, indicating a positive preference for highly-ranked schools. The coefficient on the change in school ranking variable for college-educated 25-34 year olds is not significantly different from zero indicating that schools are not a significant factor in the location decisions of this cohort, while schools appear to be an increasingly significant disamenity for their non-college educated counterparts. Interestingly, the standardized coefficients on the 2000 level of school quality are negligible for the young and middle-aged college-educated groups, indicating that changes in preferences for schools due to, perhaps, the delay of childbirth, was not an important factor in the changing location decisions of these households between 2000 and 2010.

While the small size of the coefficients on crime and school variables may be due to measurement error on these imperfectly measured variable, a number of factors weight against the importance of improvement in urban schools and reduction in urban crime as drivers of urban revival. In particular, the group that is urbanizing most rapidly over the last decade, the college-educated 25-34 year olds, is also the only group with little aversion to crime and no preferences for highly ranked schools. This young, college-educated group does experience a significant reduction in its aversion to crime from 2000 to 2010, but in this case the model suggests that it is this reduction in aversion to crime that contributes to their urbanization, rather than a decline in urban crime. Violent crime did decrease in urban relative to suburban areas over the last decade, but it only significantly contributes to the urbanization of groups other than young professionals. Moreover, others (e.g., Kneebone and Garr, 2010) have documented that this decrease in crime was even faster in the 1990s, a period over which the widespread urban revival that we document is not happening.[32] Finally, our data shows that average school quality has not

[30] SchoolDigger.com compiles test scores for schools all over the United States and provides a ranking of each school district within each 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 2004 - the earliest year available - and in 2010 in the school district that a tract falls into as our measure of school quality in 2000 and 2010, respectively.

[31] We measure per capita violent crime in a tract as 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.

[32] Ellen and O'Regan (2010) show that although higher urban crime rates in the 1970s and 1980s may have contributed to population

Table 2: Nested-Logit Residential Location Choice Regression Results
Including School and Crime

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]
House Price Index	0.07***	-0.03***	0.01	0.04***	-0.03***	0.02**
Job Opportunities – Low Inc.	-0.05***	-0.06***	0.01	-0.14***	-0.01	-0.10***
Job Opportunities – Mid Inc.	0.05***	0.08***	0.07***	0.12***	0.08***	0.10***
Job Opportunities – High Inc.	0.07***	0.04***	0.04***	0.07***	0.05***	0.07***
Avg. Travel Distance	0.01	0.00	0.00	-0.01	-0.03*	-0.03***
Museums, galleries and libraries	-0.06***	-0.06***	-0.03*	-0.04***	-0.03*	-0.04***
Golf and parks	-0.02	-0.01	-0.05**	-0.02	-0.06**	-0.02*
Gym and sports	0.09***	0.11***	-0.06***	-0.08***	0.11***	0.11***
Restaurants	0.20***	0.17***	0.10***	0.08***	-0.02	-0.10***
Bars	0.03*	0.02	0.13***	0.12***	0.07***	0.07***
Personal Services	0.04*	-0.04	0.00	-0.04	0.14***	0.17***
General Merchadise Stores	-0.01	0.01	0.03**	0.03	0.10***	0.10***
Food Stores	0.01	-0.05	0.02	-0.01	-0.07***	-0.21***
Apparel Stores	-0.04**	-0.03	0.07***	0.07***	0.06***	0.03
Violent Crime	0.00	0.02***	-0.02***	0.00	-0.01***	0.00
School Pctile Ranking	0.00	-0.01	0.02***	0.01	0.01***	0.03***
Population Density	-	-0.08***	-	-0.10***	-	-0.16***
Share of Same Type	-	-0.09***	-	-0.03**	-	-0.11***
Within-CBSA share	0.70***	-	0.66***	-	0.54***	-
R-squared	0.84		0.82		0.72	
Observations	11,179		11,114		11,462	

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]
House Price Index	-0.07***	-0.06***	0.03**	-0.02*	0.06***	-0.07***
Job Opportunities – Low Inc.	-0.03	-0.06**	-0.12***	-0.15***	-0.06***	-0.13***
Job Opportunities – Mid Inc.	0.07**	0.06**	0.25***	0.17***	0.15***	0.08***
Job Opportunities – High Inc.	0.08***	-0.01	0.00	0.03*	0.04**	0.12***
Avg. Travel Distance	0.18***	0.07***	0.10***	0.03***	0.21***	0.08***
Museums, galleries and libraries	-0.19***	-0.16***	-0.11***	-0.10***	-0.18***	-0.19***
Golf and parks	-0.20***	-0.10***	-0.13***	-0.06***	0.00	0.00
Gym and sports	0.04	0.03	0.07***	0.08***	0.01	-0.01
Restaurants	0.11	0.01	0.11***	0.13***	0.26***	0.17***
Bars	0.06	0.05	-0.02	-0.01	0.01	0.01
Personal Services	0.18***	0.22***	-0.08***	-0.15***	0.18***	0.17***
General Merchadise Stores	-0.15***	-0.16***	-0.05***	-0.05**	0.13***	0.17***
Food Stores	0.01	-0.10	-0.04	-0.08**	0.02	-0.08
Apparel Stores	0.16***	0.20***	0.18***	0.21***	0.01	-0.03
Violent Crime	-0.05***	-0.03*	-0.03***	0.00	-0.02***	-0.01
School Pctile Ranking	-0.05***	-0.06***	0.01	-0.05***	0.01*	-0.04***
Population Density	-	-0.10***	-	-0.05***	-	-0.22***
Share of Same Type	-	0.07**	-	0.11***	-	-0.07***
Within-CBSA share	0.18***	-	0.63***	-	0.57***	-
R-squared	0.16		0.74		0.48	
Observations	11,635		11,657		11,675	

Notes: * – 10% significance level; ** – 5% significance level; ***–1% significance level. The change in house prices, level of local demographic share, change in consumption, change in job opportunities, change in average distance to work, 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. The F-statistics for these regressions are above 100 for all endogenous regressors, with the exception of the nested-logit within-share, which has a first-stage F-statistics of 32 (for 25-34 college) and 60 (35-44 college) and may be weakly instrumented. Critical values for weak identification tests (Stock and Yogo (2005)) are not readily available for models with more than 2 endogenous regressors.

improved in urban relative to suburban areas over the last decade, meaning that schools may indeed contribute to the continuing suburbanization of groups other than young professionals.

The coefficients in Table 2 are in general qualitatively similar to earlier results from Table 1, and most of the change in coefficients are due to the large change in sample, not to the addition of endogeneous variables.[33] Kernel plots highlighting the contribution of different categories of variables to tract composition deliver results qualitatively similar to those in Figure 11. Figure 12 shows that bars, restaurants and services still provide the strongest contribution to changing composition of urban tracts towards 25-34 year old college-educated even after controlling for school quality and violent crime.

4.6.2 Transit

The availability of transit in urban areas is another explanation for the preferences of different groups to locate in urban areas. For instance, LeRoy and Sonstelie (1983) and Glaeser et al. (2008) argue that the poor are attracted to central city due to greater transit options.

To control for transit performance in different areas, 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. Using the fitted value from a function of transit time on the distance from each establishment, we define our measure of transit performance as the average time of a 5 mile trip using transit, starting from the tract centroid.[34] This variable unsuprisingly takes much lower values in urban areas, and in many suburbs the variable takes a value equal to walking time.

We computed this transit variable in the summer of 2014, which implies that it captures both the 2000 level and any endogenous changes since 2000, and its coefficients should be interpreted with caution. The regression results (not shown) suggest that proximity to efficient transit has attracted all groups over the last decade, and coefficients are negative and significant for 5 out of 6 groups. The coefficients are more negative for the non-college-educated group, perhaps reflecting their lower income and higher need for transit. For instance, the standardized coefficient estimate is -0.007 for college-educated 25-34 year olds and-0.012 for non-college educated 25-34 year olds, meaning that this highly urbanized variable is unlikely to explain the faster urbanization of the college-educated group.

Finally, this specification delivers coefficients and predictions on the factors driving composition changes that are very similar to those from our base specification in Table 1.

4.6.3 Commuting Analysis

One obvious unobservable in our analysis above is the location in which households in our data work. We can use the LODES commuting data to study whether this unobservable biases our results. To do this, we estimate a discrete-choice logit model that describes how households allocate across both residential and workplace tracts.[35] Preliminary results suggest that controlling for workplace location with workplace fixed effects only has a small effect on our regression coefficients.[36]

movements away from central cities, the decline in crime in the 1990s only had a weak effect on population changes, and in particular did not cause central city population increase, but appears to have reduced urban flight.

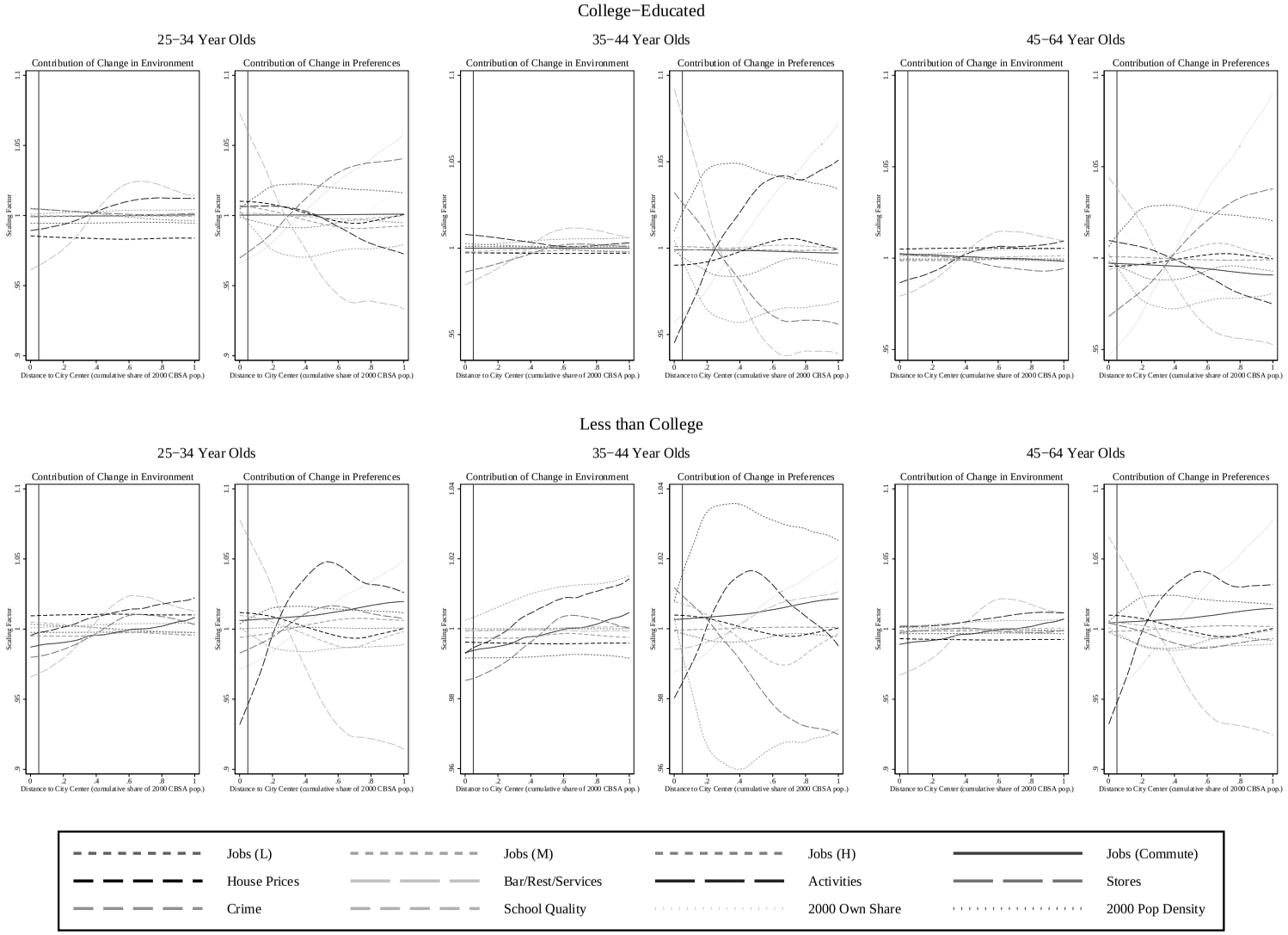
[33] One exception is the coefficient on changes in house prices, which becomes positive and significant for the middle-aged and older non-college-educated group. This result is likely due to the highly selected sample and attributable to a few CBSAs for which we have schools but no crime data.

[34] This specification is consistent with a fixed cost of walking to and waiting for transit, and constant time cost of distance once in a train. While buses may experience congestion, etc., we experimented with other more flexible polynomials and found similar results.

[35] Details on this model are provided in Appendix D.

[36] The LODES data describes the locations of employed people and is disaggregated into three age and three wage groups and not the interaction between these groups. This population breakdown is not sufficiently disaggregated to replicate our main stylized facts presented in Section 3. While we explore obtaining the underlying micro data, we reserve the residential-workplace model for this robustness exercise and

Figure 12: Factors Contributing to vs Actual Urban-Suburban Tract Growth Gradients: College-Educated 25-34 year olds Including School Quality and Crime



Notes:

4.6.4 External Validity: NETS Trip Shares and CEX Expenditure Shares

In this subsection, we compare the regression coefficients in Table 1 with National Household Transportation Survey (NETS) data on trip shares and Consumer Expenditure Survey (CEX) data on average expenditure shares in 2000 and 2010 across members of different age-education group. This exercise is one indirect way to corroborate our structural interpretation of regression coefficients on changes in amenities as parameters governing the 2010 tastes for amenities and regression coefficients on the 2000 levels of amenities as parameters governing the 2000 to 2010 change in tastes for these amenities. The intuition here is that if, for instance, young college-educated individuals have a stronger preference to co-locate with bars, then we expect that they will also take a larger share of their trips and allocate more of their total expenditures to bars than other age-education groups.[37][38] While the results of this imperfect comparison are sometimes inconclusive and our estimates of changes in expenditures often lack precision, overall NETS and CEX data lends some credence to our structural interpretation of model parameters. We focus on four categories that have a reasonable counterpart in both the CEX and the NETS: bars, restaurants, apparel stores, and food stores. Our definition of CEX expenditures for "restaurants" include all food away from home, except beer, wine and other alcohol which we classify as expenditure on "bars". NETS trip categories are sometimes more aggregated and we match the category "go out" (bar, entertainment, theater, sports event) with "bars" and the category "buy goods" (groceries/clothing/hardware store) with both food and apparel stores.[39]

We provide the CEX's UCC codes and NETS trip purpose codes corresponding to these categories, as well as details on CEX and NETS data construction in Appendix A. To maximize sample size, we aggregate quarterly CEX data over 5 years. The 2000 and 2010 data respectively reflect a five-year average of quarterly data from 1998 to 2002 and from 2008 to 2012. NETS surveys are usually about 7 years apart so we use the 2001 and 2009 versions of the NETS.

For each amenity category in each year, we compute the average expenditure share from the CEX and average trip share from the NETS across all individuals within each of our six age-education groups, using the sampling weights provided in each respective dataset. Note that the CEX only reports expenditures at the household ("consumption unit") level, so we attribute the expenditure shares of the household to its individual members. The NETS however separately records all trips on a given survey day by all members of participating households and we compute the share of trips for a given purpose out of all trips undertaken by an individual in our data.

We now compare the regression coefficients on levels and changes from our residential choice model with levels and changes in expenditure and trip shares for different age-education groups. Our hypothesis is that age-education groups who spend more on or travel more to an amenity in 2010 relative to other groups may have more positive regression coefficient in amenity change, which we interpret as a higher taste for co-locating with an amenity in 2010. Similarly, we hypothesize that an age-education group that increases its expenditures on or increases its trip share to an amenity from 2000 to 2010 relative to other groups may have a larger regression coefficient in 2000

do not make it part of our main analysis.

[37] Expenditures are, of course, endogenous and variation across groups and over time may indeed be the result of the location decisions of households. Ideally, one would also like to compute these expenditures and trip shares conditional on access to various amenities. Unfortunately, we currently only have geo-coded information for one year of the NETS and only coarse location information for CEX participants after 2006 so, for this draft, we do not control for household location. We hope to do so in future versions.

[38] A more general hypothesis is simply that preference coefficients should be larger for amenities with higher expenditure shares or higher trip shares. This hypothesis is strongly rejected by the data. For instance, food stores have very high expenditure shares for all groups, while bars have quite low expenditure shares for all groups, yet our regression results indicate that households, if anything, tend to co-locate near bars and not food stores. There are, of course, many reasons why living next to grocery stores could be undesirable and some reasons why locating in an area with many bars has advantages (e.g., ability to walk there instead of driving intoxicated). We, therefore, control for the overall attractiveness of these amenities by studying between-group differences and intertemporal changes in expenditures.

[39] There is a category for personal services in the CEX and NETS that also provides a reasonable match with our amenity categories. However the trip purpose "personal" in the NETS accounts for a very small fraction of trips (around 0.3%, about 4 or 5 times smaller than the "go out" category, the smallest that we present data for) and we are unable to precisely estimate these shares across groups.

amenity level, which we interpret as an increase in taste for co-locating with that amenity between 2000 and 2010. While we do not have enough age-education groups to perform formal hypothesis tests, visual inspection of the results provides some support for our model and uncovers consistent patterns between the CEX and NETS data.

We present the results for bars in Figure 13. The results for other amenities are in Appendix C. Bars are of special interest: Figure 9 suggest that it exerts the strongest urbanization pull on the young and college-educated and, unlike most other amenities, bars feature large variation in both coefficients, expenditure and trip shares, and changes in expenditure and trip shares across age-education groups. The three upper boxes in Figure 13 display the estimated regression coefficients on the change in the bar amenity index (reflecting the level of tastes of each group in 2010) on the left, the mean 2010 expenditure shares on bars in the middle, and the mean 2009 shares of trips to "go out" on the right. Each box features estimates for all 6 age-education groups and displays a 95% confidence intervals around these estimates. The three lower boxes display the estimated regression coefficients on the 2000 bar amenity index (reflecting the change in tastes for each group from 2000 to 2010) on the left, the mean 2000-2010 changes in expenditures on bars in the middle, and the mean 2001-2009 changes in trip shares to go out on the right. A visual comparison shows a good correspondence between regression coefficients, CEX expenditures, and NETS trip shares. In particular, the 25-34 year old college-educated both have the most positive 2010 taste coefficients on bars, the largest 2010 average expenditures shares on bars, and the largest 2009 trip shares to go out. Moreover, the 25-34 college-educated have the largest change in tastes coefficients, and also display the largest increase in expenditures on bars from 2000 to 2010 and the largest increase in the share of trips to "go out" from 2001 to 2009.

The results for the four other amenity categories shown in the appendix are more uneven than for bars, but somewhat supportive of the model. The results for restaurants are representative of the general match between CEX, NETS, and our regression results. As in bars, the CEX and NETS feature very similar patterns across groups. The model and CEX and NETS data consistently capture differences between different college-educated age groups, but sometimes regression coefficients do not agree with the CEX or NETS data on the difference between college and non-college-educated groups (for instance, the non-college-educated groups - for whom regression coefficients are usually less precisely estimated - spend less and travel less to restaurants but they do not have lower taste for co-locating with restaurants according to the model). Changes in NETS trip shares and CEX expenditure shares on categories other than going out and drinks away from home are usually measured with large confidence intervals making precise comparisons with regression coefficients across groups difficult. However, it is worth nothing that, especially in the NETS, groups that have the highest share of trips often also have the largest change in shares of trips. For instance, the 25-34 year old college-educated group has the largest expenditure shares and trip shares on both bars and restaurants (in both 2000 and 2010) and also exhibit the most positive and second most positive respective changes in expenditure on and trip sharesto bars and restaurants from 2000 to 2010.[40] Therefore these datasets support the strong correlation between levels and changes in tastes in regressions, and suggests a divergence in tastes of various age-education groups.[41]

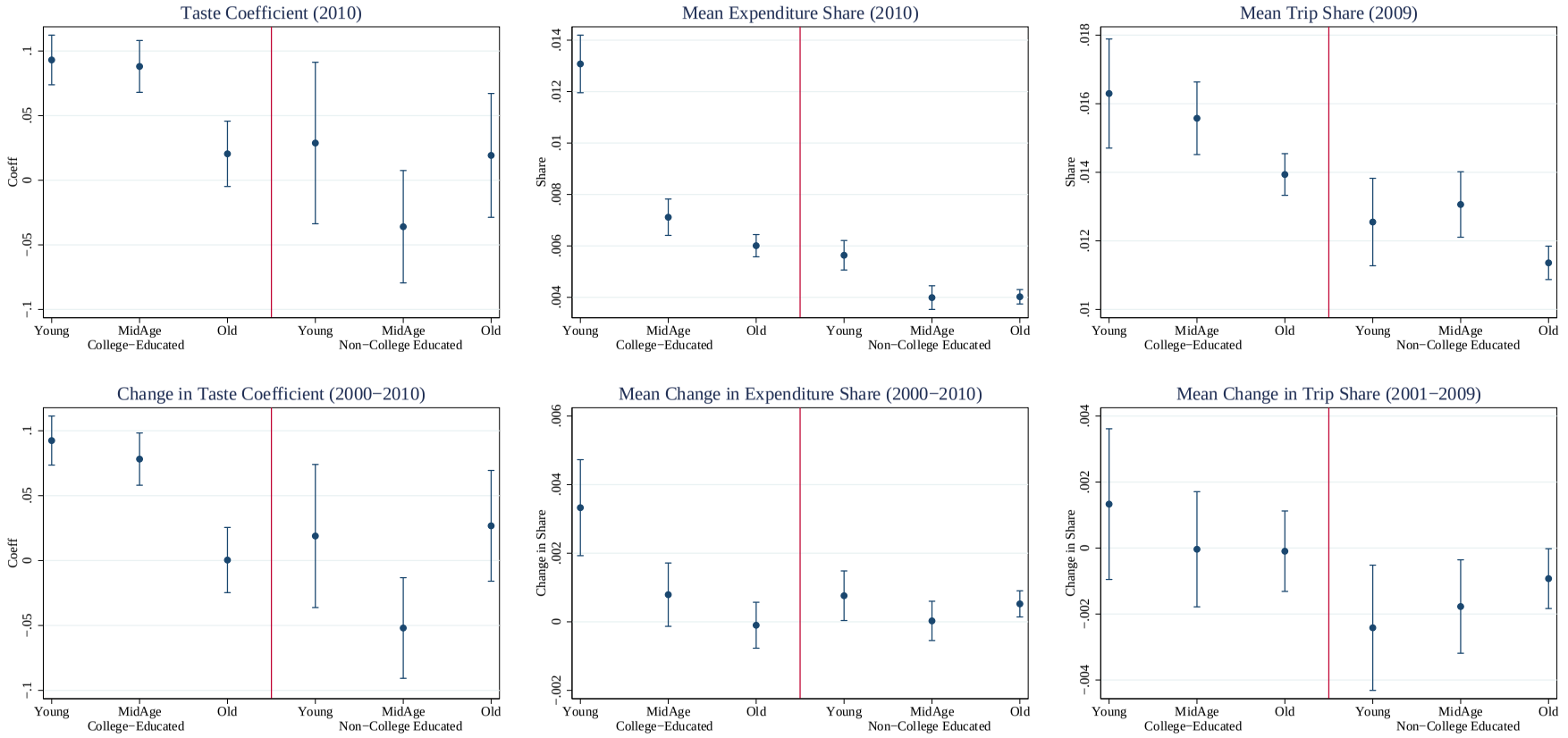
To conclude this analysis, it is important to emphasize that the share of trips whose purpose is work-related is similar than the sum of the trip shares to "buy goods," "go to restaurants," and "go out" that we report in this paper. Given the large share of trips to consumption amenities, one should not be surprised to find that they are important determinants of individual location choices.

This comparision is only meant to be illustrative, and there are many reasons why a given group may have a distaste for co-locating with an amenity despite higher expenditures or higher trip shares to that amenity relative

[40] Changes in expenditure shares on bars and trip shares to go out are positive for 25-34 year old, as are changes in trip shares to restaurants. Expenditure shares on restaurants, however, decrease for all groups, but less so for 25-34 year old and 35-44 year old college-educated.

[41] We do not attempt to back out 2000 preferences from our regression model - which would be necessary to establish a convergence in tastes from the model - by subtracting coefficients in levels from coefficients on changes, because it demands an implausible amount of precision.

Figure 13: Comparison of Regression Coefficient Estimates for the Bar Amenity Index with CEX Average Expenditure Shares on Bars and NETS Average Trips Shares to "go out".



Notes: All confidence intervals are 95%. The first column reports regression coefficients from Table 1 on changes in bars (upper box) and levels of bars (lower box) for each age-education group. The second column reports average expenditure shares on bars (upper box) and changes in average expenditure shares on bars (lower box) from the CEX. The third column reports average trip shares to 'go out' (upper box) and change in average in trip shares to 'go out' (lower box) from the NETS data. All confidence intervals are 95% intervals.

to other groups. While the NETS and CEX results lack control variables, they provide considerably more direct evidence than our regression results. Therefore, we view the finding that young professionals are increasing their expenditures on and trips to highly urbanized amenities like restaurants and bars by more than other groups as separate evidence supporting the importance of amenity levels in explaining key features of America's recent urban revival.

5 Explaining Changing Tastes

A structural interpretation of our estimation results suggests that changing tastes for some urbanized amenities play an important role in explaining the urbanization of the young and college-educated. We identify this change in preferences from a correlation between changes in the location choices and the spatial distribution of consumption amenities in 2000. This correlation may result from omitted variables, but we note that confounding factors must be both unobserved and time-varying, because the first-difference specification controls for any constant unobserved characteristics. Given the large number of (instrumented) controls for changes and levels that we include in our main analysis, as well as the additional controls introduced in the robustness checks above, we find this unlikely.[42]

In this section we investigate three potential drivers of changing preferences for the urbanized amenities driving young professionals downtown. First, what we interpret as a change in preferences for proximity to amenities could be explained by changes in the (unobserved) quality of amenities that may be larger in locations with more initial density. This could happen if, in urban areas, restaurants like Starbucks, which are popular with young professionals, are replacing restaurants like KFCs, which they avoid. Second, the family structure and income levels of the young and college-educated may be tilting towards groups that have always had a preference for urban amenities, with less demand for schooling and/or more leisure time or disposable income to spend on consumption amenities like bars and restaurants. Third, there could be a complementarity between urbanized amenities and mobile technology that benefits digitally savvy young professionals, through, for instance, mapping applications and establishment ratings aggregators.

5.1 Changing Amenity Quality

Unobserved amenity quality could correlate with our amenity indices if amenity quality has improved in urban relative to suburban areas in which case, what we interpret as a change in taste for amenity density could be driven instead by quality improvements. To test this hypothesis empirically, we use data from ESRI business analyst to create a quality control variable for restaurants, and include it into our location choice regression.

ESRI divides each neighborhood in the U.S. into market segments, and assigns a 'Market Potential Index' (MPI) to each chain in each segment based on the prosensity of a segment's inhabitants to visit a given chain relative to the average American. We are able to use the MPI in the segments containing large shares of young professionals to identify the restaurant chains that young professionals are most likely to visit.[43] The 61 chains for which MPIs are available are generally the largest national chains of both family and fast-food restaurants. The

[42] Another possibility for the size of the coefficient in amenity levels is the additional variance in parameter estimates due to collinearity with the corresponding amenity changes. This collinearity is exacerbated in our 2SLS estimation procedure where we find that levels are important amongst the exogenous predictors of changes in the first stage. This concern is mitigated by our OLS results, which are much less affected by this correlation, and also highlight the importance of the initial levels of bar and restaurant amenity indexes in explaining the urbanization of the young and college-educated.

[43] We compute the young professional MPI of each chain by averaging its MPI across all segments with both more than 50% college-educated and more than 50% 18-44 year olds. To measure the propensity of each type of individuals to visit different stores, ESRI uses the Survey of American Consumers, a proprietary dataset from GfK MRI. Of course, young professionals will report shopping near chains closer to where they live, and in this sense a quality control based on such data could spuriously correlate with the location of young professionals, and needs to be interpreted with care. While we compute an instrument for chain entry to alleviate this concern, only view results based on such marketing studies as suggestive.

three largest MPI values are Starbucks (2.17), The Cheesecake Factory (2.11) and Chipotle (1.86) while the three lowest MPI values are Logan's Roadhouse (0.22), Church's Fried Chicken (0.33) and Bob Evans Farms (0.37). Using these MPI as weights, we compute our restaurant quality index as a weighted average of the MPI ratings of all MPI-rated establishments near a tract, using the same CES weights based on transport costs as in the standard amenity indices described in section 4.1.2. This index describes the average quality of restaurants in a given tract based on the preferences of young professionals. To alleviate measurement error and concerns that changes in restaurant quality are driven by arrivals in young professionals, we instrument this change by predicting entry of each chain based on the suitability of the 2000 business environment, using a methodology similar to that for the standard amenity index (the details are in Appendix E).

Table 3: Nested-Logit Residential Location Choice Regression Results
Including Amenity Quality Control

	Including Quality Index		Base Specification - Limited Sample		Base Specification - Full Sample	
	Change	Level	Change	Level	Change	Level
Variable	[1]	[2]	[3]	[4]	[5]	[6]
Job Opportunities – Low Inc.	-0.13***	-0.03***	-0.13***	-0.04***	-0.13***	-0.05***
Job Opportunities – Mid Inc.	0.02	0.05***	0.00	0.06***	-0.03**	0.04***
Job Opportunities – High Inc.	0.13***	0.00	0.14***	0.01	0.17***	0.03***
Avg. Travel Distance	0.03***	0.01***	0.03***	0.02***	0.06***	0.03***
House Price Index	0.05***	-0.02***	0.05***	-0.02***	0.04***	-0.01***
Museums, galleries and libraries	-0.04***	-0.05***	-0.03***	-0.03***	-0.02***	-0.02***
Golf and parks	-0.03***	-0.01***	-0.02***	-0.01*	-0.01**	-0.01**
Gym and sports	-0.02	-0.02	0.01	0.02	0.00	-0.01
Restaurants	**0.12*** **	**0.09*** **	**0.16*** **	**0.13*** **	**0.10*** **	**0.09*** **
Restaurant Quality	**0.06*** **	**0.03*** **	**-**	**-**	**-**	**-**
Bars	0.06***	0.07***	0.05***	0.05***	0.09***	0.09***
Personal Services	0.09***	0.08***	0.06***	0.03	0.08***	0.06***
General Merchadise Stores	-0.02**	-0.02*	-0.04***	-0.05***	0.00	0.00
Food Stores	-0.02	-0.04**	-0.03***	-0.07***	-0.07***	-0.16***
Apparel Stores	0.00	0.00	0.01	0.03**	0.04***	0.05***
Population Density	-	-0.03***	-	-0.02***	-	-0.04***
Share of Same Type	-	-0.07***	-	-0.07***	-	-0.08***
Within-CBSA share	0.77***	-	0.78***	-	0.77***	-
R-squared	0.91		0.91		0.90	
Observations	29,449		29,449		36,103	

Notes: * – 10% significance level; ** – 5% significance level; ***–1% significance level.

An analysis of this index confirms our hypothesis that restaurant quality has grown faster in urban relative to suburban areas over the last decade. Regression results for the 25-34 college-educated group are in Table 3, show that including a quality control resolves an omitted variable bias for the restaurant level variable. The allow this comparison, Table 3 shows, side-by-side, results with the variables for change and level of restaurant quality in Column 1 and 2 and results from the same sample without these quality variables in Column 3 and 4.[44] As expected the coefficients on the change in quality is positive, indicating that young professionals have a preference for quality, and the coefficients on quality level is also postive but smaller, indicating an small increase in taste for quality. Importantly the coefficient on the change in the restaurant amenity index, which we interpret as capturing changes in tastes, drops by 35%. Results for other groups - not shown - do not feature this drop in the coefficient on the standard restaurant amenity index after including restaurant quality, and only one out of five other groups has a preference for this measure of restaurant quality tailored to young professionals. We conclude that group-specific measures of amenity quality matters and that what we interpret as a change in tastes for restaurant density may come in part from improvements in amenity quality that correlate with urban density.

[44] We cannot construct an MPI-weighted restaurant composition indexes for census tracts where we do not observe any of the restaurant chains for which MPIs are available. Columns 5 and 6 show that our coefficients are, for the most part, similar when estimated on the full sample as on the limited sample for which we can calculate the MPI-weighted restaurant composition index.

Finally, we note that restaurant quality is more urbanized in large cities, and has increased faster over the last decade in large relative to small cities. As a result, both the change and level of restaurant quality are variables that can explain the relative urbanization of young professional in large cities. While our measure of amenity quality is imperfect and only for one amenity category, the above exercise provides preliminary but suggestive evidence that, despite faster suburban relative to urban growth in amenity density, changes in urban amenity composition in the non-tradable service sector towards amenities that young professional favor may have contributed to urban revival.[45] The evolution of amenity composition as neighborhoods gentrify is likely a productive avenue for future research.

5.2 Changing family structure and income distribution of the young and college-educated

Changing household formation rates and changing income distribution within the young-college educated groups could explain their changing preferences. This could happen, for instance, if richer or solo households, who presumably spend more on single/luxury amenities like bars and restaurants, are increasing as a share of the young-college educated. To assess the importance of these trends we rely on NETS and CEX data on trip and expenditure shares of individuals grouped by the age-education categories studied above, interacted with household type and income bracket categories. Figure 14 shows IPUMS data from 2000 and 2006-2010 for college-educated 25-34 year olds, with the share in 2000 (in blue) and in 2006-2010 (red) for different household types on the right and different income brackets on the left. The five household types are 1. Solo[46], 2. Married couple with no children, 3. Household with oldest child younger than 5 years old, 4. Households with oldest child older than 5 years old, and 5. Others. The number of solo young professional is increasing, while families with young and old children are decreasing. The four income brackets in Panel B are: \$0-2,000, \$20,000-40,000, \$40,000-60,000, and \$60,000+, where income is adjusted to reflect a "per capita" equivalent using the modified OECD equivalence scale.[47]Contrary to our hypothesis, the number of low-income young professionals appears to be increasing and that of high-income appears to be decreasing. However we note that using the CPI deflator of 0.764 provided by IPUMS data, one obtains that all age-education group experience a decrease in the high-income bracket, and this decrease in the share belonging to high-income bracket is considerably more pronounced for the young and non-college-educated, than for the young and college-educated.

Figure 15 reports 2000 share of expenditures on bars and trips to go out by household type and income bracket. Panel A shows 2000 CEX expenditure shares on bars and Panel B shows 2001 NETS trip shares to "go out." The shares by income bracket are presented on the left of each panel, while those by household type are presented on the right. Appendix C provides a simlar figure for expenditures on and trips to restaurants, apparel and food stores, and other retail. The share of trips and expenditures on highly urbanized service and entertainment amenities like restaurants and bars (or trips to " go out" in the NETS) tends to increase with income, be highest for solos, and decreases as one gets married and has older children.[48] Expenditure shares on stores (apparel and food) as well

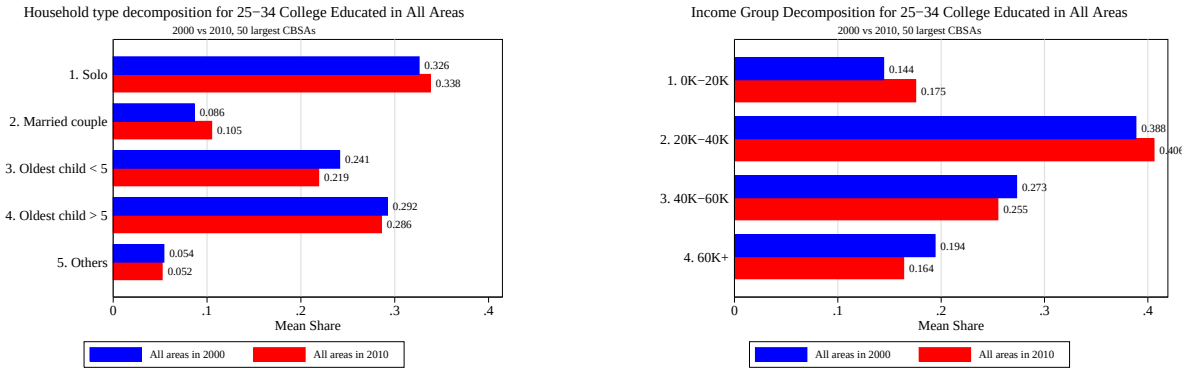
[45]We experimented with using MPI for food and apparel stores We also identified a set of food and apparel stores that had high ratings on Yelp.com in a large sample of cities and computed indices of the share of establishsments with high yelp rating. While the food and apparel MPI are available for a very limited fraction of our sample - because many tracts do not have any MPI rated food or apparel chains within 50 miles - we could produce the Yelp index more broadly. The results suggest that MPI apparel quality and highly-rated Yelp food and apparel stores are suburbanizing over the last decade. Only MPI food quality stores (e.g. Whole Foods, Traders Joes) are very slighly urbanizing and this reduces the increasing aversion of young professionals for these stores. Overall these results support the general narrative that retail stores are not driving urban revivals.

[46]Some solo individuals probably do not live alone, but instead live with unreported roommate(s).

[47]We use the modified OECD equivalence scale to obtain a "per capita" equivalent income for households with more than one member. To account for the fact some key expenditures (e.g. housing) can be shared, this equivalence scale assigns a weight of 1 to the first adult in the household, a weight of 0.5 to each additional adult, and 0.3 to each additional child. So if an observation in our data is for a household with 2 adult and 1 children with income of \$83,000, then the per capita equivalent income is 83,000/(1+0.5+0.3) = \$46,111. We normalized income to 1999 dollars using the CPI adjustement provided by IPUMS data.

[48]We ignore the category household type "others" in our discussion of the results. It includes unmarried couples (inconsistently defined

Figure 14: Share of Individuals in IPUMS data belonging to each Household Types and each Income Bracket in 2000 and 2006-2000.



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 2006 to 2010. 2006-2010 income is discounted to 1999 dollars by a factor 0.764 (recommmended in IPUMS data) and all income is adjusted for household size using the OECD equivalence scale.

as trips to buy goods display the reverse pattern, though expenditures on apparel stores do not tend to vary with income. These differences are generally significant in the CEX expenditure data, but not the NETS data where the trip shares are measured less precisely.

We perform a simple decomposition to evaluate the power of changes in the distribution of the young and college-educated across household types and income groups to explain changes in expenditure shares or trip shares for each amenity category. For simplicity, suppose that all households have equal total expenditures normalized to one unit. The following decomposition is the same for each age-education group and each amenity, so we do not index these in our presentation. Denote each of N types by n (either one of 5 household types or 4 income brackets) and the share of individuals of that type at time t by $s_{n,t}$ (based on either the CEX or NETS sample). Denote the expenditure or trips of type $n$on a given amenity at time t by $x_{n,t}$. We can write the aggregate expenditure or trip share on this amenity in any period t as $x_t = \sum_{n=1}^{N} s_{n,t} x_{n,t}$:

The change in the expenditure or trips from 2000 to 2010 can therefore be expressed as

$$x_{2010} - x_{2000} = \underbrace{\sum_{n=1}^{N} s_{n,2010} \left(x_{n,2010} - x_{n,2000}\right)}_{\text{Component [1] due to change in within-type expenditure/trip patterns}} + \underbrace{\sum_{n=1}^{N} \left(s_{n,2010} - s_{n,2000}\right) x_{s,2000}}_{\text{Component [2] due to change in type composition}} \tag{7}$$

Our objective is to assess the importance of this second component of the decomposition in equation 7. If this component is highly correlated with the actual change in expenditures or trip shares, then changing household types or income distribution within an age-education group have the potential to explain the changes in preferences that we estimate in subsection 4.6.4

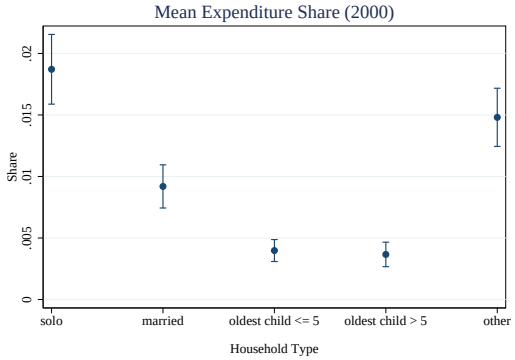
The result of this decomposition, no shown, suggest that changes in the share of individuals in each household types and income brackets generally contribute to determining changes in trip shares and expenditure shares. For instance, when we decompose the change in expenditure shares on bars, which saw a significant increase for 25-34 year old college-educated between 2000 and the 2010, the sign of component 2 in equation 7 capturing the impact of changes in composition is positive for both the income and household type decomposition. The same is true for the decomposition of trip shares to "go out." The main caveat in this exercise is that the changes in shares of each

across years in the CEX for instance) and households with more than 2 adults and no children under 18 (e.g., 25 year olds living with their parents).

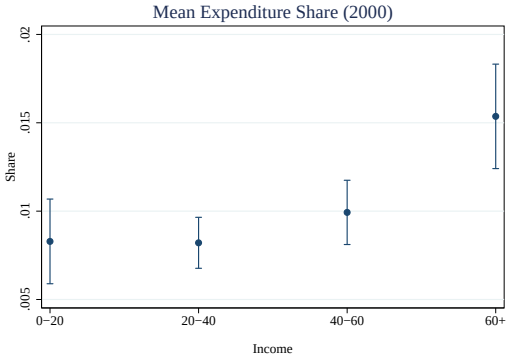
Figure 15: Average Expenditure Shares on Bars and Trip Shares to "Go Out" by Household Types and Income Bracket for 25-34 Year Old College-Educated Individuals.

Panel A: CEX 2000 Expenditure Shares on Bars

By Household Types

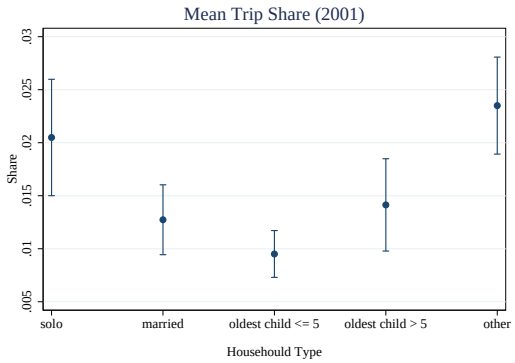


By Income Brackets

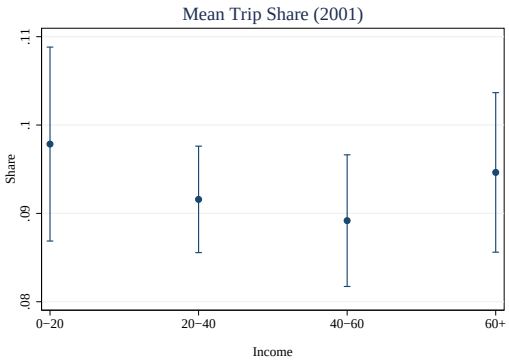


Panel B: NETS 2001 Trip Shares to Go Out

By Household Types



By Income Brackets



Notes: All confidence intervals are 95%.

household type and income groups from 2000 to 2010 in the NETS and the CEX are neither precisely estimated nor always consistent with IPUMS changes in shares. [49] For instance, the share of solo people in the NETS data is much smaller than in IPUMS data (7% vs 33%) and decreasing from 2000 to 2010. The household type shares in CEX data are closer to IPUMS: the CEX share solo is about 1/3 smaller than in IPUMS but also slightly decreasing, the CEX share married without children is increasing, and the CEX share married with children is decreasing. In both CEX and NETS, there is an increase in individuals in both the lowest and highest income bracket, although the NETS data features stronger income growth at the top. Perhaps unsurpringly, the NETS data performs poorly on the household decomposition and better on the income decomposition, and the exact reverse is true for the CEX. The correlation across all 4 amenity categories and all 6 age-education groups for the CEX between the actual change in expenditure and that predicted household type change is 0.73, but only 0.11 for income. In the NETS these correlations are 0.31 for the household type decomposition and 0.62 for the income decomposition.

Overall, these results demonstrate that smaller families and higher income groups spend more on and travel more to urbanized service and entertainment amenities like restaurants and bars. Recent trends towards smaller families for young professionals can potentially explain an increase in preferences for colocating near these amenities from 2000 to 2010. A trend towards higher income of young professionals would have the same impact, but such trends are difficult to assess given the sensitivity of the income levels to the choice of CPI deflator.[50] Finally, we cannot assess the importance of these trends with any degree of precision, because the CEX and NETS data do not precisely measure changes in shares of each household type and income brackets.

5.3 Changing mobile technology and review platforms

One last hypothesis that could explain the young professionals' stronger tastes for urban amenities is their complementarity with recent innovation in mobile technology, providing broad access to mapping applications and establishment ratings aggregators. This hypothesis is speculative and hard to test directly, but one approach is to use NETS data to compute measures of the share of establishments near a tract that are independent (i.e., not part of a chain). Presumably these independent establishments stand to benefit more from easily accessible review portals, and may drive the complementarity between technologies and urbanized amenities.

We identify independent establishments in the NETS data as those that do not have at least 5 other establishments with the same name.[51]We measure a composition index reflecting the local share of establishments that are independent (see appendix section E for calculation details). As expected, the share of independent restaurants is considerably larger in urban areas, especially in large cities. The share of independent restaurants is also increasing faster in urban relative to suburban areas, although this increase is faster in small relative to large cities. However, adding this variable in level and changes to our main regression in Table 1 has little effect on results and for instance across all 6 age-education groups only 4 coefficients out of 12 are significant at the 10% level. In particular, coefficients in both levels and changes are close to zero and insignificant for the 25-34 college-educated. So, while

[49]We use the sampling weights provided by the NETS and CEX data, but we cannot be certain that sampling correctly captures the proportion of such small groups.

[50]It is also possible to more directly test the importance of household types and income groups on the urbanization of young professionals through a Bartik-type shift share analysis. For instance, solo households are considerably more urbanized than families with children and the relative growth of solo households and decline in families with children could push young professionals downtows. A shift share analsysis for 25-34 year old college predicts their overall urbanization, but does not sucessfully predict which CBSAs will experience young professional growth. Moreover, the urbanized households types are not growing faster for the 35-44 year college-educated, which are also urbanizing. While this analsys is quite rough and does not include controls, it shows that changes in family composition are only one possible driver of urban revival, and perhaps unlikely to be the main driving forcel.

[51]In section 5.1 we used regular expressions to try to capture as many restaurant chains as possible. In this case however, there are too many chains to allow manual coding, and we can only require that chain establishments have exaclty the same name. This procedure is less prone to errors in this case, because our methodology does not rely on any chain in particular, but rather on whether each establishment is part of a larger group with the same name. The NPD Group, a marketing agency, report that 53.8% of restaurants in the Spring of 2010 are independent. In the NETS sample we identify 49.6% of restaurants as independent. .

a large share of independent restaurants is a feature of the downtown of large cities, it does not appear to predict the urbanization of young professionals. Of course, the share of independent restaurants only allows for a very coarse test of our hypothesis. A better test of the importance of mobile platforms and review systems in explaining urban revival would exploit spatial variation in the timing of the introduction of various applications or platforms (e.g. yelp), but such variation is hard to isolate.

6 Additional hypotheses

One prominent hypothesis behind the re-sorting of households is reduced access to homeownership following the housing crisis and recession of 2007-2009. Given that rental units (almost always multifamily) are more urbanized that owner-occupied units (generally single-family homes), a decline in accessibility to home ownership that dispropotionately affected young professionals could pull them out of the suburbs and push them into urban areas.

There is much evidence that in the aftermath of the housing crisis, credit score requirements for access to mortgage credit became more stringent. For instance, the average FICO credit score of mortgages acquired by the Fannie Mae and Freddie Mac rose from 725 in 2007 to more than 760 by 2010 (Parrott and Zandi 2013).[52] Presumably, this reduction in credit availability has been disproportionately harmful to younger individuals about to enter the housing market, and who may have been driven away from home ownership towards rental options. Consistent with this story, Rappaport (2015) documents the rapid increase in multifamily construction starting in 2010, and the increased propensity of young adults to live in multifamily units as opposed to single-family homes following the housing crisis.

The main flaw in this hypothesis is the timing of the housing crisis: the 2000s includes more years of historically easy mortgage credit than of restricted credit. Using IPUMS data and a methodology similar to that in the household type decomposition of Subsection 5.2, we decompose the growth of 25-34 year old and 35-44 year old college-educated individuals by tenure type (owners and renters) from 2000 to 2010 (results not shown). We confirm that renters are more prevalent in urban areas, and that the younger group is more likely to rent. However, we find that homeowners have grown *faster* nationally than renters in both age groups.[53] Therefore, the premise of the housing market hypothesis that young professionals have been forced into renting from 2000 to 2010 is not supported by the data. In fact, further analysis reveals that ownership rates among young professionals have increased from 2000 to 2010, in both urban and suburban areas.

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

7 Discussion

Urban revival currently gathers considerable media attention and interest from the general public. We have shown 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 move disproportionately to suburban areas, college-educated 25-44 year olds have flocked to downtown areas.

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

[53] The number of 24-35 year old college-educated owners has grown by 19%, versus 8% for renters. For 35-44 year old college-educated, the number of owners has grown by 11% over the last decade, versus 6% for renters.

In this paper, we evaluate the importance of various explanations for this trend. In our main analysis, we find that diverging preferences for consumption amenities - such as retail, entertainment, and service establishments - explain the diverging location decisions of the young and college-educated relative to their non-college-educated peers and their older college-educated counterparts. In the same model, we find limited evidence that factors like changes in urban relative to suburban neighborhood characteristics, tastes for living in close proximity to job locations, or willingness-to-pay for housing help to explain why the young and college-educated are moving downtown in big cities, while the rest of the country is moving to the suburbs. In complementary analyses, our data provides corroborating evidence for these changing tastes for amenities and rejects other hypotheses, such as changes in mortgage lending practices during the housing crisis

The diverging preferences for consumption amenities to which we attribute urban revival are identified from a correlation between changes in the location choices of individuals in different age-education groups and the spatial distribution of consumption amenities in 2000. It is possible that this correlation is the result of some unobserved factor that we do not control for in our model. We note, however, that confounding factors must be both unobserved and time-varying, because the first-difference specification controls for any constant unobserved characteristics. Given the large number of (instrumented) controls for changes and levels that we include in our main analysis, as well as corroborating evidence of diverging preferences in expenditure and trips data, we find this unlikely.

It is, of course, important to identify the source of such changing preferences. We have explored a few avenues and plan to further this work. We see that part of the change in what we interpret as a change in preferences for proximity to amenities is in fact a change in amenity quality that is correlated with amenity density. 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. Preliminary evidence based on the expenditure and trip choices of different sub-categories of the young and college-educated suggests that the changing composition of this demographic group helps to explain their shift downtown. The idea here is that young professionals now have higher disposable income than in 2000 (due to either changing household formation rates and increasing returns to college), and that downtown amenities are luxury goods. We continue to assess this hypothesis by estimating the differential location tastes of different sub-categories of the young and college-educated in the PUMA-level data (estimating our tract-level choice model for these sub-categories will require restricted-access micro data).

That being said it is striking that the classic factors used to explain household residential location decisions (jobs, housing, crime, and schooling) struggle to explain urban revival. If the key factor at play is indeed a changing preference for urban consumption amenities, then there are important consequences for the sustainability of the urban revival trend and its welfare implications. Since these amenities are endogenous, their concentration will grow with local demand and may act as an anchor for the new generation of college-educated households, keeping them downtown even as they form families and as their demand for space and schooling rises. If we believe that tastes are diverging between the college-educated and their non-college-educated peers, then these consumption amenities will compensate the young and college-educated for the high housing prices that they will increasingly face in gentrifying downtown neighborhoods, but will offer little compensation for the non-college-educated households already living in downtown neighborhoods. These poorer households will either be displaced or have to pay the high housing costs to continue to live in downtown locations where the businesses offer fewer of the consumption 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. University of Toronto and Federal Reserve Bank of Chicago.

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 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.

Brueckner, Jan K., Jacques-Francois Thisse, and Yves Zenou, "Why is central Paris rich and downtown Detroit poor?: An amenity-based theory," *European Economic Review*, 1999, *43* (1), 91–107.

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 restless in a knowledge economy*, CEOs for Cities, 2005.

_ , "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.

_ **, Gilles Duranton, and Matthew A. Turner**, "Speed," Technical Report, Sciences Po 2016.

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

Diamond, Rebecca, "The determinants and welfare implications of U.S. workers' diverging location choices by skill: 1980-2000," *Job Market Paper, Harvard University December*, 2012.

Duranton, Gilles and Diego Puga, "Chapter 8 - Urban Land Use," in J. Vernon Henderson and William C. Strange Gilles Duranton, ed., *Handbook of Regional and Urban Economics*, Vol. 5 of *Handbook of Regional and Urban Economics*, Elsevier, 2015, pp. 467–560.

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 and Katherine O'Regan, "Crime and urban flight revisited: The effect of the 1990s drop in crime on cities," *Journal of Urban Economics*, 2010, *68* (3), 247–259.

Fee, Kyle and Daniel Hartley, "The relationship between city center density and urban growth or decline," Technical Report 2012.

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.

_ and Matthew E. Kahn, "Sprawl and urban growth," in J. Vernon Henderson and Jacques-François Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4 of *Cities and Geography*, Elsevier, 2004, chapter 56, pp. 2481–2527.

_ , 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.

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.

Ihrke, David K., Carol S. Faber, and William K. Koerber, *Geographical mobility: 2008 to 2009*, US Department of Commerce, Economics and Statistics Administration, US Census Bureau, 2011.

Jaffee, Dwight and John M. Quigley, "The Future of the Government Sponsored Enterprises: The Role for Government in the U.S. Mortgage Market," NBER Working Paper 17685, National Bureau of Economic Research, Inc 2011.

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 analysis of qualitative choice behavior," Technical Report 1972.

_ and others, *Modelling the choice of residential location*, Institute of Transportation Studies, University of California, 1978.

Miller, Claire Cain, "Where Young College Graduates Are Choosing to Live," *The New York Times*, October 2014.

Monte, Fernandino, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration and Local Employment Elasticities," November 2015. Georgetown University and Princeton University.

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.

Parrott, Jim and Mark Zandi, "Opening the Credit Box," Technical Report, Moodys Analytics and the Urban Institute, New York and Washington, DC 2013.

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.

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

Stock, James H. and Motohiro Yogo, "Testing for weak instruments in linear IV regression," *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, 2005.

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

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, P., "The Interdependence of Residential and Workplace Choice: Implications of Multiple Worker Households for Residential Mobility and Location Choice," in "42nd Annual Meetings of the North American Regional Science Association International, Cincinnati, Ohio" 1995.

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, many tracts are split or consolidated and their boundaries are changed to reflect population growth or decline. The Longitudinal Tract Data Base (LTDB) provides census tables for constant 2010 tract boundaries from 1970 through 2000. The combination of population and area weighting between 2000 and 2010 in the LTDB has a high degree of accuracy, but the possible source of error is that blocks in 2000 are sometimes split into different portions that are assigned to different 2010 tracts. In the LTDB, the block population to 2010 tracts is allocated in proportion to the area of these block fragments. Since blocks are usually small and have few residents, this aspect of the estimation is unlikely to cause much error.

For the stylized facts on recent urban growth, we assemble a database of constant geography census tracts using this LTDB and data from the NHGIS to estimate decennial censuses of 1970 to 2000 and ACS 2008-2010 within 2010 boundaries. The ACS usually uses legal boundaries as of January 1 of the last year of the estimate period, so ACS 2008-2012 aggregates, which was estimated in 2012, is already based on 2010 boundaries. However, whenever we use the ACS data that was estimated before 2010, such as ACS 2005-2009 aggregates, we interpolate the data to the 2010 boundaries using the LTDB.

CBSA Definitions Core Based Statistical Areas (CBSAs) refer collectively to metropolitan and micropolitan statistical areas. The general concept of a CBSA is that of a core area containing a substantial population nucleus, 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.

A.2 LODES data

The LODES data derives from the Longitudinal Employer-Household Dynamics data infrastructure, and is available for each year from 2002 to 2011. The data consists of three parts: origin-destination data (OD), workplace area characteristic data (WAC), and residence area characteristic data (RAC). OD data characterize residence-workplace flow of workers at the census block level while WAC and RAC data capture the characteristics of workers working or living in a census block. The LODES data available for general public use is processed using noise infusion, small cell imputation, and synthetic methods in order to protect the confidentiality of workers.[54]In this paper, we use the 2002 and 2011 OD and WAC data.

We use the OD data for the commute pattern analysis of subsection 3.3 and the residence-workplace model of subsection 4.6.3. The OD data records the count of persons working and living in a given pair of census blocks

[54]Specifically, adjusted establishment-level employment value is obtained by multiplying true employment value by a random noise factor. The random noise factor is specific to employer and establishment and distorts true employment value within a pre-determined range. The adjusted establishment-level employment is then aggregated to the census block level. For each job in a census block, a residential location is drawn from a set of coarsened residences covering the entire United States, based on a Dirichlet multinominal posterior distribution. The posterior distribution depends on a likelihood distribution and a prior distribution. In particular, the likelihood distribution is derived from the observed commuting of workers with similar characteristics (such as age, sex, earnings, and industry) between coarsened residences and the census tract where the job is located. Once a coarsened residential location is selected, the job is mapped to a census block within the selected coarsened residential location, based on the population distribution for the previous decennial census. Detailed description of procedures used in LODES data for disclosure control and confidentiality protection is available at http://lehd.ces.census.gov/doc/technical_paper/tp-2006-01.pdf and ftp://ftp2.census.gov/ces/wp/2014/CES-WP-14-38.pdf

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 $3,333/month, and greater than $3,333/month). For our analysis we aggregate the OD data at the tract level, and in fact the LODES documentation discourages the use of block-level data, because of the noise infusion procedure.

We use the WAC data to compute the job variables described in section 4.1.3. The data provides counts of jobs in each census block not only by age and income groups, but also by industry sectors. Specifically, the WAC data reports job counts by demographic (age or income) group separately for each of the 20 NAICS sectors listed in Table A.1. This breakdown by industry allows us, after aggregating the data to tracts, to compute the income group-specific Bartik instruments that we use in the paper. Note that we exclude federal workers from both our OD and WAC data.

Table A.1: List of NAICS sectors in the LODES data

NAICS sector	Description
11	Agriculture, Forestry, Fishing and Hunting
21	Mining, Quarrying, and Oil and Gas Extraction
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale Trade
44-45	Retail Trade
48-49	Transportation and Warehousing
51	Information
52	Finance and Insurance
53	Real Estate and Rental and Leasing
54	Professional, Scientific, and Technical Services
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management and Remediation Services
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment, and Recreation
72	Accommodation and Food Services
81	Other Services [except Public Administration]
92	Public Administration

A.3 NETS data

The 2012 National Establishment Time-Series (NETS) Database includes over 52.4 million establishments with time-series information about their location, industries, performance and headquarters over the period 1990-2012. The NETS dataset is based on annual snapshots of U.S. establishments constructed by Duns and Bradstreet. Duns and Bradstreet collects information on each establishment through multiple sources, including: phone surveys, Yellow Pages, credit inquiries, business registrations, payment experiences, public records, court and legal filings, government registries (e.g., with county clerks and Secretaries of State), news and media, and the internet, to name a few. Walls & Associates converts Duns and Bradstreet's archival establishment data into a time-series database of establishment information by creating crosswalks between establishments categorizations (SIC/NAICS), undergoing quality control, and estimating fields in the NETS database. Fields are estimated when data is unavailable in the Duns and Bradstreet archival data, such as in cases where an otherwise complete historical record of a

business is lacking collected information for a given year. Our analysis uses the 2012 release of the NETS data, including the file NETS2012_MISC, which contains the latitude and longitude of each establishment, as well as the NETS2012_MOVE file, which contains information on the location of establishment moves. The NETS data records the exact address for 74.5% of establishments (2009), while only the zip code is known for 24.1% of establishments. When establishments have a location at the zip code level, we assign them to the centroid of the respective zipcode.

Neumark et al. (2007)assess the reliability of NETS data by comparing it to other more aggregated establishment datasets i.e., the Quarterly Census of Employment and Wages (QCEW), the Current Employment Statistics (payroll) survey (CES), the Size of Business data (SOB) and the Business Employment Dynamic data (BED). Their conclusion support our use of the NETS data to compute long differences in establishment changes. Moreover, they report that NETS has better coverage of very small establishments (1-4 persons, which is often the case of urban service amenities) and has consistent coverage with other datasets when comparing all other establishment sizes.

In this paper, we only use NETS establishment data, which sets the bar lower than Neumark et. el. (2005) employment comparisons. Due to rounding issues and imputed data NETS one would expect the NETS data to perform poorly in capturing year-on-year change, but possibly better for the long differences (2000 to 2010) that we are using. Indeed, Neumark et al. (2005) report that: "The correspondence bewteen NETS and QCEW yearly first-difference employment changes by industry and county is not very strong, with a correlation of only 0.528. However if we look at employment changes over periods of at least a few years, this problem is substantially mitigated, as the correlation rises to 0.864" [55] It is also instructrive to look at the aggregate growth of chain establishments. For instance, according to Stock and Wong (2015), Chipotle, one of the most popular restaurant chains with young professionals, had nearly 100 stores in 2000 and grew to about 1000 stores in 2010. This pattern of growth is generally reflected in the NETS data, which reports Chipotle having 21 locations in 2000 and around 800 in 2010. These numbers highlight both the difficulry in capturing all chains in the NETS data (one has to use establishment names), the possible lag in recording new locations, and the ability of the NETS data to capture the general pattern of growth.

Table 5 reports the number of establishments nationally in 2000 and 2010 in each of our 9 amenity category, as well as the SIC codes used to define these categories.[56]

A.4 Zillow house price indexes

We construct the house price index is constructed using publically available data from Zillow.com. The Zillow House Value Index (ZHVI) for all homes (i.e. single family, condominium, and cooperative) is computed monthly at the zip code-level. The index is available for 10,452 zip codes in 2000 and 11,118 zip codes in 2010. For each zip code, we compute a yearly value of the index by average over all months. To map zip codes into census tracts, we use a census tract to zip code crosswalk file provided by U.S. Department of Housing and Urban Development. The crosswalk also contains the share of tract residential addresses that are located in each component zip code that overlays the tract; we use this residential address share as weights when construction our tract-level house

[55] Business relocations are intuitively difficult to record, but Neumark et al. 2005 find that the NETS data accurately tract "significant moves", or cross-county and cross-zip code moves, but is less accurate when reporting within-city moves. For instance they find that "58.5% (237/(452-47)) of the valid business relocations that we identified from the Los Angeles Times could be found in our NETS dataset [...] We are able to confirm only 27% (21/77) of within-city moves, whereas we are able to confirm 70% (177/252) of between-city, within-state moves, and 74% (37/50) of cross-state moves."

[56] The NPD Group, a marketing agency, reports 579,416 restaurants in the Spring of 2010 . Couture (2013) report 273,000 restaurants on Google Local in States accounting for 50% of the US population and report a similar number (268,000) from the National Restaurant Association (suggesting close to 550,000 restaurants nationally). The number in the NETS 416 807 therefore appears to be missing some establishments captured in other datasets.

Table A.2: NETS Establishment Counts and SIC codes

Category	2000 Establishment Count	2010 Establishment Count	SIC Codes
Museums	35972	52961	84120000, 84120100 - 84120102, 84129901 - 84129903, 842200 - 842202, 823100 - 823104
Golf and parks	10438	9727	7992, 7996
Gym and Sports	134613	193238	7991, 7997, 7999
Restaurants	437570	416807	581200 - 581209
Bars	64948	75261	581300 - 581302
Personal Services	385745	544486	723, 724, 729901, 729902
Food Stores	281269	335802	54
Apparel Stores	197909	239863	56
General Merchandise	43468	54797	53

price index. Specifically, for tracts that fall entirely in to component zip codes for which an index is available, we compute the tract-level index as the weighted average of the home value index across all component zip codes with the residential address shares as weights. For tracts that do not fall entirely into zipcodes for which an index is available, we drop the zipcodes with missing data and normalize the residential share in zipcodes with available data to 1. If a tract does not fall into a zipcode for which data is available, 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), covering 498 CBSAs.

A.5 UCR Crime data

The crime data comes from the Uniform Crime Reporting (UCR) city-level crime data from 1990, 2000 and 2010. The crimes reported are 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 the FBI. Thus, we lack coverage if any city did not report. All CBSAs have many cities but most have only one data point for the principal city. In 1990, there were 9222 cities reporting, which increased to 11,044 in 2010. This increase is partially attributable to new cities being incorporated between 1990 and 2010.

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. So for each year, we assign a FIPS code - either a Place or a County Subdivision - to each city, and overlay each census tract on the city map.Census tracts that do not overlap with any cities reporting crime data are discarded. We then assign the crime total for each city to the tracts that overlap with it assuming that population and crime are uniformely distributed within tracts and within cities. Finally, we obtain total violent crime in each census tract by summing over its intersections with different cities, and divide by tract population to obtain per capital crime. The final data set contains crime data for 54,745 tracts in 1990 and 57,095 tracts in 2010.

A.6 Consumption Expenditure Survey (CEX) data

The Consumer Expenditure Survey (CEX) program consists of two surveys, the Quarterly Interview Survey and the Diary Survey, that provide information on the buying habits of American consumers, including data on their expenditures, income, and consumer unit (families and single consumers) characteristics. The survey data are collected for the Bureau of Labor Statistics by the U.S. Census Bureau. In this paper, we use the public-use micro-data from the Diary Survey for years 1998 – 2002 and 2008-2012, which records, for each respondent, all expenditures on small, frequently purchased items over two consecutive one-week periods, as well as characteristics, income and weights for the consumer unit. In particular, each expenditure is classified by a Universal Classification Code (UCC), which we use to identify consumption categories. The CEX expenditure categories that we use to match our amenity categories are as follow (with name of category and UCC codes in parenthesis):

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

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 in 2008-2012.

A.7 National Household Transportation Survey (NETS) data

The National Household Travel Survey (NETS) provides comprehensive data on travel and transportation patterns in the United States, in which detailed data is collected on daily trips taken in a 24-hour period for each respondent via a travel day diary. We use the 2001 and 2009 NETS surveys. Each trip is categorized by a WHYTO code, which we use to identity the trip purpose. The trip purposes that we use to match our amenity categories are as follow (with WHYTO code and description in parenthesis):

1. Restaurants (Meals, get/eat meal, coffee/ice cream/snacks, WHYTO 80, 82, 83)
2. Bars (Go out/hang out: entertainment/theater/sports event/go to bar , WHYTO 54)
3. Food Stores (Buy goods: groceries/clothing/hardware store, WHYTO 41)
4. Apparel stores (Buy goods: groceries/clothing/hardware store, WHYTO 41)

We compute the average share of trips for each trip purpose as the number of trips for that purpose divided by all trips made by the respondent on the travel day. We use weights at the person level (final person weight) when calculating population estimates of mean trip shares. In the NETS, household income is reported in brackets. We use the midpoint of each bracket, and 167, 000 for the top bracket "100,000+", as an estimate for household income. [57]

Our sample size for the 24-35 year old college-educated (smallest group) is 6228 individuals in 2001 and 7309 in 2009.

B Is urban revival a result of population growth or of changing composition?

Our finding that the general population is suburbanizing while the college-educated population is urbanizing suggests that changes in socio-economic composition within CBSAs are an important feature of urban revival. To assess the relative importance of changing population density versus changing composition as drivers of urban revival, we decompose the differences between urban and suburban growth into four components: the change in urban composition, the change in suburban composition, the change in urban population and the change in suburban population. To perform this decomposition, we denote by $s^d_{urb,00}$ the share of the total population of a given CBSA in 2000 that belongs to group d and lives in the urban area of that CBSA. $s^d_{sub,00}$ is similarly defined for suburban areas. Denote the number of individuals in group d living in the urban area of that CBSA in 2000 as $pop^d_{urb,00}$, and the number of individuals in group d in the suburban area as $pop^d_{urb,00}$. To refer to the general population in a CBSA - i.e., to the sum of all group d - we use the superscript $d = all$. The notation is the same for 2010 variables. Using this notation, $\frac{pop^d_{urb,10}/pop^d_{urb,00}}{pop^d_{sub,10}/pop^d_{sub,00}}$ measures the ratio of urban to suburban growth for group d,

[57] In the 2009 NETS, all children under the age of 5 are excluded from the survey. While the NETS provides a variable that indicates the age range of the youngest child in the household, we sometimes cannot infer the age of the oldest child if there are household members who did not complete the person interview. We assume that if a household has children who did not complete the interview and we know that their youngest child between 0-5, then their oldest child is also less than 5.

and it takes a value larger than 1 in CBSAs experiencing urban revival.[58] So for each group and each CBSA, our decomposition is:

$$\frac{pop^d_{urb,10}/pop^d_{urb,00}}{pop^d_{sub,10}/pop^d_{sub,00}} = \left(\frac{s^d_{urb,10}}{s^d_{urb,00}}\right)\left(\frac{s^d_{sub,00}}{s^d_{sub,10}}\right)\left(\frac{pop^{all}_{urb,10}}{pop^{all}_{urb,00}}\right)\left(\frac{pop^{all}_{sub,00}}{pop^{all}_{sub,10}}\right). \tag{A.1}$$

It is instructive to first consider at the correlation between urban revival (the ratio of urban to suburban growth for group d, $\left(\frac{pop^d_{urb,10}/pop^d_{urb,00}}{pop^d_{sub,10}/pop^d_{sub,00}}\right)$) and the four explanatory terms in the decomposition in equation A.1. For the 25-34 college-educated group, which has experienced the fastest urbanization over the last decade, urban composition $\left(\frac{s^d_{urb,10}}{s^d_{urb,00}}\right)$ has a correlation of 0.94 with urban revival. All other correlations are relatively small. This suggests that urban revival happens mostly through changing demographic and socio-economic composition within urban areas, rather than through urban population growth or changes in suburban composition. For instance, CBSAs experiencing urban revival do not display faster urban population growth $\left(\frac{pop^{all}_{urb,10}}{pop^{all}_{sub,00}}\right)$).

We also compute the mean value of each element in equationA.1, across the 50 largest CBSAs. The mean ratio of urban population in 2010 to urban population in 2000 $\left(\frac{pop^{all}_{urb,10}}{pop^{all}_{urb,00}}\right)$ is equal to 0.99. In other words, the downtown population of large cities barely changed, *on average,* from 2000 to 2010. For the same set of CBSAs, the mean ratio of suburban population in 2000 to suburban population in 2010 $\left(\frac{pop^{all}_{sub,00}}{pop^{all}_{sub,10}}\right)$ is 0.89, capturing a significant increase in the suburban population of large cities. Such an increase provides a strong force *against* urban revival as we define it. The average change from 2000 to 2010 in the share of a CBSA's population that is, 25-34 year old college-educated and lives in urban areas $\left(\frac{s^d_{urb,10}}{s^d_{urb,00}}\right)$ is 1.43 confirming a strong shift in urban composition towards young professionals. There is much less change in suburban composition over the same period, and the average value of $\left(\frac{s^d_{sub,00}}{s^d_{sub,10}}\right)$ is 0.97.[59] While these results highlight clear patterns, they also hide interesting underlying variation. For instance, a rust-belt city like Cleveland has experienced urban revival in the face of a rapidly declining urban population (2010 to 2000 ratio of 0.88), thanks to huge improvements in urban composition (2010 to 2000 ratio of 1.78 for the 25-34 year old college-educated group).[60] To summarize, urban composition is changing fast enough to generate a strong trend towards the urbanization of young professional in large cities, despite stagnant urban and rising suburban populations.

[58] Note that we express growth as a ratio x_{10}/x_{00} instead of $(x_{10} - x_{00})/x_{00} = x_{10}/x_{00} - 1$ as elsewhere in the paper.

[59] For the 35-44 year-old group these numbers are 1.21 and 0.99 and for the 18-24 year-old group we find 1.59 and 0.84. It particularly interesting to note that for the 65+ group, these numbers are 1.4 and 0.71. Clearly, then, urban areas have experienced population shifts towards 65+ (or 45-65) college-educated that are as fast as those for young professionals, which may explain the conventional widsom that baby-boomers are returning to urban areas. However, the population of older college-educated Americans has grown even faster in the suburbs, and therefore does not appear to display a strong new preference for downtown living. This large growth in educated baby-boomers everywhere is of course of function of the large relative size of this generation.

[60] Detroit is the only city in which young professionals are not growing faster downtown relative the suburbs so it is interesting to consider its recent growth dynamics. In fact Detroits young professional composition is favoring urban areas. For the 25-34 year old college-educated group, Detroit has a 2010 to 2000 urban composition ratio of 1.02, which is the worse performance among the 50 largest CBSAs, but still an improvement. The suburbs are doing even worse, however, with a 2000 to 2010 suburban composition ratio of 1.10, which is the 3rd worse among the 50 largest CBSAs. So in Detroit composition change favor urban areas. Unfortunately, Detroit has experienced the largest urban population drop among the 50 largest CBSAs, with a 2010 to 2000 ratio of 0.77. Even Detroit's stable suburban population ranks it 47th out of 50 in terms of suburban growth. Clearly Detroit is stuggling. One bright spot is that younger 18-24 college-educated group, a very small cohort, is actually growing faster in urban areas relative to the suburbs, thanks to huge composition changes. This may annouce future trends.

C Additional Results

C.1 OLS Regression Specification

In this subsection we present regression results for the OLS nested-logit specification (although we instrument the within-CBSA share, which otherwise captures almost all the variation in the data. The dependent variable and all regressors are as in Table 1 in the main text. The OLS results are in Table A.3. We first note that almost all coefficients are closer to 0 in the OLS specification, with the key exception being the coefficient on changes and levels of middle-income job, which become larger in magnitude. It is interesting to note that while the IV coefficient on changes in house prices was strongly negative for 4 out of 6 groups, these groups are now almost indifferent in the OLS. Coefficients in the OLS specification are often of the same sign as in the IV, with middle-income jobs again being a notable exception, with highly positive coefficients in OLS and sometimes negative coefficients in IV. High-income jobs in level and changes also sees some sign switches and appears to be a much less attractive feature in the OLS specification. Young college-educated have notably more positive coefficients in levels of amenities than other groups and the old and non-college educated feature negative coefficients in levels of amenities in the OLS that does not show in the IV regressions which wrongly predicts the urbanization of that group. The factors pushing the young and college-educated into urban areas are quite similar in OLS and IV, and for instance replicating Figure 9 in OLS shows that three of the five factors providing the strongest urbanization push are the same as in IV (in OLS the level of middle-income jobs is first, levels of bars and restaurants are fourth and fifth). New addition to the top five are levels of gyms (second) and changes in house prices (third).

C.2 Additional comparison of NETS, CEX and regression coefficients.

Here we produce figures similar to Figure 13 in the paper which compares, for the amenity category "bars", the regression coefficients from Table 1 in changes and levels with average expenditure shares from the CEX and average trip shares from the NETS in changes and levels. We have results for restaurants, food stores, apparel stores. Note that in Figure A.2 on food stores and figure A.3 on apparel stores, the NETS trip share data is the same, and embraces a larger category "buy goods" that includes food, clothes and hardware, and is therefore not directly comparable.

C.3 Additional results on average expenditure shares and trip shares by household types and income brackets.

In this section we report additional results on average expenditure shares from the CEX and average trip shares from the NETS by household types and income brackets. Figure A.4 shows results for restaurants, similar to Figure 15 for bars in the main text. Figure A.5 reports CEX expenditures on apparel stores in Panel A, CEX expenditures on food stores in Panel B, and NETS trip shares to "buy goods" (groceries, clothes, hardware) in Panel C.

D Commuting Analysis

In section 3 we used the LODES commute data to show that job location cannot explain all of the high-income residential shift towards downtown. The reason is that even when holding distance to workplace fixed, high-income workers in large cities still live closer to the CBD in 2011 than in 2002. We now formalize this argument by specifying and estimating a residential-workplace choice model. This model permits the addition of a workplace

Table A.3: OLS Nested-Logit Residential Location Choice Regression Results

Panel A: College Educated

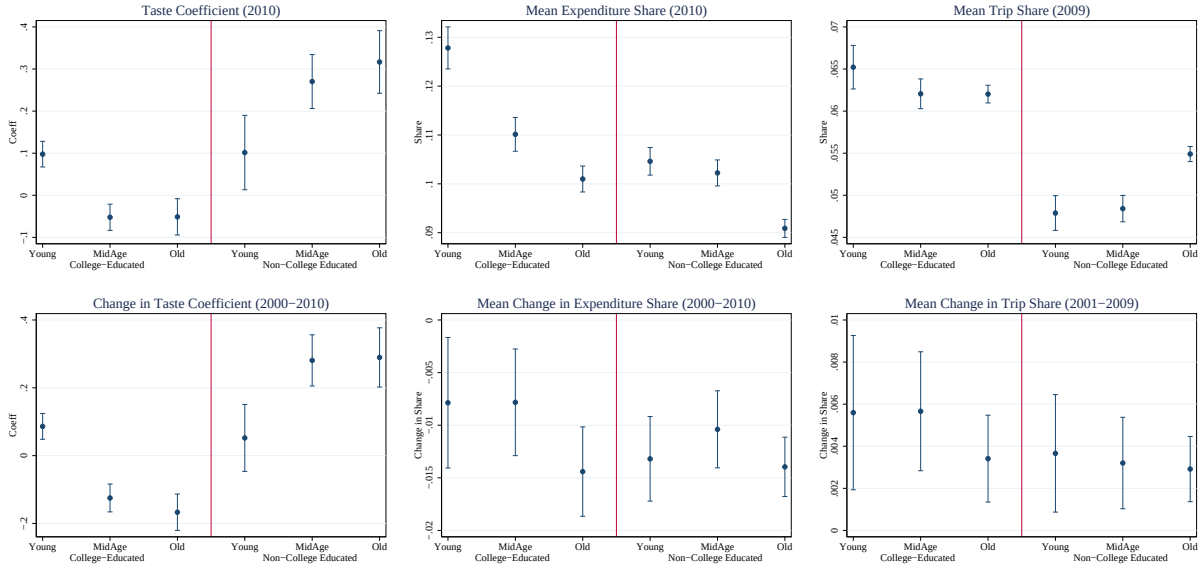
	25-34 Year Olds		35-44 Year Olds		45-65 Year Olds	
	Change	Level	Change	Level	Change	Level
Variable	[1]	[2]	[3]	[4]	[5]	[6]
House Price Index	0.061***	-0.004***	-0.004***	0.022***	0.000	0.013***
Job Opportunities – Low Inc.	-0.069***	-0.041***	-0.067***	-0.173***	-0.007**	-0.140***
Job Opportunities – Mid Inc.	0.054***	0.109***	0.102***	0.236***	0.048***	0.233***
Job Opportunities – High Inc.	0.029***	-0.053***	-0.009***	-0.032***	-0.006**	-0.056***
Avg. Travel Distance	-0.002***	0.000	-0.004***	-0.004***	-0.008***	-0.012***
Museums, galleries and libraries	-0.003***	-0.009***	0.003***	-0.003**	0.000	-0.007***
Golf and parks	-0.002***	0.002***	-0.003***	0.002**	-0.003***	0.000
Gym and sports	0.010***	0.015***	0.014***	0.018***	0.026***	0.040***
Restaurants	0.006***	0.004**	0.002**	-0.014***	0.002**	-0.017***
Bars	0.001**	0.004***	0.006***	0.003**	0.002*	-0.012***
Personal Services	0.010***	0.002	0.014***	0.013***	0.019***	0.013***
General Merchadise Stores	-0.004***	-0.014***	-0.003***	-0.013***	-0.006***	-0.019***
Food Stores	0.003***	-0.002	0.006***	-0.009***	0.007***	-0.026***
Apparel Stores	0.006***	0.003*	0.010***	0.007***	0.009***	0.005**
Population Density	-	-0.005***	-	-0.015***	-	-0.036***
Share of Same Type	-	-0.038***	-	-0.042***	-	-0.069***
Within-CBSA share	0.912***	-	0.879***	-	0.787***	-
R-squared	0.976		0.971		0.946	
Observations	36,103		36,122		37,025	

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 – Low Inc.	-0.152***	-0.157***	-0.180***	-0.230***	-0.106***	-0.095***
Job Opportunities – Mid Inc.	0.104***	0.187***	0.282***	0.392***	0.152***	0.188***
Job Opportunities – High Inc.	0.017***	-0.080***	-0.067***	-0.103***	-0.037***	-0.076***
Avg. Travel Distance	0.019***	0.046***	0.009***	0.026***	0.009***	0.045***
House Price Index	0.001	-0.014***	0.000	0.025***	0.001	0.001
Museums, galleries and libraries	-0.013***	-0.041***	-0.005***	-0.031***	-0.006***	-0.030***
Golf and parks	-0.006***	-0.003**	-0.008***	-0.005***	-0.007***	-0.004***
Gym and sports	0.003	0.004	-0.002*	-0.013***	0.002*	-0.012***
Restaurants	0.008***	-0.008*	0.004***	-0.019***	0.003***	-0.017***
Bars	0.002	-0.012***	-0.005***	-0.031***	-0.002**	-0.010***
Personal Services	0.031***	0.040***	0.021***	0.025***	0.024***	0.040***
General Merchadise Stores	-0.008***	-0.028***	0.000	-0.001	-0.004***	-0.012***
Food Stores	0.028***	0.056***	0.020***	0.046***	0.015***	0.024***
Apparel Stores	0.014***	0.015***	0.018***	0.028***	0.013***	0.003
Population Density	-	-0.051***	-	-0.019***	-	-0.030***
Share of Same Type	-	-0.004	-	0.020***	-	-0.041***
Within-CBSA share	0.639***	-	0.835***	-	0.851***	-
R-squared	0.854		0.944		0.940	
Observations	37,391		37,478		37,565	

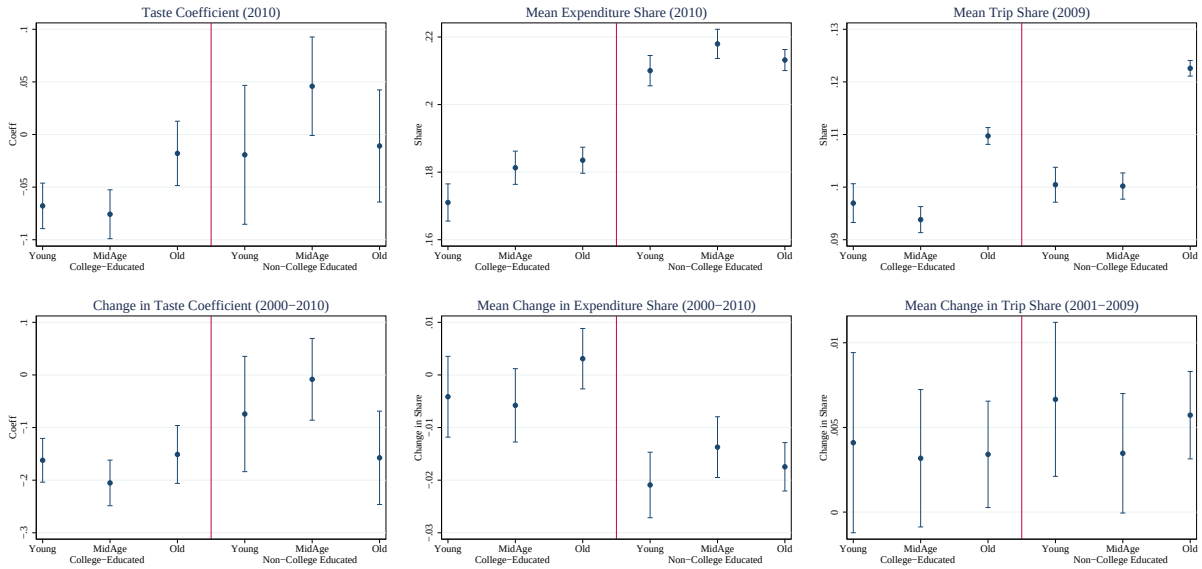
Notes: * – 10% significance level; ** – 5% significance level; ***–1% significance level. The only instrumented variable in this specification is the change in within-CBSA share.

Figure A.1: Comparison of Regression Coefficient Estimates for the Restaurant Amenity Index with CEX Expenditure Shares and NETS Trips Shares on Restaurants.



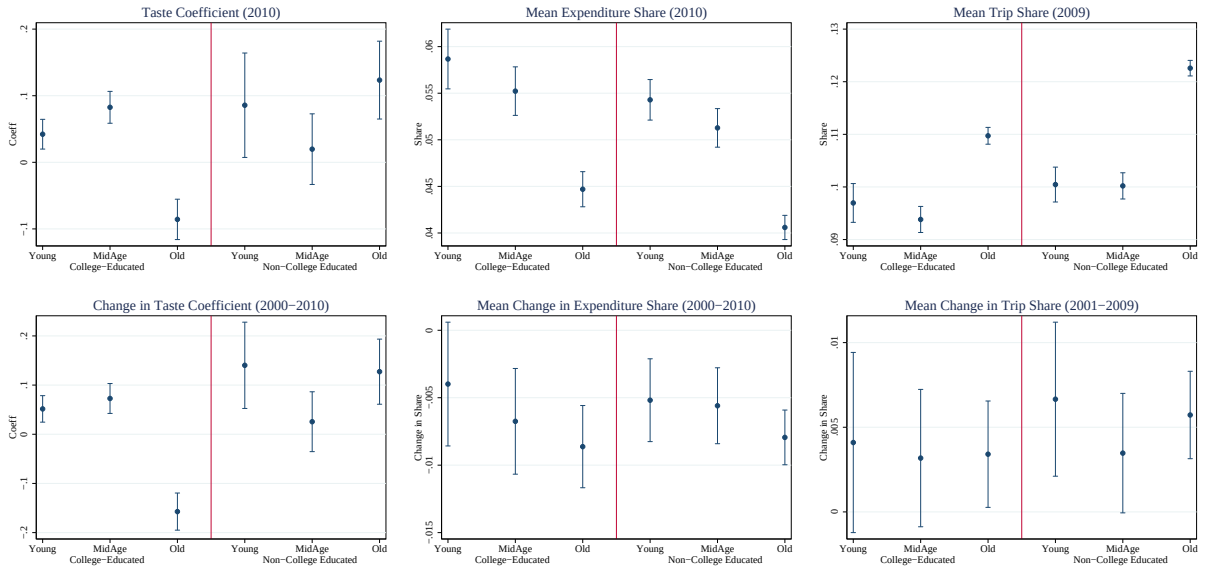
Notes: The first column reports regression coefficients from Table on changes in bars (upper box) and levels of restaurants (lower box) for each age-education group. The second column reports average expenditure shares on restaurants (upper box) and changes in average expenditure shares on restaurants (lower box) from the CEX. The third column reports average trip share to restaurants (upper box) and changes in average trip share to restaurants (lower box) from the NETS data. All confidence intervals are 95% intervals.

Figure A.2: Comparison of Regression Coefficient Estimates for the Food Store Amenity Index with CEX Expenditure Shares on Food and NETS Trips Shares on "Buy Goods" (clothes, food, hardware).



Notes: The first column reports regression coefficients from Table on changes in food stores (upper box) and levels of food stores (lower box) for each age-education group. The second column reports average expenditure shares on food stores (upper box) and changes in expenditure shares on food stores (lower box) from the CEX. The third column reports average trip share to "buy goods" (upper box) and changes in average trip share to "buy goods" (lower box) from the NETS data. All confidence intervals are 95% intervals.

Figure A.3: Comparison of Regression Coefficient Estimates for the Apparel Store Amenity Index with CEX Expenditure Shares on Apparels and NETS Trips Shares on 'Buy Goods' (clothes, food, hardware).



Notes: The first column reports regression coefficients from Table on changes in apparel stores (upper box) and levels of apparel stores (lower box) for each age-education group. The second column reports average expenditure shares on apparel stores (upper box) and changes in expenditure shares on apparel stores (lower box) from the CEX. The third column reports average trip share to "buy goods" (upper box) and changes in average trip share to "buy goods" (lower box) from the NETS data. All confidence intervals are 95% intervals.

fixed-effect, and delivers within-work tract preference coefficients for residential characteristics that are convincingly free of simultaneity with job location.

D.1 Commute Model

The model is similar to that in section 4, but now the location decision of a person is a discrete choice of a single residence-workplace pair. Each person i chooses its residential location j and workplace location k in year t to maximize its indirect utility function V^i_{jkt}:

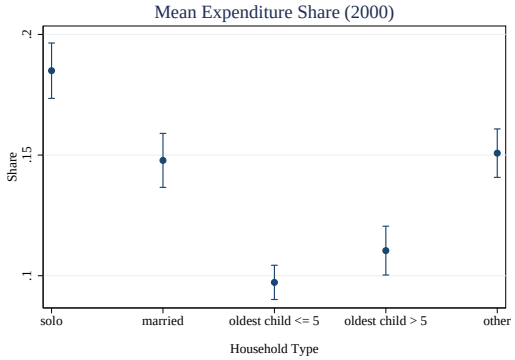
$$\max_{j,k} V^{id}_{jkct} = \alpha^d_t \mathbf{X_{jct}} + \beta^d_t \mathbf{X_{kct}} - \omega^d d_{jkc} + \xi^d_{jct} + \psi^d_{kct} + \mu^d_{jkc} + \theta^d_{ct} + \psi^{id}_{ct}(\sigma^d) + (1-\sigma^d)\varepsilon^{id}_{jkct} \tag{A.2}$$

$\mathbf{X_{jt}}$ and $\mathbf{X_{kt}}$ are vectors of observable time-varying characteristics of residences and workplaces, respectively. d_{jk}, a variable absent from the simple residential choice model, denotes the travel distance from residential location j to workplace location k, and ω^d reflects group d's marginal disutility from commuting. ξ^d_{jct} and ψ^d_{kct} represents the unobserved group-specific, time-varying quality of each residential location and workplace location. To ease the notation, we omit all time-invarying residential, workplace, and residential-workplace unobserved characteristics, because they eventually drop out in first-difference. Exactly as in the residential choice model,θ^d_{ct} represents an unobserved time-varying quality of CBSA c for individuals in group d. We assume a nested-logit error structure, where $\psi^{id}_{ct}(\sigma^d)$ and ε^{id}_{jkct} are a random individual- and time-specific taste shocks for CBSA c and residential-

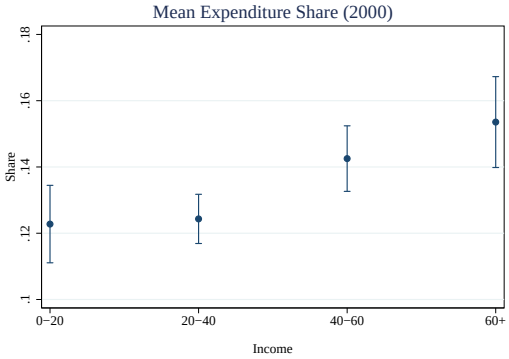
Figure A.4: Average Expenditure Shares and Trip Shares to "Restaurants" by Household Types and Income Bracket for 25-34 Year Old College-Educated Individuals.

Panel A: CEX 2000 Expenditure Shares on Restaurants

By Household Types

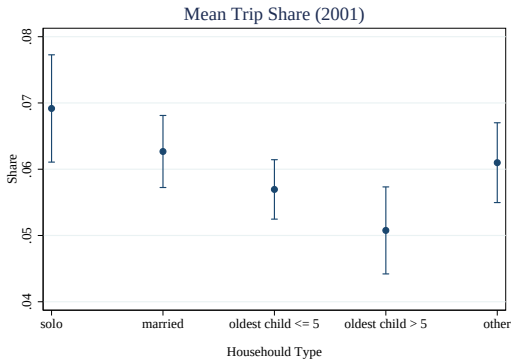


By Income Brackets

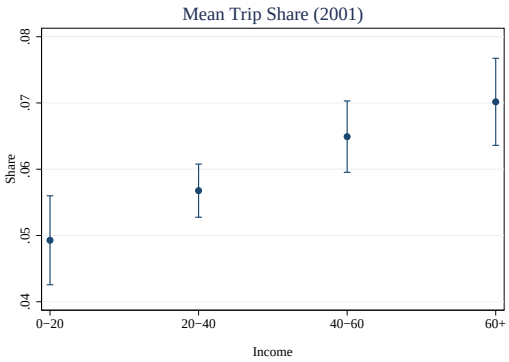


Panel B: NETS 2001 Trip Shares to Restaurants

By Household Types



By Income Brackets

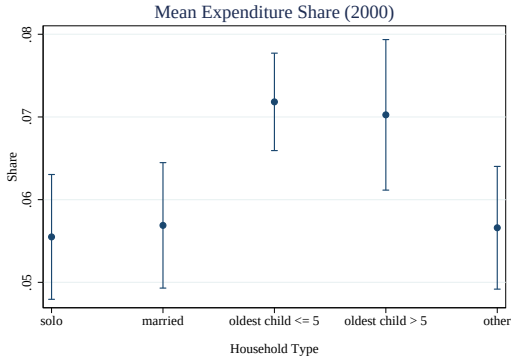


Notes: All confidence intervals are 95%.

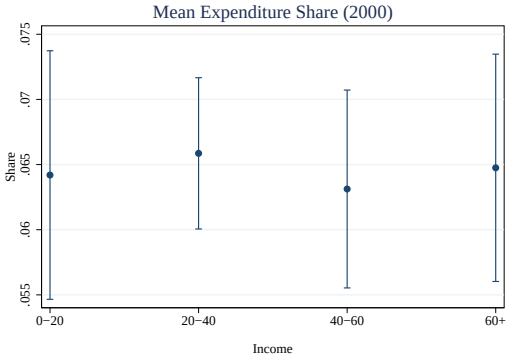
Figure A.5: Average Expenditure Shares Apparel Stores and Food Stores, and Trip Shares to "Buy Goods" (groceries, clothes, hardware) by Household Types and Income Bracket for 25-34 Year Old College-Educated Individuals.

Panel A: CEX 2000 Expenditure Shares on Apparel Stores

By Household Types

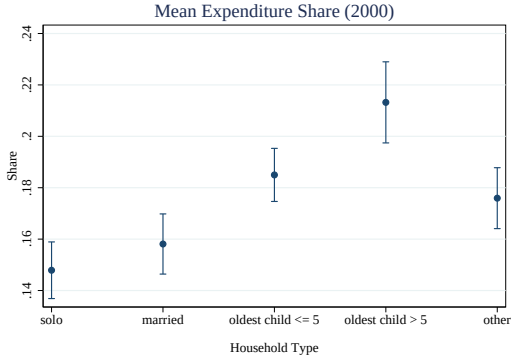


By Income Brackets

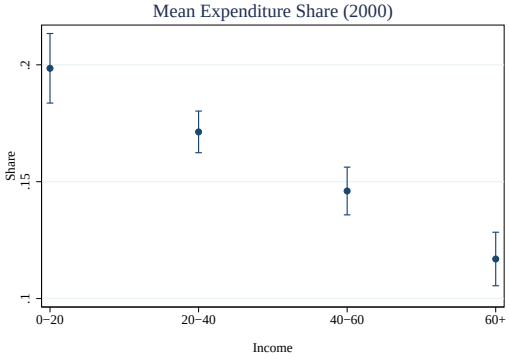


Panel B: CEX 2000 Expenditure Shares on Food Stores

By Household Types

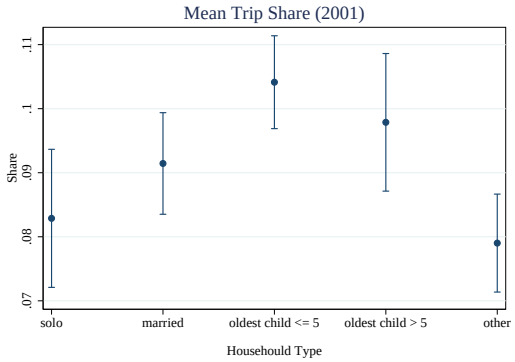


By Income Brackets

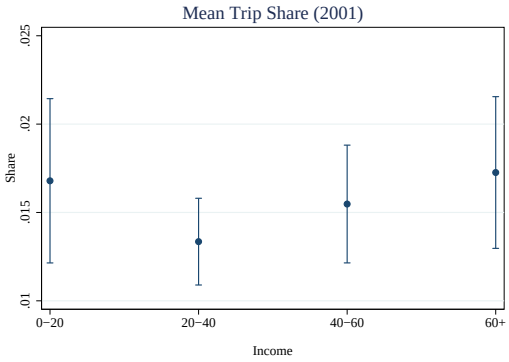


Panel B: NETS 2001 Trip Shares to Buy Goods

By Household Types



By Income Brackets



Notes: All confidence intervals are 95%.

workplace tract pair jk, respectively.[61][62]We solve the model exactly as in section 4 and obtain:

$$\widetilde{\Delta \ln s_{jk}^d} = \alpha_{2011}^d \Delta \tilde{\mathbf{X}}_{\mathbf{jc}} + \Delta \alpha^d \tilde{\mathbf{X}}_{\mathbf{jc,2002}} + \beta_{2011}^d \Delta \tilde{\mathbf{X}}_{\mathbf{kc}} + \Delta \beta^d \tilde{\mathbf{X}}_{\mathbf{kc,2002}} - \Delta \omega^d d_{jkc} + \Delta \tilde{\xi}_{jc}^d + \Delta \tilde{\psi}_{kc}^d + \Delta \tilde{\theta}_c^d + \sigma^d \widetilde{\Delta s_{jk|c}^d} + \epsilon_{jkct}^d$$

Instead of estimating workplace characteristics directly, we add a workplace fixed-effect σ_{kc}^d which captures both observed and unobserved group-specific and time-varying workplace characteristics.[63] The resulting estimating equation is:

$$\Delta \ln \left(s_{jk}^d\right) = \alpha_{2011}^d \Delta \tilde{\mathbf{X}}_{\mathbf{jc}} + \Delta \alpha^d \tilde{\mathbf{X}}_{\mathbf{jc,2002}} + \sigma_{kc}^d - \Delta \omega^d d_{jkc} + \Delta \tilde{\xi}_{jc}^d + \Delta \tilde{\theta}_c^d + \sigma^d \Delta s_{jk|c}^d + \epsilon_{jkct}^d \qquad \text{(A.3)}$$

.

D.2 Commute Model Variable Definition

Before estimating the model, we describe all the variables in equation A.3 were not in the residential model of Section 4.

Commute Shares The dependent variable in the commute model is the 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 year t in CBSA c. We obtain these numbers from the LODES data in 2002 and 2011 for high-income, medium-income and low-income workers. Let c be the CBSA of tract k and $L_c$be the set of tracts located in CBSA c. The the share of CBSA c workers who live in tract j and work in tract k in year t:

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

Commute time In the current draft, we proxy for the commute time between the workplace and residence tract d_{jkc} using a flexible quadratic function of the Haversine distance between workplace and residence tracts. That is, we include both the level and the squared distance between tracts j and k in the estimating equation. In future drafts, we will replace this proxy with a snapshot of driving and transit times collected from Google Maps.

[61]We assume that (i) that people select both their place of work and residence simultaneously and (ii) each person gets an independent residential-workplace pair taste draw in each period. We use long differences in our estimation, making each of these assumptions more plausible. The transportation literature has explored richer substitution patterns allowing for sequential decision-making (Waddell et al., 2007) and joint-location decisions for households with multiple workers (Waddell, 1995). By ignoring each of these factors, we are overestimating the flexibility of workers in moving to their optimal workplace and residential location pair, perhaps moreso for older workers who are more likely to live in larger households and be tied to residential locations that are convenient to the workplaces and schools of family members.

[62]The logit distribution on the random taste shocks imposes the independence of irrelevant alternatives (IIA) property on within-CBSA individual location choices. This property implies that, when agents substitute away from one option within a CBSA tract-pair choice set, they substitute to all other options in equal proportions - regardless of how similar those alternatives are to the alternative that agents are substituting away from.

[63]We focus on estimating the effect of residential characteristics separately from the effect of changes in workplace location, but our model can also be fully estimated without the workplace tract fixed-effect. In this case, the number of jobs in tract k - possibly excluding own workers in own tract j - becomes a workplace characteristics. The coefficient on this variable provides a measure of the impact of workplace reallocation in space on residential choices. This measure is valid under an independent of irrelevant alternative (logit) assumption. That is, the effect of an increase in jobs in tract k on 2002 to 2011 changes in the number of individuals in tract pair jk depends only on the initial 2002 share of people in jk, and is directly proportional to that original share. In this model, $\Delta \ln \left(s_{jk}^d\right)$ depends on the characteristics of tract j and tract k, but not on that of any other tracts.

Alternatively, we could infer a time varying measure by using tract-to-tract commuting times in 2000 and 2010 from the Census Transportation Package (CTP).

Residence Tract Characteristics Our CBSA residential characteristics are the same as in the residential choice model of Section 4. Note that the variables for job opportunities and average distance to work now take an interpretation as a purely residential characteristics. These variables capture the possibility that households choose residential locations based on their proximity to employment locations other than their own, as such job opportunities may become relevant to future career events. We control for local demographic shares in 2002 defined by income groups instead of age-education group, and we derive these shares from the LODES data.

D.3 Commute Model Identification

As in section 4, the identification strategy for the commute model relies on first-differencing, the addition of a rich set of controls, and the set of instrumental variables described in that section. However, the specification in equation A.3 provides and additional, sharper way of controlling for the simultaneous determination of workplace and residential location changes. This simultaneity problem is straightforward; for instance we expect high-income workers to move to areas that experience an influx of firms hiring them, to reduce their commute costs. The reverse is also true; we expect firms hiring high-income workers to move to areas that experiences an influx of these workers, as a mean of attracting talent. Moving closer to a young, educated talent pool is often the stated objective of employers like Amazon, Twitter or Google when they move to new downtown offices (Johnson and Wingfield (2013)). Our work-tract fixed-effect specification solves this simultaneity problem by delivering the within-work tract impact of residential characteristics i.e., by considering the change in residential choice of people working within the same tract in 2002 and 2011. In this case, changes in residential location are not affected by a change in workplace location. In terms of equation A.3, the workplace fixed-effect captures all unobserved changes in workplace characteristics ($\Delta\tilde{\psi}^{d}_{kc}$).

D.4 Commute model results

TBD

E Amenity Indices Methodology

In this Appendix we provide additional details on the methodology to derive the amenity indices, the restaurant quality index and the instruments for these variables.

Consumption amenity index

The consumption amenity indices measure the availability of 9 different categories of establishments around the centroid of each census tract. For each category a the amenity index in tract j is the invert of a CEX price index:

$$A_{aj} = \frac{1}{\left(\sum_{i=1}^{I_j}(p_a + 2(t_{ij}))^{1-\sigma}\right)^{1/1-\sigma}}, \tag{A.4}$$

where p is the average price of visit to an establishment in amenity category a,[64] t is the travel cost of a one-way trip to establishment i from the tract centroid j, I_j is the set of all establishments in category a within 50 miles of a tract, and σ is the elasticity of substitution equal to 8.8.[65] These amenity indices are discussed in further details in Couture (2013).

We compute travel costs t_{ij} using data from Google Maps. We start by computing the linear distance from tract j's centroid to an establishment i. To go from linear to actual travel distance, we use an average ratio of travel time to travel distance by car, transit and foot, computed from a random sample of 200 NETS establishments at various distances from that tract's centroid. To go from travel distance to travel time, we proceed in three steps. First we create a random trip sample to 200 establishments at different distances from each tract's centroid. Second, we use Google Maps to obtain distance/time pairs for these trips and then use these pairs to estimate log-on-log functions relating trip time to trip distance in each tract (as in Couture et al. (2016) for car and transit. For the indices by foot that we use in the paper, we use Google Maps constant speed of 20 minutes per mile. Third, we use these functions to assign a travel time to each establishment for each mode, starting from a tract's centroid. To go from travel time to travel cost we use a value of time of \$12 per hour or about 50% of the average US wage as suggested in Small and Verhoef (2007). For car trips we add fuel cost of \$5 per hour to the transport cost.

Restaurant quality index and independent establishment index.

The restaurant quality index of Subsection 5.1 is a weighted average of ESRI's MPI ratings for all 61 rated restaurant chains near a tract. The CES weights are lower for establishments farther away from a tract's centroid, and for tract j the quality index Q is computed as follows:

$$\mathrm{Q}_j = \frac{\sum_j MPI_i \times (p + t_{ij})^{(1-\sigma)}}{\sum_j (p + t_{ij})^{(1-\sigma)}} \tag{A.5}$$

Where MPI_i is the MPI rating of restaurant chain i, p is the average price of visiting a restaurant, t_{ij} is the total transport cost of a trip to establishment i from tract j, and σ is the elasticity of substitution equal to 8.8.

The independent establishment index of Subsection 5.3 is a weighted average of the share of independent establishments near a tract. We compute it using equation A.5, except that instead of using an MPI rating we assign a rating of 0 to all establishments that are part of chains with at least 5 members, and a rating of 1 to all other "independent" establishments.

Instruments Our instrument for the 2000 to 2010 change in the amenity index is a prediction based on the 2000 business environment. The instruments' computation for each amenity index proceeds in two steps:

1. First, we regress SIC8-level establishment entry from 2000 to 2010 in each tract on variables capturing the pre-existing commercial environment in 2000. To compute the instruments for the restaurant quality index, we predict entry at the chain level instead of the SIC8 level.[66]

[64] We derive these average prices using expenditure categories in the CEX data that most closely matches 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 stores, \$36.5 for food stores and \$60.4 for apparel stores.

[65] 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.

[66] There are various reasons why entry might be correlated with the existing commercial environment. Looking within a small radius in close proximity to a given location, we expect that entry will be decreasing in the concentration of establishments offering similar services, due to competition and cannibalization concerns. On the other hand, once we control for the density of existing businesses in close proximity to a point of entry, we expect that entry will be increasing in the broader density of existing establishments under the same chain, since this will indicate proximity to the chain's upstream suppliers or distribution centers and some pre-existing market knowledge. In addition to these within-chain scale economies, we also account for sector-level co-agglomeration externalities, in the form of positive spillovers from local

2. Second, we use the fitted values of the regression in step 1, which predict establishment entry in each tract, to obtain predicted changes in each amenity index. This predicted change in the index is our instrument.

We now describe the first step regression which predict exit and entry. We describe the chain-level regressions, because the SIC8 level regressions (e.g. predicting entry of Korean restaurants) are a straightforward simplification of the chain-level regressions without the chain level variables. We model the entry and exit decision of each chain as a function of the business environment, more precisely of the number of establishments at different distance from the tract centroid that are in the same chain, in the same SIC8 code but not in the same chain, in the same SIC6 code but not in the same SIC8 code, and in the same SIC4 code but not in the same SIC6 code.[67] We experimented with including controls for the presence of establishments that are plausibly wholesalers for a given chain or SIC8 but ultimately none of these coefficients were significant and we removed these variables from our regressions. Define n_{jt}^{C} as the number of establishments in a chain C in tract j in period t. Let $n_{jt,dist}^{sic\#(C)}$ be number of establishments in the same SIC# code as chain C (where # takes value 4, 6, 8 and 10) within distance interval 2^{dist} to 2^{dist+1} (where $dist$ takes values from 0 to 3) from the centroid of tract j. Note that SIC10 codes are not defined by the Department of Labor; we assign an $SIC10(C)$ code to each chain to simplify the notation. We obtain the following estimating equation, in which each observation is a tract:

$$
\begin{aligned}
n_{j10}^{C} - n_{j00}^{C} = \alpha^{C} + \sum_{dist=1}^{3} \beta_{dist}^{sic\#(10)} \left(n_{j00,dist}^{C} - n_{j00,(dist-1)}^{C} \right) + \\
\sum_{dist=1}^{3} \beta_{dist}^{sic8C} \left(\left(n_{j00,dist}^{sic8C} - n_{j00,dist)}^{C} \right) - \left(n_{j00,(dist-1)}^{sic8C} - n_{j00,(dist-1))}^{C} \right) \right) + \\
\sum_{dist=1}^{3} \left(\sum_{\#\in\{6,4\}} \beta_{dist}^{sic\#(C)} \left(\left(n_{j00,dist}^{sic\#(C)} - n_{j00,dist)}^{sic(\#+2)(C)} \right) - \left(n_{j00,(dist-1)}^{sic\#(C)} - n_{j00,(dist-1)}^{sic(\#+2)(C)} \right) \right) \right) + \varepsilon_{j}^{chain}.
\end{aligned}
\tag{A.6}
$$

Table A.4 provides information on the variables in equation A.6 that are significant predictors of entry and exit across all regressions. Remind that we use chain-level predictions to compute the instrument for changes in restaurant quality, and SIC8 level predictions for changes in the 9 standard amenity indices. The table displays results aggregated over all 1078 SIC8 codes - corresponding to 1078 regressions - for establishments included in our 9 amenity indices. The first column provides the percentage of SIC8 codes in which a control variable - such as the number of establishments in the same SIC6 but not the same SIC8 within 2-4 miles of a centroid - has a positive and significant coefficient at the 10% level, and the second column provides the percentage of negative and significant coefficients for that variable. The results highlight the local nature of entry and exit decisions, with coefficients on variables computed within 0-1 miles from a tract's centroid more likely to be significant that those capturing establishments farther away. The first striking result in Table A.4 is the importance of competition and cannibalization concerns; for 93% of SIC8, the presence of establishments in the same SIC8 in 2000 within 0-1 miles significantly reduces entry in a tract from 2000 to 2010. The results also points to the importance of agglomeration forces with establishments in related but not similar product space. The 0-1 miles coefficients on establishments in the same SIC6 but other SIC8s are positive and significant in 47% of cases and negative and significant in only 6% of cases, and at the SIC4 level these numbers are 52% positive and significant versus only

activity from non-competing or differentiated firms within the same industry. In addition to these direct effects, we expect that the existing landscape captures location-specific barriers to entry, such as existing density and either natural or regulatory supply constraints, as well as direct effects.Finally, proximity to wholesalers may lower the cost of entry for establishments.

[67] To simplify the organization of the data, we create variables for the SIC8 level and chain-level regressions sample from two different sets of establishments. Variables for the SIC8-level regressions are created from all establishments within each amenity category, while variables for the chain-level regressions includes the subset of restaurants that are MPI rated (which assumes that chains make their entry decision based only on the presence of other chains.)

10% negative and significant.

Table A.5 presents similar results for entry and exit regressions on the 61 restaurant chains for which we have MPI ratings. The results feature similar local cannibalization and agglomeration concerns, with coefficient on the number of establishments between 0-1 mile of a tract's centroid within the same chain significantly reducing entry for 89% of chains, and coefficients on same SIC4 and same SIC6 much more likely to be positive and significant. Interestingly, the same chain coefficients features a strong reversion from negative to positive sign as distance between a tract and an establishment increases, implying that while chains avoid locating near an existing outlet, they still prefer to locate in markets that they have already penetrated. A similar but weaker pattern is also present at the SIC8 level in Table A.4. [68]

In the second step of our methodology we use the fitted value of these entry and exit regressions to compute predicted changes in the amenity index. To compute the instrument for category a we start from the vector of all establishments in that category in 2000, and use the fitted value from the first step regression to add predicted establishments to the centroid of each tract and therefore obtain a vector of "predicted" 2010 establishments. From that vector of predicted establishments, we compute a predicted amenity indices for 2010 using equation A.4. The difference between the predicted amenity index in 2010 and the actual amenity index in 2000 is our instrument for the change in the amenity index from 2000 to 2010 in amenity category a in tract j. We use the same procedure to obtain the instrument for restaurant quality based on fitted value of the chain-level entry and exit regressions.[69]

[68] Note that contrary to the SIC8-level regressions, in the chain-level regressions the coefficient for the number of establishments with the same SIC8 code within 0-1 mile is *positive* and significant in 52% of cases. This suggests that restaurant chains are more differentiated in their products and services relative to all other establishments, which could explain why having a very similar firm of the same SIC8 code very nearby switches from a cannibalizing force (predicting establishment exit) in the SIC8-level regressions to an agglomerating force (predicting entry) in the restaurant chain-level regressions.

[69] When adding fitted value of the change in establishment at the chain level, each chain receives a different MPI rating. In some rare cases, the predicted 2010 establishment vector includes a "negative" amount of a chain, in cases with strong prediction of exit and few existing chains in 2000. It is impossible to compute and index with negative MPI weights and therefore we limit the number of such occurences by aggregating chains into ten MPI deciles when we compute the instrument. If a bin with negative value persists we simply remove it.

Table A.4: Tract-level Predicted Establishment Entry at the SIC8 Level.

	Percentage of SIC8-Specific Coefficients		
	Negative and Significant	Positive and Significant	Not Significant at 90% 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 level regression in Equation A.6 and provide the percentage of cases, out of all 1078 SIC8 codes within our 9 amenity categories, in which each variable is significant at the 90% level.

Table A.5: Tract-level Predicted Establishment Entry at the Chain Level

	Percentage of Chain-Specific Coefficients		
	Negative and Significant	Positive and Significant	Not Significant at 90% 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 level regression in Equation A.6 and provide the percentage of cases, our of 61 restaurant chains for which we have MPI data, in which each variable is significant at the 90% level.