Does Rail Transit Investment Encourage Retail Activity?

Final Report

METRANS Project 11-04

August 2013

Principal Investigator:
Jenny Schuetz

Price School of Public Policy
University of Southern California
Los Angeles CA 90089

METRANS Transportation Center
USC CSULB

Disclaimer

Abstract

Over the past 20 years, California has made substantial investments in intra-metropolitan passenger rail infrastructure, expanding existing systems and building new ones. Such investment has the potential to encourage the growth of mixed-use transit-oriented development, defined as a high-density mix of residential and commercial uses within walking distance of rail stations. Previous studies have attempted to identify the impacts of rail transit investments on land values, residential real estate, and population characteristics. However, little research to date has examined whether rail investment stimulates economic development in the form of retail activity. In this paper, I use geocoded data on the location of retail establishments in California from 1992-2009 to examine whether newly opened rail stations have resulted in a net growth in surrounding retail activity. Results indicate that new rail stations were located in areas with initially high employment density, somewhat outside the city centers, with relatively low household incomes and property values. The impact of new stations on nearby retail activity has been highly varied: while new station openings are associated with increased retail employment in the San Francisco MSA, retail employment decreased around new stations in Sacramento, and did not significantly change in Los Angeles or San Diego.

Table of contents

List of figures and tables

Acknowledgements

Funding for this project was provided by the METRANS Transportation Center at USC and the Lusk Center for Real Estate. Thanks to Gen Giuliano, Peter Gordon, Chris Redfearn and Lisa Schweitzer for helpful comments. Marie Sullivan provided invaluable assistance with the GIS analysis. Thanks also to Yuting Hu, Sanggyun Kim and Vincent Reina for excellent research assistance.

Section 1) Introduction

According to advocates of "New Urbanism", one of the goals of public investment in rail transit is to encourage the growth of mixed-use transit-oriented development, defined as a high-density mix of residential and commercial uses within walking distance of rail stations. In theory, by reducing transportation costs for residents in the surrounding neighborhood, new transit stations should result in increased land values, spurring higher density development and higher-value uses. A fairly broad empirical literature has attempted to identify the impacts of rail transit investments on surrounding neighborhoods, measured by a variety of outcomes, including land values, housing prices, population and housing density, employment composition and population characteristics (Baum-Snow and Kahn 2005; Bollinger and Ihlanfeldt 1997; Ewing and Cervero 2010; Giuliano and Agarwal 2010; Kahn 2007; Winston and Maheshri 2007). However, almost no research to date has examined the impacts of rail investment on one of the key components of mixed-use development, namely retail activity.

As a land use class, retail should be highly compatible with neighborhoods surrounding rail stations: the increased pedestrian traffic generated by transit riders should increase retail business, whereas high traffic may be considered a negative amenity for residential development. Moreover, municipalities offer give planning preference to commercial developments over residential ones because of the fiscal benefits generated by business taxes (Gruen 2010). The presence of retail services in a neighborhood, such as grocery stores, pharmacies and restaurants, has important quality of life implications for residents, for instance whether they can purchase healthy food choices at reasonable prices (Hayes 2000). Therefore the extent to which rail investments improve the quantity or quality of retail services in surrounding neighborhoods is potentially a valuable social benefit that might serve to justify public expenditures. This is particularly true for rail investments in low-income and minority neighborhoods, which tend to have fewer retail and household service establishments, and where those that do exist offer inferior quality goods at higher prices (Alwitt and Donley 1997; Bartie et al 2007; Carr and Schuetz 2001; Sloane et al 2005). Moreover, because, low-income households are typically less likely to own automobiles, they may face greater barriers to reaching commercial centers outside their immediate neighborhoods.

In this project, I will examine how public investments in rail transportation in five large metropolitan areas in California have affected the quantity and type of retail services in surrounding neighborhoods. As shown in Table 1, over the past 20 years, California's largest metropolitan areas have expanded existing rail systems or built new ones, offering an opportunity to investigate changes in retail activity surrounding new rail stations and along rail lines. There is considerable variation in the types of neighborhoods that received new transit stops, both within and across MSAs, offering the chance to investigate differential impacts of transit on retail by prior economic and demographic characteristics, such as resident income and racial composition. Of particular interest is whether investment that took place in historically poor and underserved areas has spurred additional economic growth in these areas.

The urban economics literature provides several models of retail firm location that provide a theoretical framework for why the presence of public transportation should affect the amount and composition of local retail outlets. Hotelling (1929) develops a simple spatial model of retail location, which suggests that the density of stores within a given geographic area depends on store fixed costs, buyer density, transportation costs, and frequency of purchases. Intuitively, high store fixed costs serve as a barrier to entry, reducing the number (density) of stores. Higher density of buyers within an area increases returns to stores and encourages entry.

High transportation costs will discourage consumers from traveling long distances, leading to smaller market areas. Store networks will also be denser for more frequently purchased items, such as perishable or frequently consumed goods.

Under this model, investments in rail infrastructure could affect retail activity in several ways. The addition of a new rail station to a neighborhood decreases transportation costs between neighborhoods that are connected by the rail line, thus expanding the market area of stores at each stop. Within the neighborhood surrounding the station, if the rail station attracts additional transit users, it will effectively increase the buyer density in the neighborhood, and so should lead to an increase in the number (density) of stores nearby. A larger consumer base may also encourage product differentiation, leading to greater diversity of store types by goods and services offered. Previous empirical studies on the impacts of rail stops on other outcomes, such as property values, population and employment, have shown that the extent of impacts depend crucially on increases in ridership (Baum-Snow and Kahn 2005; Kahn 2007; Bollinger and Ihlanfeldt 1997). Therefore whether rail stations result in an increase in the quantity (or quality) of surrounding retail will depend on the number of transit users at that location. Moreover, an increase in retail establishments near a newly built train stop could represent either a net increase in retail activity through new store creation or redistribution from other, less accessible sites, as stores relocate closer to the rail station.

In this analysis, I combine geocoded data on retail establishments from the National Establishment Time Series (NETS) database with the location and opening dates of more than 500 rail stations throughout California's large metropolitan areas. I examine differences in neighborhood characteristics for the one-quarter mile areas around rail stations, comparing new station neighborhoods to areas around older station as well as designated census tracts that are never within one-half mile of any rail stations. I also test whether the opening of new stations has been associated with a growth in retail activity immediately around the stations, relative to control tracts and older station areas. Results indicate that new rail stations were located in areas with initially high employment density, somewhat outside the city centers, with relatively low household incomes and property values. Analysis of retail employment suggests that the impacts of new stations on retail activity near the stations has been highly varied: while new station openings are associated with increased retail employment in the San Francisco MSA, retail employment decreased around new stations in Sacramento, and did not significantly change in Los Angeles or San Diego. The differences in growth of nearby retail may reflect baseline heterogeneity in characteristics of both stations and surrounding neighborhoods, such as proximity to CBD and ridership levels.

The remainder of this paper is organized as follows. Section 2 provides background on the rail systems that form the setting for the analysis and lays out hypotheses for how station characteristics might affect surrounding retail growth. Section 3 describes the data and empirical methods. Section 4 presents results of descriptive statistics, while Section 5 offers directions for future research and discusses policy implications.

Section 2) Background on California rail systems

This paper focuses on fixed-line transit systems in California's four largest metropolitan areas: Los Angeles-Long Beach, Sacramento, San Diego, and San Francisco-San Jose. As of the end of the study period in 2009, these four MSAs had a total of 520 stations belonging to eleven different fixed-line transit systems. Most of the systems are light rail lines, there are some heavy rail lines (BART, Caltrain, and the LA Metro Red and Purple lines) and one bus rapid transit

(BRT) system (the MTA's Orange Line). This section describes the growth of rail systems during the 1992-2009 period, and discusses how the characteristics of the stations and the rail systems might affect their impact on surrounding development patterns.

Data assembly

Data on the location and characteristics of transit stations was assembled from a variety of sources, primarily the websites of the various transit operators. Station names and locations were collected from system maps, along with indicators for all systems and lines using each station. Addresses were geocoded to obtain latitude and longitude coordinates and match stations with census tracts. Information on the initial year of rail service at each station was obtained from transit operator websites and supplemental sources, including Amtrak's Great American Stations website and contemporaneous media coverage.[1]

Expansion of California rail systems

Most of California's metropolitan areas are relative newcomers to rail transit systems, compared to East Coast cities: Baum-Snow and Kahn (2005) find that San Francisco is the only California city in the group of seven "legacy" MSAs with rail transit in 1970. But all four of the state's largest MSAs had some rail stations as of 1990 (Table 1), and nearly all of the transit systems expanded between 1992 and 2009.

The new stations represent various types of growth. LA's Metropolitan Transit Authority (MTA) built several new subway and light rail lines and a new BRT line. Most of these stations connect neighborhoods within the City of Los Angeles, but two lines extend to the nearby cities of Long Beach and Pasadena. The other main rail system in the LA region is the Metrolink commuter rail system, operated by the Southern California Regional Rail Authority. The entire Metrolink system was developed between 1992 and 1995, and serves all five counties in the LA

[1] http://www.greatamericanstations.com/

CMSA (Los Angeles, Orange, Riverside, San Bernadino, and Ventura). The Sacramento Regional Transit District operates a light rail system, begun in 1987 and expanded during the study period. San Diego's Metropolitan Transit System operates three light rail lines under the San Diego Trolley; two lines were built in the 1980s, the third was added in 1990 and expanded throughout the study period. The San Diego region is also served by the North County Transit District's Coaster commuter rail, built in 1995.

The San Francisco-San Jose MSA has the oldest systems and the most extensive transit network in California. The San Francisco Municipal Railway System, or MUNI, is a system of buses, trolleys and cable cars, some of which have operated along the same routes for nearly a century, but with expansions continuing through the 1990s and 2000s. MUNI stops are only within the city of San Francisco. The Bay Area Rapid Transit (BART) is a heavy rail system running throughout the region, including the East Bay cities of Oakland and Berkeley; the earliest BART stations were built in the 1970s, within expansions continuing to current years. The Santa Clara Valley Transportation Authority operates three light rail lines, primarily in the city of San Jose, but extending to several other cities in Santa Clara and San Mateo counties. These stations opened between 1987 and 2005. A regional commuter rail, Caltrain, runs through the spine of the Bay Area, from San Francisco to San Jose and extending south to Gilroy. Most Caltrain stations were open before 1992, with a few later expansions. The newest and smallest rail system in the Bay Area is the Altamount Commuter Express, which began operating in 1998, from San Jose north to Fremont and extending east to Stockton.[2] The Bay Area has by far the highest share of transit riders among the four MSAs in this study, although ridership varies across counties as well: roughly one-third of commuters in San Francisco County ride public

---

[2] The three easternmost stations on ACE – Stockton, Lathrop/Manteca and Tracy – fall outside the counties in the San Francisco-San Jose CMSA as so are excluded from analysis in this paper.,

transit to work, and 11 percent in Alameda County, compared to under seven percent in Los Angeles County and about three percent in San Diego and Sacramento Counties.[3]

As shown in Table 2, approximately one-quarter of all stations and 13 percent of newly built stations serve multiple rail lines (this includes multiple lines within the same transit system and multiple systems), raising some empirical challenges in defining the date of station "opening". For the purposes of analysis, any station that served one or more of the listed transit systems prior to 1992 is considered to be an existing station through the entire period of analysis. Newly opening stations from 1992 to 2009 either did not exist or did not serve any of these rail stations prior to 1992. Although not part of the analysis, 31 of these stations are also served by Amtrak; Amtrak passenger service either pre-dated 1992 or was added to new stations after the local rail service began. The first year of rail service for some pre-1992 stations could not be determined, particularly for the MUNI stations. A further complication to identifying an "opening" year, some of the transit systems made use of pre-existing stations built as part of earlier passenger or freight rail systems (Table 3). A review of station histories provided by Amtrak indicates that many of these stations were built prior to 1940 as part of inter-city passenger rail services, most notably belonging to the historic Atchison, Topeka and Santa Fe Railway. Although operating service to some of these stations had ceased well before the current transit systems formed, some portion of the station and track infrastructure remained.

As discussed in more detail below, the reuse of existing and historic stations has several potentially important implications for retail growth subsequent to station re-opening. First, the locations of these stations were not selected based on current economic conditions; they may not be in the optimal location relative to current distribution of population and non-retail activities.

[3] Transit ridership shares calculated from 2010 ACS data; numbers represent percent of workers aged 16 or more who do not work at home.

Second, the built environment surrounding older stations is more likely to be already developed than the surroundings of greenfields stations, which may make these areas more costly to adapt (although older stations may have higher density of retail prior to station opening, as a legacy of the existing development patterns). Third, for older stations that are considered architecturally or historically important, there may be historic preservation mechanisms that hinder nearby redevelopment (a number of the older stations are listed on the National Registry of Historic Places).

Could station characteristics impact nearby retail development?

As an empirical setting for studying the impact of rail stations on nearby retail activity, California's cities offer considerable diversity, not only in opening dates, but in the characteristics of stations and transit systems. Variation in station characteristics may plausibly be expected to impact nearby economic activity, directly through physical context and indirectly through differences in transit ridership (see Voith 2005). Broadly speaking, the type of station attributes that may affect retail growth can be grouped into four categories: station physical characteristics, neighborhood physical characteristics, neighborhood economic & demographic characteristics, and neighborhood regulatory and political environment.

Comparing two stations along Metrolink's Orange County Line illustrates how station and neighborhood physical conditions might impact nearby development. The Fullerton station, originally developed in 1923 by the Union Pacific Railroad, is built in Spanish Colonial style, with decorative architectural features. It is located in a fairly dense commercial area near the city's downtown, and ties directly into the surrounding street grid. The Laguna Niguel/Mission Viejo station does not have a station building, just a boarding platform, surrounded by large surface parking lots, lying between two major freeways. While the Fullerton station could easily be incorporated into a New Urbanist-style, pedestrian-friendly transit oriented development, the location and physical characteristics the Laguna Niguel station would make such development difficult, if not impossible. More generally, the physical features of the station and its surrounding neighborhood offer different opportunities for nearby development. Most light rail stations and tracks lie above ground, creating a break in the street grid on either side of the station, relative to underground subway systems. Like the Laguna Niguel stop, many stations along commuter rail lines are surrounded by large surface parking lots, to accommodate park-and-ride passengers. Several rail lines, such as the MTA Green Line, run down the median of freeways, again forming a barrier to immediately contiguous development.

A priori, it is unclear whether newly opening stations built in older, denser neighborhoods should experience more retail growth than stations built in low-density, undeveloped areas. On the one hand, neighborhoods that already have high density of either residential or commercial uses have existing infrastructure and may provide more critical mass of consumers (see Brooks and Lutz (2013) for the example of persistent development patterns around Los Angeles streetcar lines). On the other hand, building new stores might require redevelopment of existing structures, a more time-consuming process. Older neighborhoods may also have more fragmented land ownership, requiring costly and lengthy parcel assembly for large-scale projects. Therefore the relationship between prior land use patterns and retail growth must be examined empirically.

In general, we would expect economic and demographic characteristics to affect the likelihood of retail growth around a station. Households with more disposable income or wealth represent desirable consumers for retailers. Previous research has shown that some population characteristics, such as share of college-educated population, share of households with children, and share of owner-occupied homes, are negatively correlated with retail activity, potentially because these households prefer neighborhoods that are exclusively residential (see Schuetz et al 2012, Waldfogel 2006). Higher income households are also more likely to own cars, and may find rail transit less of an amenity than households who depend on public transit.

Two attributes of stations which I cannot directly observe are likely to affect the prospects for retail growth around new stations. First, the regulatory and political environment will affect the potential not only for retail, but for any type of transit-oriented development. If a new station is opened in an area not zoned for commercial activity, or zoned for low density development, then little retail growth is likely to follow. Conversely, local officials may offer incentives for projects near transit, relative to other locations; such incentives may include density bonuses in designated areas, reduced permitting fees or fast-tracking of development approvals. If developments near stations receive preferential treatment, it is unclear whether TOD projects represent a net increase in jobs or merely a shift away from other locations. I do not have zoning data, so will rely on proxies such as non-retail employment density prior to station opening and composition of the housing stock to infer zoning constraints. Possible approaches to testing whether TOD projects receive preferential treatment will be discussed in more detail in the empirical strategy section below. Second, stations with higher ridership should be more attractive to retailers, because of the higher volume of consumers. Time-varying ridership data per station is not available. My closest proxies are the number of lines serving each station and the density of nearby stations; rail transit will be a more desirable mode of transportation at stations that are part of a denser network, offering access to more destinations. Future analysis will use neighborhood-level measures of commuting mode share as additional proxies for ridership levels.

Section 3) Methodology and data description

This paper seeks to determine whether investments in public transportation have improved access to retail services by increasing the amount of retail activity in neighborhoods surrounding new rail stations. The empirical strategy is a modified difference-in-difference approach, which compares the change in retail employment density near newly opening transit stations, before and after the stations open, to changes in retail employment density for two sets of control neighborhoods. The transit literature generally finds that most potential riders will come from a one-quarter mile radius of the station, with impacts declining rapidly beyond that (see Kolko 2011), so I define the primary treatment area to be within one-quarter mile radius of each newly opening station.[4] Stations in existence the entire study period (1992-2009) form one set of control areas. The second set of control areas are census tracts outside a half-mile boundary of any transit station (new or existing), so not likely to be directly affected by station openings, but within the same PUMA as at least one rail station. Limiting control areas to the same PUMA should yield comparison groups that have similar demographic and economic characteristics to the treatment areas, as well as similar proximity to spatially-specific amenities (access to highways, distance to central business district, etc.). Figures 1 and 2 show the quarter mile boundaries for areas around selected transit stations, as well as the control tracts that fall outside the larger half-mile boundary.

The general form of the regression to be estimated is shown below:

$$\textit{Eq. (1)} \qquad EmpDens_{it} = \beta_0 + \beta_1 Station_i + \beta_2 PostStation_t + \beta_3 X_{it} + \varepsilon_{it}$$

where i indexes the neighborhood and t indexes the year. The dependent variable, *EmpDens*, is the number of retail employees per square mile in the subject neighborhood. *Station* is a dummy

[4] The simplest analysis will use half-mile radius circles around stations as treatment areas; other estimates will choose census blockgroups or tracts with new stations as the treatment groups, to allow use of economic and demographic control variables.

variable indicating whether a station ever exists the neighborhood and *PostStation* indicates the presence of a new station after opening. That is, for control areas, both *Station* and *PostStation* equal zero at all times. For older station areas, *Station* equals one while *PostStation* equals zero at all times, while for new stations *Station* always equals one and *PostStation* switches from zero to one after the station opens. Thus the coefficient β_2 is the main coefficient of interest, estimating the change in retail employment density for areas near stations after the station opens. X is a vector of control variables, such as non-retail employment density, population density and demographic characteristics. Specifications also include a linear year trend, to control for time-varying factors across all locations, such as macroeconomic conditions, as well as PUMA fixed effects, which control for time-invariant factors affecting the larger neighborhoods.

Data on retail employment densities and non-retail employment are calculated using the National Establishment Time Series (NETS) database for California, 1992-2009. This dataset contains the business name, geocoded address, NAICS industry code, number of employees, and firm type (i.e. single- or multiple-establishment firms). Using GIS, I identify all retail establishments within one-quarter mile area of sites where stations exist or will exist during the study period, and aggregate employment within station areas, as well as control tracts. Retail establishments are defined as those with NAICS two-digit code 44-45, as well as food service establishments (NAICS 722). Current analysis looks at the density of retail employment overall; future refinements will include analysis by retail segment (i.e. clothing, general merchandise, food and beverage) and by establishment size and by firm type, to test whether different types of retailers have different sensitivities towards transit. One hypothesis is that consumers using transit to conduct their shopping are more likely to purchase small, easy-to-carry items (such as food and clothing) than large, heavy items that would be better transported by car (electronic appliances, household items or furniture).

A key question of interest is whether the effects of new rail stations vary by prior demographic, economic or physical characteristics of the neighborhood. As Redfearn (2009) points out, estimating average effects across all transit stations may obscure substantial differences in localized impacts. Anecdotal evidence from Los Angeles suggests that development patterns around new transit lines have varied considerably: new residential and commercial activity has occurred around Metro's Gold Line in Pasadena and the Red Line stations in Hollywood, while very little development has occurred around the Blue Line stations in relatively lower-income neighborhoods of South LA. To test whether baseline characteristics influence the trajectory of retail around new stations, the basic estimates from Equation 1 will also be estimated separately for each of the four MSAs. Future robustness checks will stratify the analysis by initial neighborhood income, racial/ethnic composition, population density, and composition of surrounding housing stock.

Data on population characteristics are taken from the 1990 and 2000 census and the 2005-2009 American Community Survey. Because station areas are defined as the one-quarter mile circle surrounding stations, they do not align with census tract boundaries. GIS tools were used to identify which census tracts fall within the one-quarter mile treatment area for each station, and the amount of land each tract represents within the treatment area. Census variables for each tract are weighted by land shares to calculate average characteristics for each treatment area. Particular variables included are population density, share of population that is black and Hispanic, median household income, median value of owner-occupied housing, share of housing units built before 1940 and share of housing units in single-family structures. Variable definitions and sources are shown in Table 5, summary statistics are shown in Table 6.

Several geographic variables are also included in the analysis. Distance from each station and from the centroid of each control tract to the nearest Central Business District are calculated using latitude and longitude coordinates. The CBD is defined as the census tract with the highest total employment density, using the NETS data. One CBD is identified for each of the designated central cities within the MSA, using OMB definitions. Proximity of stations and control tracts to major highways are also calculated using GIS. Initial retail density and growth in retail are likely to vary relative to proximity to both the CBD and highways, also some rail stations are built along the freeway median, which is likely to impede nearby development. As discussed in the previous section, rail transit should be more attractive to riders when stations are part of a dense network, the analysis also includes a measure of station density, specifically the average distance of each station to its three nearest neighbor stations. Higher distances to nearby stations indicate lower network density.

One potential concern with identifying the effects of new rail stations on retail activity is that transit investments may occur simultaneously with changes in local land use or fiscal policy, designed to enhance the use of transit (for instance, an increase in allowable density of development or reduced tax rates). Unfortunately, collecting systematic data on those policies for all 254 neighborhoods in our sample that receive new train stations, and comparing them with similar policies for neighborhoods that do not receive rail stops, will not be feasible, so it will not be possible to include this in the regression models. A planned extension of the research is to conduct case studies of selected stations – including some stations that have seen large increases in employment as well as little employment growth – to examine the history of station development, including changes in land use or approvals process. In these case studies, I will interview officials from the relevant transit agencies and city planning departments, and developers who have built TOD projects, to ask them about changes in zoning or variances requested and received. I will also review public documents on zoning near station areas, before and after station development. And conducting larger-scale qualitative or quantitative studies on the relationship between transit investment and land use or fiscal policies is an important area for future research.

Section 4) Empirical results

Where are stations located?

The location of new stations with respect to other stations and the CBD varies across MSA and across systems w/in MSAs. Considering only stations built before 1992, San Francisco's transit network has the highest station density (the median station averages 0.74 miles to its nearest three stations), while LA's network is the least dense, at nearly 10 miles between stations. By the end of the study period, station density had increased by large amounts for all MSAs but Sacramento. Within the two largest MSAs, network density also varies across transit systems: average distance from BART stations to neighbors is nearly ten times that from MUNI stations (0.21 compared to 1.91), and Metrolink stations are five times farther from nearby stations than the MTA rail stops. Similarly, Metrolink is a more suburban system than MTA rail, with the median station located 11.5 miles from the CBD, compared with 3.6 miles for MTA stations. In general, suburban systems stations are less dense than transit systems near the center city (the correlation between distance to neighboring station and distance to CBD is 0.92).

Another illustration of the variation in station density is shown in Figures 1 and 2. Downtown Los Angeles has seven MTA rail stations, all within less than one mile of at least one other station. Indeed for several stations, there are overlaps in the quarter-mile radius of the station, which is expected to draw the most riders and should see the greatest development impact. By contrast, the two Metrolink stations in Ventura County are approximately four miles apart, with no overlap in potential catchment areas.

To examine more systematically the characteristics of neighborhoods around new rail stations, Table 7 presents descriptive characteristics of new station areas, compared to old station areas and the set of control census tracts with no stations. Columns 1-3 show the mean and standard deviation for several key characteristics, while Columns 4-6 present differences between the three neighborhood groups. In terms of distance from CBD, new stations fall between older station areas (which are close to the CBD) and control tracts. The areas where new stations were built were on average 10 miles from older stations, closer than the control tracts that did not receive stations. The population density of new station areas, as of 1990, was lower than that of older station areas, but compared to the population density of control tracts. However, new station areas had much higher initial (non-retail) employment density than control tracts, and were in fact comparable to older station areas in baseline employment density. A question is whether planners behind the location decisions believed that placing new stations in employment centers would maximize ridership, by serving an existing base of consumers, or whether it was easier to gain neighborhood acceptance for stations in heavily commercial areas (commercial landlords might be more supportive of stations than residential owners). Similarly, both old and new station areas had comparable median incomes, both significantly lower than control tracts. New station areas also had lower initial housing values than either old station areas or control tracts. The racial and ethnic composition of residents varies less across neighborhood groups: new station areas had somewhat more Hispanic residents than older station areas, but about the same as control tracts, and roughly the same share of black residents. In terms of housing stock composition, new station areas fall between older station areas (high percentage of pre-1940 housing) and control tracts (mostly new housing). New station areas had the lowest share of single family detached housing, with control tracts having the highest.

In general, the comparison of new station areas to older stations and control tracts suggests that station placement was not exogenous to initial physical, economic and demographic characteristics. New stations were placed in highly dense commercial areas outside the city center, with relatively low population densities and low property values – areas that should have the potential for additional commercial development.

<u>Does retail density increase after new stations open?</u>

To determine whether the opening of new stations is associated with a change in nearby retail employment, I first present differences in means of employment density, before and after stations open, then estimate difference-in-difference regressions.

Table 8 compares average retail employment densities across the three neighborhood groups, as well as pre- and post-station densities for new station areas. The first section of the table presents results for all metropolitan areas combined. Retail employment density is higher in areas around old and new stations than in control tracts: 523 retail employees per square mile in control tracts, compared to 2268 employees around older stations and 2439 retail employees in the quarter mile radius around new stations. Moreover, retail employment density in the areas about new stations appears to increase in years after the stations opened, rising from 2066 employees per square mile to 2668 employees. But this average effect for the full set of new station neighborhoods conceals large differences across MSAs, as shown in the lower four sections of Table 8.

In the Los Angeles and Sacramento MSAs, retail employment density within one-quarter mile of new stations actually decreases after these stations open. In Los Angeles, retail density around new stations both before and after opening is lower than retail density around older stations; this reflects the place of new Metrolink stations in suburban parts of Southern California, as well as some of the MTA rail stations in less dense areas. Prior to new stations opening, the average station area had 1902 retail employees per square mile, which drops by nearly 400 employees in the years after the stations open. The drop is statistically significant at the one percent level. In Sacramento, new station areas initially had higher retail employment density than older station areas (1859 employees compared to 1513), but after the stations open, retail employment drops by approximately 460 employees per square mile, although significant only at the 10 percent level. Retail employment density around new stations in San Diego increase by approximately 300 employees per square mile after stations open, although the change is not statistically significant. The only MSA in which new station openings are associated with a positive and significant change in retail employment is the San Francisco-San Jose metropolitan area. Retail employment around new stations is initially 2265, similar to that around older stations (2324), but increases to 4810 after station opening – more than doubling retail employment within a constant land area.

The difference in means tests summarized in Table 8 offer some insight into raw changes in retail employment around station areas, but do not prove that the station opening causes these changes. In particular, they do not address the counterfactual question of whether retail employment might have been increasing or decreasing in comparable areas during the same time period (there is no “post” opening period for control areas or older stations). To conduct a more robust test of the relationship between station opening and surrounding retail employment, I move to a regression framework, with results shown in Table 9. The results largely confirm the results of the difference in means analysis: across the combined sample, both old and new station areas have higher retail density than control tracts, and retail employment increases following new station opening. But regression results also confirm heterogeneous effects across MSAs, with positive effects of station opening only in San Francisco, while the other areas see no change or a decrease in retail following new station opening.

The first two columns in Table 9 show results of regressions including only the dummy variables for station presence ever in the study period and post-opening of new stations. Both regressions also include a linear time trend, to account for secular changes in retail employment over the period. Column 2 also includes PUMA fixed effects, so that coefficients should be interpreted as the effect of difference between treatment and control areas within the same PUMA. Coefficients in both columns indicate that station areas (old and new) have higher retail density than control tracts within the PUMA, and that retail employment in new station areas is higher after the station opens. Column 3 adds controls for distance to CBD, proximity to pre-1992 transit stations, and density of non-retail employment. Adding these controls causes the sign on station to flip, due to strong positive correlation between the station dummy and total employment density. The magnitude of the coefficient on the post-station dummy decreases but remains significant. Not surprisingly, retail employment is great in areas with higher total employment; workers in nearby industries form an important segment of retail consumers. Retail employment density decreases with distance from the CBD, consistent with monocentric cities models.[5] The coefficient on proximity to older transit stations is not statistically significant in Column 3. Column 4 adds controls for various population characteristics, causing the sign on station to flip back to positive, and not affecting the coefficient on post-station. Population density, which like non-retail employment density indicates the presence of potential consumers, is positively associated with retail density. Share of older housing stock is also positively associated with retail density, while median household income is negatively associated with retail density. Although higher income should imply greater purchasing power, higher income households are often opposed to nearby commercial activity; investigation of whether this could reflect zoning pressures is an area for future analysis.

As with the difference in means, the aggregate results conceal variation across MSAs. Columns 5-8 estimate the model from Column 4 separately for each of the four MSAs. In Los Angeles, station areas have higher retail density than control tracts within the same PUMA, and the coefficient on post-station opening for new stations is negatively but not statistically different from zero (and indeed the magnitude is quite close to zero). Station areas in Sacramento also have higher average retail density than control areas, while post-station opening is associated with lower retail density, controlling for other factors. The estimates for San Diego find no significant differences in retail density by station presence or post-station opening, relative to control tracts. Only in San Francisco is there a positive coefficient on the post-station dummy, while the coefficient on station presence is not statistically different from zero. The estimated coefficients on control variables are largely consistent across MSAs, suggesting that the fundamental drivers of retail activity are similar across cities.

[5] Because the CBD measures are calculated based on multiple central cities per MSA, this actually measures distance to the nearest employment sub-center, and does not assume monocentricity.

Section 5) Conclusions and recommendations

Over the past twenty years, local and regional governments in California have made substantial investments in new or expanded rail transit systems. One of the justifications for public funding of transit systems offered by these officials is that the areas near transit stations with benefit from greater economic development, increases in jobs, property values, and other amenities. In this paper, I describe the locations selected for new rail stations in California's four largest MSAs and examine whether the opening of new stations has been associated with a growth in retail activity immediately around the stations.

Results indicate that new rail stations were located in areas with initially high employment density, somewhat outside the city centers, with relatively low household incomes and property values. These areas therefore offer potential for physical and economic development surrounding the stations. However, there is considerable variation in the baseline characteristics of neighborhoods selected for retail, both within and across MSAs, that may affect the prospects for nearby development. In addition, the new stations vary in their physical characteristics, intra-MSA location, and the type of rail networks to which they belong. Stations that attract higher ridership will presumably be more attractive targets for retail or residential development, because of their ability to draw in consumers. Analysis of retail employment suggests that the impacts of new stations on retail activity near the stations has been highly varied: while new station openings are associated with increased retail employment in the San Francisco MSA, retail employment decreased around new stations in Sacramento, and did not significantly change in Los Angeles or San Diego.

These results naturally raise the question: what makes San Francisco different? In particular, are there features of the urban structure, the transit system, or the population that might make access to transit a more valuable amenity in the San Francisco metro area than in the other regions studied? While the analysis is to preliminary to draw firm conclusions, several differences between San Francisco and other California cities stand out. As noted in Section 2, the transit systems in San Francisco city – especially MUNI – are much older than in other parts of the state. Much of the current built environment was developed around the trolley lines that still form the basis for MUNI's routes. Employment is highly concentrated in the central business district and nearby downtown, compared to more polycentric employment patterns in Los Angeles city, for instance. A number of Bay Area municipalities have been at the forefront of land use policies that try to curb new development in outlying suburbs, resulting in a higher share of residents in older, centrally located neighborhoods (see Gruen 2010, chapters 6 and 7). Perhaps due to the combination of centralized employment and old, dense transit networks, transit ridership in San Francisco city is much higher than in any other California county. Prior research has suggested that when new rail lines are built, riders are drawn not from vehicle commuters, but from former bus riders – a shift within transit rather than a net increase. Although at the moment these are just hypotheses to be tested, all of these factors may contribute to make rail stations a more valuable amenity in San Francisco, leading to greater employment growth in the vicinity of new stations.

The next stage of research will be to examine in more detail why stations have created such heterogeneous impacts across MSAs. Estimating regressions separately by MSAs is a fairly crude proxy for variation in the underlying neighborhood or station characteristics that should impact the neighborhood's ability to attract retail. Thus future analysis will stratify the sample by characteristics such as intra-MSA location, neighborhood income & racial composition, as well as by density of built environment. Of particular interest is whether initially low-income and minority neighborhoods have benefitted from new rail stations, or whether the benefits accrue mostly to higher-income areas. Additionally, looking at the impacts separately by rail systems within MSAs could create some insight into whether high-density networks serving central cities, like MTA rail and MUNI, have different impacts on nearby retail than suburban commuter rail systems, like Metrolink and Caltrain.

Another area for further exploration is whether the effects of rail investment vary over time, because of the time needed for retail establishments to obtain necessary permits for construction/renovation and business operation, and if there is an adjustment period before residents switch transit modes to greater use of rail. I will test for variation in effects with age of station by using continuous time variables measuring time before/after station openings instead of dummy variable for post-station opening. I will also examine trends in retail activity before station opening, as a means of testing for speculation or capitalization of effects before the station became operational.

References

Alwitt, L. and T. Donley. 1997. Retail Stores in Poor Neighborhoods. Journal of Consumer Affairs 31: 139-164.

Bartie, P. and T. Blakely, J. Pearce, K. Witten. 2007. "Neighborhood Deprivation and Access to Fast Food Retailing: A National Study." American Journal of Preventive Medicine 32(5): 375-82.

Baum-Snow, Nathaniel and Matthew Kahn. 2005. Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000. Brookings-Wharton Papers on Urban Affairs.

Bollinger, Christopher R. and Keith Ihlanfeldt. 1997. The Impact of Rapid Rail Transit on Economic Development: The Case of Atlanta's MARTA. Journal of Urban Economics 42: 179-204.

Brooks, Leah and Byron Lutz. 2013. Vestiges of Transit: Urban Persistence at a Micro Scale. Working paper.

Carr, James and Jenny Schuetz. 2001. Financial Services in Distressed Communities: Framing the Issue, Finding Solutions. Washington, DC: Fannie Mae Foundation.

Ewing, Reid and Robert Cervero. 2010. Travel and the Built Environment. Journal of the American Planning Association 76(3): 265-294.

Giuliano, Genevieve and Ajay Agarwal. 2010. Public Transit as a Metropolitan Growth and Development Strategy. Brookings Institution working paper.

Gruen, Claude. 2010. New Urban Development: Looking back to see forward. Piscataway NJ: Rutgers University Press.

Hayes, Lashawn R. 2000. "Do the Poor Pay More? An Empirical Investigation of Price Dispersion in Food Retailing." Princeton Department of Economics, Industrial Relations Working Paper No. 446.

Hotelling, H., 1929. Stability in competition. Economic Journal 39: 41–57.

Kahn, Matthew. 2007. Gentrification Trends in New Transit Oriented Communities: Evidence from Fourteen Cities that Expanded and Built Rail Transit Systems, Real Estate Economics 35 (2): 155-182.

Kolko, Jed. 2011. Making the Most of Transit: Density, Employment Growth, and Ridership around New Stations. Public Policy Institute of California working paper.

Redfearn, Christian. 2009. How Informative are Average Effects? Hedonic Regression and Amenity Capitalization in Complex Urban Housing Markets. Regional Science and Urban Economics 61:519-41.

Schuetz, Jenny, Jed Kolko and Rachel Meltzer. 2012. Are Poor Neighborhoods Retail Deserts? Regional Science and Urban Economics.

Sloane, D.C., L.B. Lewis, L.M. Nascimento. 2005. Assessing Healthy Food Options in South Los Angeles Restaurants. American Journal of Public Health 95(4): 668-673.

Voith, Richard. 2005. Comment on Baum-Snow and Kahn. Brookings-Wharton Papers on Urban Affairs.

Waldfogel, Joel. 2008. The median voter and the median consumer: Local private goods and population composition. Journal of Urban Economics 63: 567-582.

Winston, Clifford and Vikram Maheshri. 2007. On the Social Desirability of Urban Rail Transit Systems. Journal of Urban Economics 62: 362-382.

Figure 1: Treatment areas around transit stations, Downtown Los Angeles

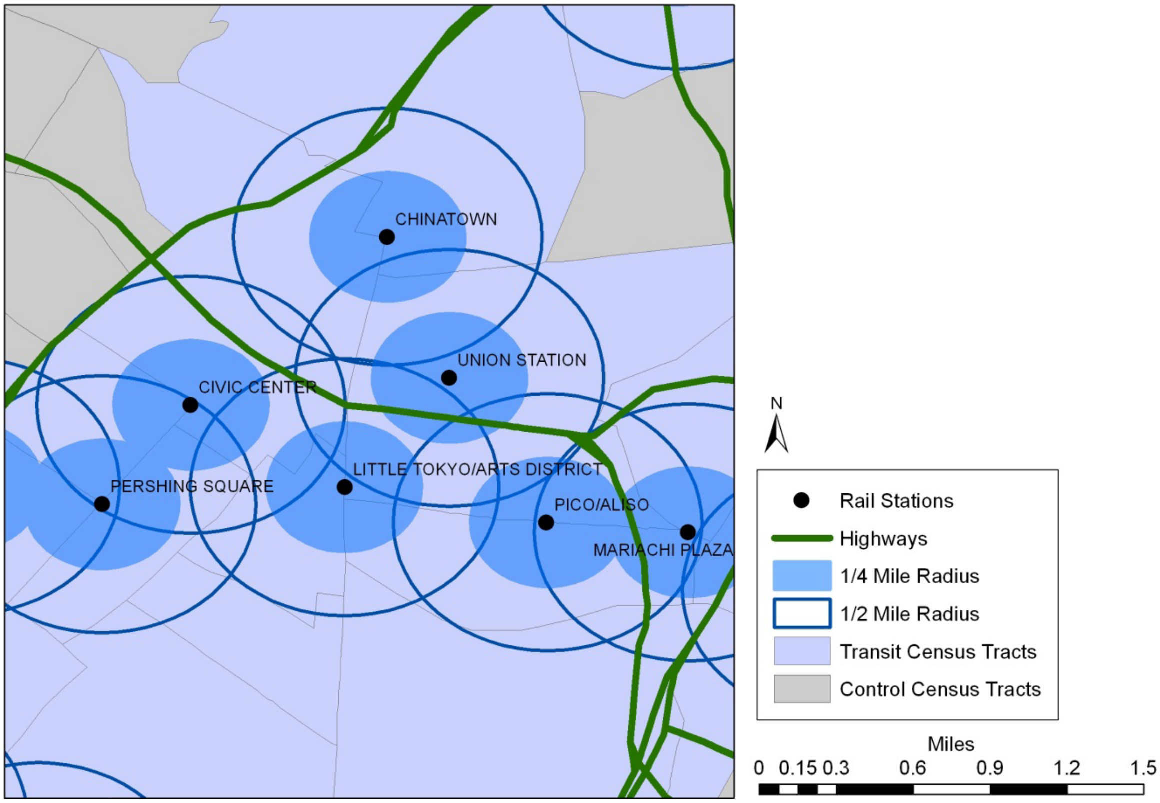

Figure 2: Treatment areas around transit stations, Ventura County

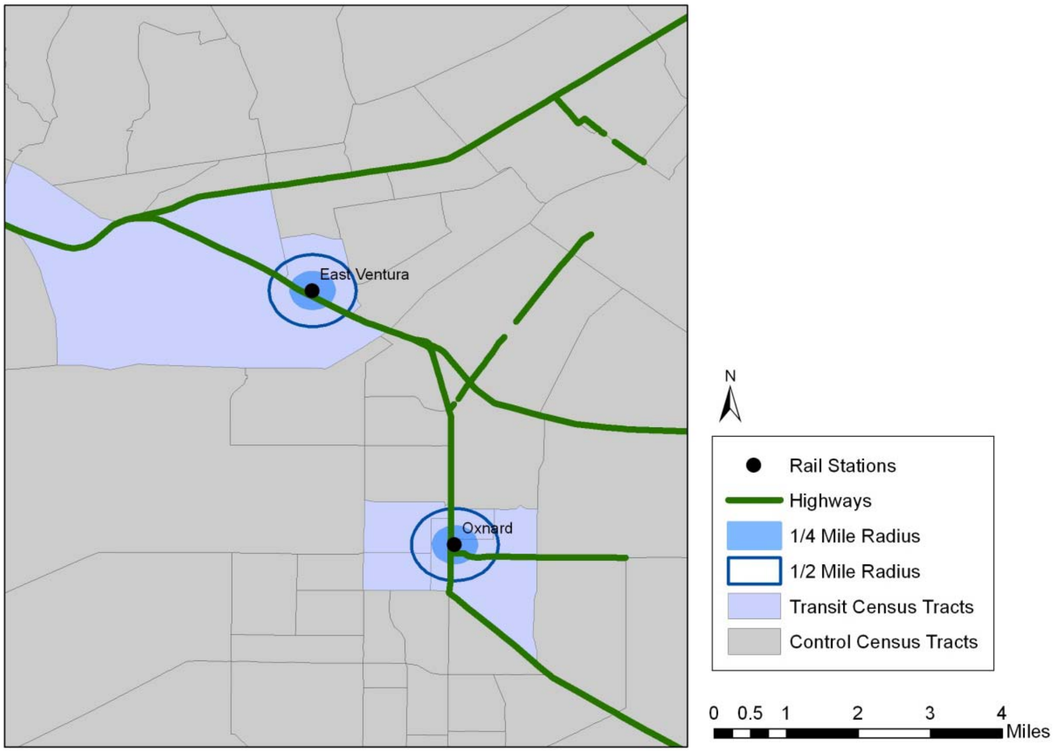

Table 1: Fixed-line transit systems in California

Metro area	System	Stations < 1992	New stations, 1992-2009	Total
Los Angeles	LA Metro Rail	22	48	70
	LA Metro BRT	0	13	14
	Metrolink*	0	55	56
Sacramento	Sacramento Light Rail	30	18	48
San Diego	Coaster San Diego	1	5	8
	San Diego Trolley	34	18	52
San Francisco/San Jose	BART	35	9	44
	MUNI fixed lines	96	48	144
	Caltrain	25	7	33
	San Jose Light Rail	31	30	62
	Altamount Commuter Express (ACE)	5	3	8
All metros	Amtrak	7	na	31
Total		266	254	520

Source: Data assembled from transit operator websites. Columns for total station and stations before 1992 double-count stations served by multiple lines. Column for new stations, 1992-2009, has no double-counted stations.

Table 2: Stations served by multiple lines or systems

Number of lines	Old stations	New stations	All stations
1	65.4%	87.0%	76.0%
2	22.9%	11.0%	17.1%
3	6.0%	1.2%	3.7%
4	3.0%	0.4%	1.7%
5+	2.6%	0.4%	1.5%

Table 3: Reuse of existing stations by new transit lines

Station built	# stations
< 1900	5
1900-1920	7
1921-1940	10
1941-1960	7
1961-1980	6
1981-1990	9
Total	44

Table 4: Median station location, by MSA and rail system

	Distance, pre-92 stations	Distance, all stations	Distance CBD
SF/SJ	0.74	0.36	3.98
MUNI	0.25	0.21	4.20
BART	2.16	1.91	5.47
CalTrain	1.79	1.67	4.35
SAC	0.91	0.87	3.51
SD	1.78	1.11	4.46
LA	9.54	1.73	7.57
MTA rail	2.74	1.09	3.61
Metrolink	28.64	5.26	11.50
All stations	1.44	0.87	4.45

Distance to other stations is average distance to three nearest stations. All distances reported in miles.

Table 5: Variable definitions and sources

Variable	Definition	Source
Station characteristics		Assembled by author
stabuilt	Year station originally built	
yropen	Year current service began	
numlines	# rail lines/systems operating at station	
oldnear3	Average distance to 3 nearest stations, 1990	
near3	Average distance to 3 nearest stations ever	
Geographic characteristics		
landsqmi	Land area (sq mi)	Census
distcbd	distance (miles) to CBD	Author calculations
centcity	= 1 if central city, = 0 otherwise	OMB 2000 MSA defns
disthwy	distance (miles) to nearest class 1 highway	GIS calculations
Employment		NETS 1992-2009
est	# retail establishments	
emp	# retail employees	
totemp	# employees, all industries	
est5	Stations: retail est w/in 1/2 mile	
emp5	Stations: retail emp w/in 1/2 mile	
totemp5	Stations: emp (all ind) w/in 1/2 mile	
Population characteristics		Census 1990, 2000, ACS 2005-2009
popdens	Pop/sq mi	
white	% white	
black	% black	
hisp	% Hispanic	
ownocc	% owner-occupied housing	
hsgpre40	% housing built < 1940	
sf	% single-family housing	
medhhinc	Median HH income (2000 $)	
medvalue	Median housing value (2000 $)	

Table 6: Variable summary statistics

	Mean	Std. Dev.	Min	Max	Obs
Station characteristics					
yropen	1990.87	9.49	1972.00	2009.00	520
near3	1.57	2.12	0.10	16.22	520
oldnear3	5.40	9.96	0.10	55.72	520
distcbd	5.87	5.45	0.06	28.29	520
numlines	1.38	0.90	1.00	7.00	520
Employment & location characteristisc					
empdens	797.63	2181.34	0.00	69238.77	62,370
estdens	98.81	193.50	0.00	5393.44	62,370
distcbd	8.78	6.07	0.06	62.00	62,370
oldnear3	13.10	14.31	0.10	62.25	62,370
totempd	5681.75	24559.64	0.00	740281.80	62,370
Population characteristics					
popdens	9,059	7,767	0	61,088	10,395
medhhinc	72,374	33,348	0	261,852	10,385
medvalue	438,405	231,896	0	1,295,561	10,359
medrent	1,222	404	0	2,592	10,358
white	61.46	22.47	0.00	100.00	10,345
black	8.02	13.04	0.00	97.64	10,345
hisp	29.07	25.17	0.00	100.00	10,345
forborn	27.16	15.10	0.00	88.18	10,345
hsgpre40	10.46	15.98	0.00	100.00	7,394
ownocc	57.39	25.03	0.00	100.00	10,330
sf	64.39	26.86	0.00	100.00	10,330

Table 7: Neighborhood characteristics of new station areas (1992/1990)

	(1)	(2)	(3)	(4)	(5)	(6)
	New station	Old station	None	New - old	New - None	Old - none
distcbd	7.30	4.50	9.30	2.80 ***	-2.00 ***	-4.80 ***
	(6.76)	(3.26)	(6.03)			
oldnear3	10.02	0.99	14.46	9.03 ***	-4.43 ***	-13.47 ***
	(12.67)	(1.09)	(14.53)			
popdens	8,513	11,226	8,266	-2,713 ***	247	2,960 ***
	(8,484)	(7,568)	(7,377)			
totempd	22,891	23,114	2,552	-223	20,338 ***	20,562 ***
	(65,388)	(62,462)	(4,184)			
medhhinc	59,834	59,576	74,668	258	-14,834 ***	-15,091 ***
	(20,879)	(26,626)	(33,086)			
medvalue	353,379	391,591	404,210	-38,212 ***	-50,831 ***	-12,619
	(158,198)	(190,141)	(199,532)			
black	8.95	10.46	8.35	-1.51	0.60	2.11 **
	(12.35)	(12.47)	(15.31)			
hisp	26.17	20.64	24.21	5.53 ***	1.96	-3.57 ***
	(22.44)	(18.51)	(22.76)			
hsgpre40	16.27	30.72	9.45	-14.45 ***	6.82 ***	21.27 ***
	(18.00)	(21.71)	(15.07)			
sf	43.60	48.97	66.21	-5.37 **	-22.61 ***	-17.24 ***
	(26.12)	(27.92)	(25.18)			
n =	254	266	2945			

Census variables are reported for 1990, employment density for 1992. Distance to CBD and distance to pre-1992 stations are constant over time. Station areas for new and old stations are 0.25 radius around station. Control areas are non-station census tracts in PUMAs with at least one station. Standard errors shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Retail employment density, by presence of station (1992-2009)

All metropolitan areas

	No stations	Old stations	New stations			
			All years	Before opening	After opening	Post - pre
Mean	523	2,268	2,439	2,066	2,668	602 ***
St dev	(872)	(4,309)	(5,552)	(3,798)	(6,383)	
n =	53,010	4,788	4572	1,738	2,834	

Los Angeles

Mean	576	2643	1623	1902	1510	-391 ***
St dev	(919)	(2837)	(2279)	(2295)	(2264)	
n =	33,210	396	2,070	593	1,477	

Sacramento

Mean	276	1513	1701	1859	1398	-461 *
St dev	(403)	(2555)	(2443)	(2669)	(1911)	
n =	3186	540	324	213	111	

San Diego

Mean	443	2400	2128	1896	2215	318
St dev	(699)	(4509)	(2213)	(1973)	(2293)	
n =	5814	612	432	117	315	

San Francisco/San Jose

Mean	477	2324	3622	2265	4810	2595 ***
St dev	(892)	(4627)	(8363)	(4946)	(10334)	
n =	10800	3240	1746	815	931	

Retail employment density is calculated as employees per square mile, either for the one-quarter mile radius around a station or for control census tracts. Standard errors shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Does retail employment growth increase after station opening?

Dep var:	ln(Retail emp/sq mi)				LA	SAC	SD	SF-SJ
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
station	1.236***	0.812***	-0.123***	0.0976**	0.239***	0.431***	0.268	-0.061
	(0.050)	(0.049)	(0.043)	(0.042)	(0.047)	(0.146)	(0.169)	(0.071)
post_trt	0.149***	0.381***	0.118***	0.104**	-0.037	-0.874***	-0.163	0.470***
	(0.054)	(0.052)	(0.046)	(0.044)	(0.056)	(0.155)	(0.174)	(0.071)
distcbd			-0.0137***	0.00725***	0.00911***	-0.0460***	0.0325***	-0.005
			(0.002)	(0.002)	(0.002)	(0.014)	(0.009)	(0.006)
ltotempd			0.845***	0.721***	0.736***	0.711***	0.765***	0.668***
			(0.004)	(0.005)	(0.007)	(0.020)	(0.015)	(0.011)
oldnear3			-0.003	0.0115***	0.00517**	0.0549***	-0.007	0.0771***
			(0.002)	(0.002)	(0.002)	(0.013)	(0.008)	(0.006)
lpopd				0.362***	0.334***	0.182***	0.348***	0.468***
				(0.006)	(0.008)	(0.022)	(0.020)	(0.013)
loldhsg				0.0373***	0.0597***	-0.149***	0.131***	0.003
				(0.005)	(0.007)	(0.024)	(0.018)	(0.008)
linc				-0.133***	-0.0913***	-0.404***	-0.361***	-0.232***
				(0.016)	(0.017)	(0.088)	(0.067)	(0.038)
year	0.0160***	0.0152***	-0.00396***	-0.00601***	-0.00231**	-0.00732**	-0.00439*	-0.0146***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.002)
PUMA Fes	N	Y	Y	Y	Y	Y	Y	Y
Observatior	62,370	62,370	62,370	61,954	35,412	3,986	6,786	15,770
R-squared	0.077	0.336	0.624	0.651	0.68	0.62	0.589	0.638

Robust standard errors shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1