Reconsidering the Impact of Access to Transit on Local Land Markets

Final Report

METRANS Project 11-08

September 2015

Principal Investigator:
Christian L. Redfearn

Sol Price School of Public Policy
University of Southern California
Los Angeles, CA 90089-0626

Disclaimer

The contents of this report reflect the views of the author, who is responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, the California Department of Transportation and the METRANS Transportation Center in the interest of information exchange. The U.S. Government, the California Department of Transportation, and the University of Southern California assume no liability for the contents or use thereof. The contents do not necessarily reflect the official views or policies of the State of California, USC, or the Department of Transportation. This report does not constitute a standard, specification, or regulation.

Abstract

We exploit the polycentric nature of the Los Angeles Metropolitan area to learn about the impact of new passenger rail stations on land use in the surrounding areas. By using the many centers in the Los Angeles MSA, we are better able to control for variation in trend growth in population and employment density. The parallel trend assumption required of differences-in-differences approach appears to fail under commonly-used controls. Making use of the centers as units of analysis reveals significant growth in both employment and population density around new stations. These results are useful for policy makers interested in assessing the indirect benefits of investment in new stations. The results are also informative for those using the DiD approach in urban settings. While the effects of new stations are significant and positive, there is marked heterogeneity across stations – suggesting that more research is needed to understand the link between new stations and subsequent changes in land use.

Table of Contents

List of Figures

List of Tables

Disclosure

The project was funded in entirety under this contract to California Department of Transportation.

Acknowledgements

This paper has benefited from the comments of Genevieve Giuliano and the research support from METRANS. The paper has also benefited from Yuting Hou who has provided research assistance. Thanks also to the participants at the UCI-UCLA-USC Joint Urban Economics Symposium, September 27^{th}, 2013 and, in particularly, to Matthew Kahn for his discussant comments.

1. Introduction

Differences-in-differences (DiD) is a commonly-used approach to assessing program evaluation in a wide variety of settings. It is the backbone of research in medical sciences, but also has wide use in the social sciences (see for example Card and Krueger (1994)). Regardless of the application, a standard set of assumptions are required to recover the true underlying population parameters. A central assumption of DiD is that trends in the outcome variable in the control and treatment groups are common. If not – if underlying trends differ between the treatment and control groups – then measured impact of the treatment may be biased. In the case of urban spatial data, there are several reasons to question the usual assumptions, but in particular this parallel trend assumption. In this paper, we consider the use of DiD in the case of new transit rail stations in Los Angeles, California. An important policy question facing policy makers is how these new stations have induced changes in land use around them. DiD is a natural tool that could be used to assess these changes. At issue is the whether or not the necessary assumptions hold such that the results can be interpreted correctly.

The motivation for studying changes in passenger rail access it twofold. The first is driven by the need to assess the impact of investment in new stations and rail. Over the last the last several decades, large investments have been made in passenger rail. Baum-Snow, Kahn, and Voith (2005) reports that "25 billion dollars were spent between 1970 and 2000 in 14 major cities in the United States." In the years since, new lines and stations have been added to the passenger rail systems in New York City, Washington D.C., and Los Angeles, among others. Though ridership is viewed as one metric of direct benefits from this investment, indirect benefits may accrue in the areas around stations as these locations become more accessible to the rest of the network and the rest of the metropolitan area.

Accessibility and commuting costs are central influences on firm and household location choices. Classic urban economics focuses broadly on the trade-off between commuting costs and land consumption that results in highest land use density around the central business district (CBD) (Alonso 1964, Mills 1967, Muth 1969). The same logic can apply for other locations within the metropolitan area: if the change in access around the stations results in more demand and, under usual assumptions, prices for surrounding land change. Under the classic model of a "featureless planes," this results in (re)development and higher land-use density around the new stations. In practice, regulation can restrain development. It is an open empirical question as to what changes around new passenger rail stations. Moreover, it is a significant challenge to be able to attribute changes in population and/or employment density around stations to the independent effect of a new station on land use.

The goal of this research is to do just this: to assess the impact of new passenger rail stations on changes in land use intensity. We do this using several sources of data that focus on passenger rail Los Angeles County, California.[1] We use the National Employment Time Series (NETS) data that tracts establishments at specific, geocoded, locations. We use four cross sections of the NETS data (1995, 2000, 2005, and 2009) to identify the independent effect of new stations. We also use U.S. Census data on population and the Southern California of Associated Governments (SCAG) data on employment as well. We then use the location and opening dates of the Los Angeles Metropolitan Transportation Authority (MTA) Metro passenger rail stations.

We use one additional source of data that will becomes useful as we proceed through the DiD analysis. We make use of the employment centers established by Redfearn (2007). These 41 centers are the "poly" in the polycentricity that Los Angeles exhibits. Clearly, a CBD exists, but more than 90 percent of the MSA employment resides outside the traditional CBD in downtown Los Angeles. But rather than be scattered and dispersed, employment outside the CBD is concentrated in nodes. 3.2 million jobs – about half of all jobs in the metropolitan area are in employment centers. These local agglomerations are found across the Los Angeles MSA (Giuliano and Small 1991, Giuliano and Redfearn 2007) and many other MSAs (Bogart and Ferry 1999, McMillen 2001, Craig and Ng 2001, McMillen and Smith 2003, Lee 2007). These employment centers act as local points of employment gravity and, in the case of Los Angeles, appear to exhibit differential growth in land use intensity. There is variation among centers and systematically different growth between areas in centers and outside centers. These centers are important to acknowledge formally and incorporate in the DiD analysis. A "standard" approach to DiD that includes only distance to the CBD as a control for the broad secular decline in em-

[1] In this preliminary report, we focus on just the Metro stations of Los Angeles County. We also have access to the NETS data for Sacramento, San Diego San Francisco, and San Jose metropolitan areas. Our analyses will be extended to this MSAs in future research.

ployment density as distance increases fails to control for local density. By better controlling for the underlying trends in population and employment density, estimated effects of stations are more plausible. We find that new stations add significantly to employment growth. Population growth around stations also looks to be positively correlated with new passenger rail stations. But, these results are preliminary at this time.

In the rest of the paper, we first briefly discuss the theoretical basis for monocentricity and polycentrity in the context of passenger rail in Section 2. In this section, we also discuss how DiD is used in these settings to identify the independent effect of new passenger stations on land use. In Section 3, we cover the various data sources and illuminate the ways in which the data are utilized. We cover preliminary results in Section 4. We discuss what the results imply with regard to land use and new passenger rail stations. In this section, we also extend the results to shed light on the importance of the – oft imposed, but not tested – assumptions that undergird use of DiD in practice. This research is underway and we also discuss possible avenues for improving the results.

2 Monocentricity, Polycentricity, & The Access to Passenger Rail

The motivation for studying passenger rail stations should be clear: in a congested urban area, a second mode with an alternative right of way could be a valuable investment. New York City's wealth and density that depend on a robust subway system are prima facie evidence of benefits, but New York is exceptional. Indeed, the literature on rail access is large but hardly conclusive. That is, though New York City's rail system may be viewed as essential to its urban form, other systems – like Washington, D.C.'s Metro and the Bay Area's Rapid Transit system (BART) – have their critics about their continuing need for subsidies. We do not address this debate, but instead to look at an important dynamic within the larger conversation. To be specific, one claimed benefit of investment in passenger rail is the induced change in demand for locations around new stations.[2] Indeed, "value capture" – the ability to tax some of the new value created from new stations – is viewed as part of the way in which the rail investments are to be recovered

[2]To be clear, the economics of passenger rail stations, like any network, are nonlinear. The value of an additional station very much depends on the value of the rest of the next. Likewise, the entire network becomes more valuable when a new node is added. The focus of this paper is the direct impacts of changing land-use density around new stations, not the indirect network effects.

(Suzuki, Cervero, and Iuchi 2013). In order for “value creation” to work, land prices must rise. And, if significant value is to be created, typically land use changes must follow.

In previous work, Redfearn (2009a) found little significant evidence that house prices systematically rose in the time period following the opening of a new passenger rail station along two lines of the Los Angeles Metro system. However, he looked only at proximal single-family residential homes. It entirely possible that the stations most likely benefit from the new stations are those in existing employment centers around which there are few single-family houses. Indeed, those stations around which single-family residences were a significant portion of the land-use, may be those neighborhoods that make densification difficult via regulation and entitlement challenges.

In this paper, we focus on employment rather than houses and house prices. At issue in the case is how the independent effects of stations may be identified. The first obvious issues is sample selection with regard to the choice of station location. It may be possible that stations might be located in areas ripe for employment growth. If so, measured effects might overstate a general effect of new stations. In the case of Los Angeles, much of this is mitigated due to the nature of the rights-of-way. Most of the rail lines are above ground, following long existing rights-of-way. Moreover, the inherent nature of the network and the spacing of rail stations implies that station location choices are constrained.

The constrained nature of rail stations is especially true with regard to politics. Though all development in urban Los Angeles is now viewed through a political lens, Altshuler and Luberoff (2003) goes to great length to illustrate that a narrow view of economic optimization guides station location choice in Los Angeles. Kahn (2007) formalizes station location choice in examining patterns of gentrification around passenger rail stations. He finds mixed results across 14 different metropolitan passenger rail systems, but find no evidence of impact on residential gentrification in Los Angeles. He cites Altshuler and Luberoff (2003) to help explain how politics may trump selection in a way that could lead to bias: “...localities frequently minimized controversy by siting new transit lines in existing rail or freeway corridors, chosen for their availability rather than their optimality from a patronage standpoint. This strategy was used, for example, in Portland and Sacramento and for portions of the Los Angeles’ new Blue Line Altshuler and

Luberoff (2003)." For these reason, we don't feel that sample selection from cherry-picking the best potential passenger rail station locations is a primary challenge.

Rather, the primary empirical challenge is the control for heterogeneity in those factors that contribute to population and employment density – the outcome variables of interest. At issue here is that employment and population growth may not be distributed randomly over space. If not, and if stations are located in places that grow at faster rates than areas do not receive new stations, then the underlying factors that drive the faster growth may be incorrectly attributed to the introduction of new stations. This sort of empirical challenge is common in the social sciences, in which average treatment groups are different from average control groups. One approach for handling this type of empirical set up is "double-differencing" or "differences-in-differences" (DiD).

Card and Krueger (1994) and the responses that followed that research produced a large number of papers using differences-in-differences. Of course, several key papers followed as well that were critical of the generic use of DiD to recover particular parameters and interpret them directly as evidence of a policy intervention (Abadie 2005, Bertrand, Duflo, and Mullainathan 2004, Donald and Lang 2007, Imbens 2004)). A central problem facing empiricist who use a DiD approach is the underlying assumption of parallel trends. Equation (1) highlights the challenge:

$$(1) \quad Y_i = \alpha + \beta T_i + \gamma t_i + \delta(T_i \cdot t_i) + \varepsilon_i$$

In Equation (1), Y_i is the outcome variable – in our case changes in employment and population density; α is a constant term; βT_i is the treatment group fixed effect (the average effect of all stations on changes in density); γt_i is the time trend, common to control and treatment groups; and $\delta(T_i \cdot t_i)$ is the true effect of treatment. Holding fixed the locations of any station at any point in time, interactions between the date and the station fixed effects measure the independent impact of new stations. The parallel trend assumption makes clear if $cov(\varepsilon_i, T_i \cdot t_i) = E(\varepsilon_i(T_i \cdot t_i)) = \Delta$, then $E[\hat{\delta}_{DD}] = (\gamma^T + \gamma) - \gamma^C = \gamma + \Delta$ and the measured interaction γ in Equation (1) is biased by Δ.

A good example of the differences-in-differences approach is Kahn (2007); Kahn with Baum-Snow (2000, 2005) provide two other examples of rigorous analysis of new passenger rail impacts.

We follow these to arrive at our basic model:

$$\text{(2)} \quad density_i^{e,p} = d_i^{year} + d_i^{city} + d_i^{station} + d_i^{newStation} + \beta X_i + \varepsilon$$

This represents a "standard" approach to identifying the independent effect of new stations on surrounding density. In this case, $density^e$ reflects employment density; the other dummy variables capture the average effects within years, within cities, and within close proximity to a passenger rail station. ("Close" will be defined in the results section below. Various specifications are examined.) The other terms represented in the βX reflect other covariates, such as demographics and distance to the CBD.

The time dummies in Equation (2) are obvious and meant to reflect the fact that larger economic forces are work in the metropolitan area that should not be attributed to new stations. The station dummies capture the average treatment effect – before and after a new station is established. These are the notional treatment effects, that these areas may be persistently different than other areas. The city dummy here represents the hypothesis that cities within the metropolitan area vary in their growth rates and, again, should not be attributed to new stations. It is here that we exploit the novel use of employment centers as a potential control for other variation across geographies that remain approximately fixed over time. Cities are smaller units than the metropolitan area, but hardly small in many cases. Los Angeles County contains 88 incorporated municipalities. Los Angeles City has 3.7 million residents; Long Beach, the second city in L.A. County has 462,000 residents. Thirteen municipalities have populations between 100,000 and 200,000. The remaining 73 municipalities have an average population of 37,000. The economic dynamics within the large cities may be quite heterogeneous.

An novel alternative set of controls for variation in the underlying fundamentals of growth are the employment centers of Redfearn (2007). These 41 center are defined as contiguous clusters of Census tracts that collectively are significantly different than the areas that surround them.[3] The same logic applies for the location fixed effects, but in this case provide a much tighter geography

[3]The basic search algorithm in this paper begins with local maxima. Then tracts contiguous to the maxima are added iteratively. The optimal collection of tracts are those that provide the best fit of those in and those outside a center boundary.

to capture employment concentration. An alternative specification is then:

$$(3) \quad density_i^{e,p} = d_i^{year} + d_i^{center} + d_i^{station} + d_i^{newStation} + d^{center} \cdot d^{station} + \beta X_i + \varepsilon$$

A final specification addresses the generic βX in Equations (1) and (3). These variables are meant to capture all the other variables that influence employment densities. This is obviously a long list that is generally not readily available. Standard proxies for these include demographic data like income, race, ethnicity, human capital, and other variables. These variables are only moderately successful at predicting economic activity. As such, we take one additional set of models to make use of the persistence of employment and population over time (Redfearn 2009b). This specification embeds the covariates in βX above by using lagged employment density. That is,

$$(4) \quad density_i^{e,p} = \alpha + density_{i,-1}^{e,p} + d_i^{center} + d_i^{newStation} + d^{center} \cdot d^{newStation} + \beta X_i + \varepsilon$$

In this way, all those factors thought to be relevant to employment density – regulation, in particular – are controlled for via the lagged dependent variable. Systematic changes in employment density over time arise then from new stations and changes in particular subcenters and the areas around them.

3 Data

In order to explain the various data sources we use in these analyses, it may be most fruitful to look at several different, but related maps. The first is to look at the larger Los Angeles metropolitan area to place the study area in better context. Figure 1 shows the solid area all locations within four miles of a passenger rail station on the MTA Metro system. The Pacific Ocean is in the lower Southeast third of the map. The lines are major highways. The map includes all of Los Angeles and Orange Counties, but only the denser portions of Riverside, San Bernardino, and Ventura Counties. Figure 2 shows the location of the L.A. Metro rail lines and stations within the study area. The stations that are “X’ed” out are those stations that represent new stations during our sample period.

Figure 1. Working Data for Station Access & Employment

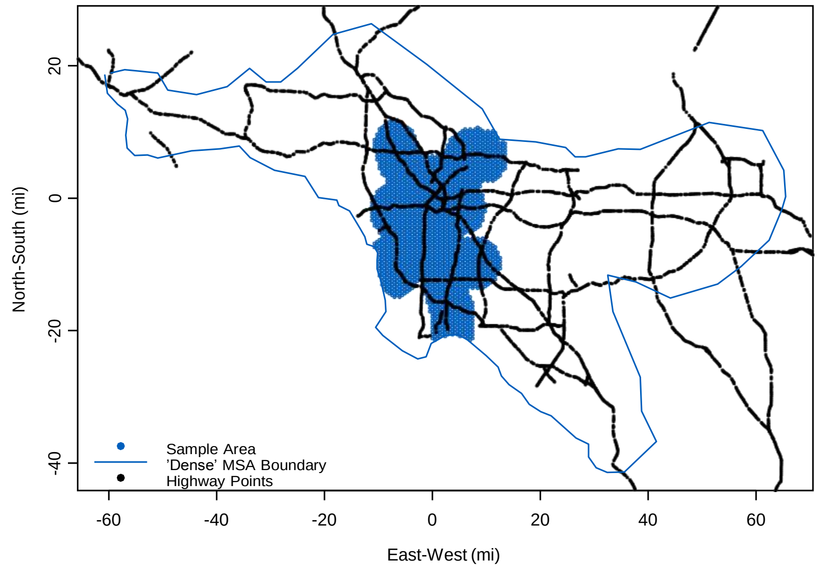
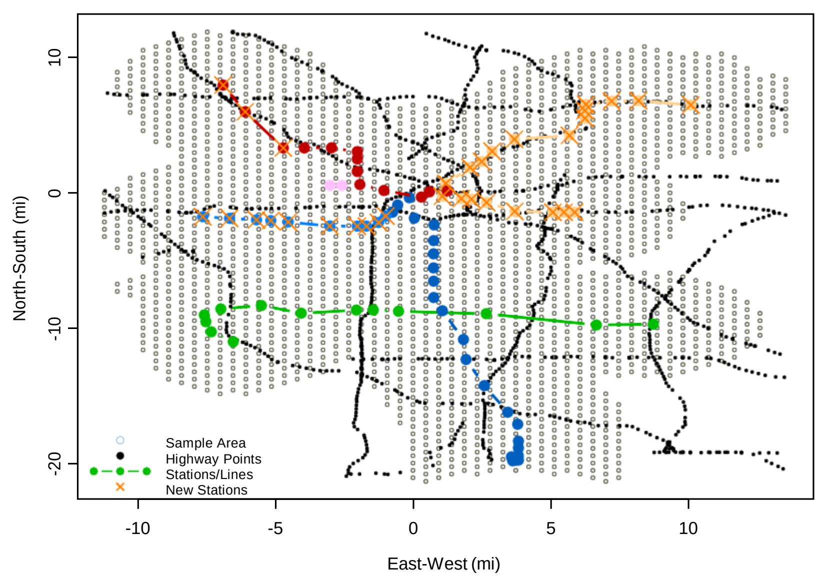


Figure 2. Working Data for Station Access & Employment



The novel use of employment centers as controls may be motivated by looking at them geographically. Figure 3 shows the same study area map in Figure 1, but now overlays the center boundaries defined by the contiguous Census tracts that comprise the centers (Redfearn 2007). In a region as large Los Angeles, there are many significant nodes of dense employment – the

Figure 3. Working Data for Station Access & Employment

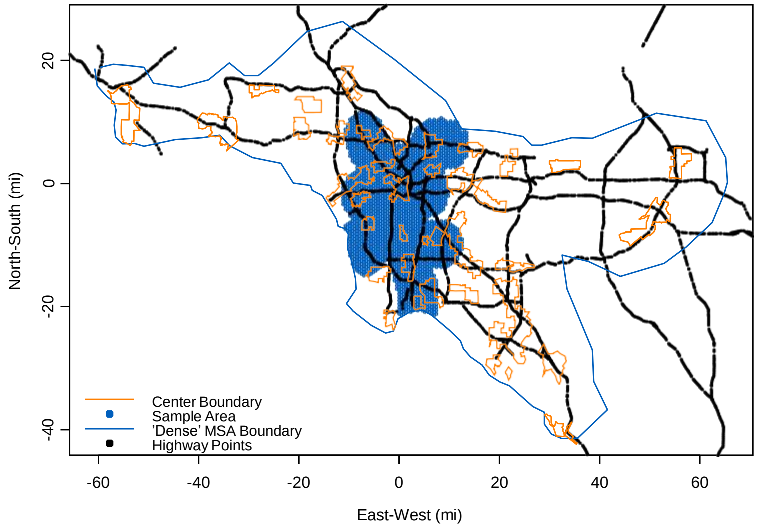
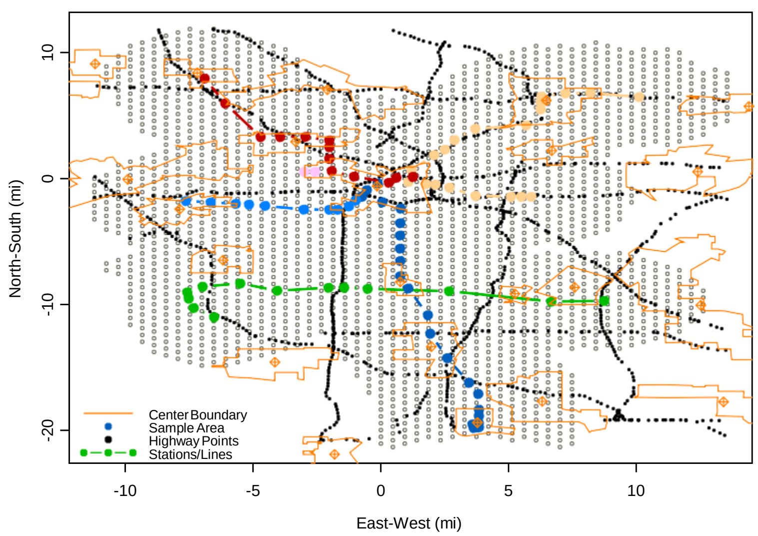


Figure 4. Working Data for Station Access & Employment



CBD is but one. The map makes clear that centers exist throughout the MSA and well outside the study area. Figure 4 overlays the centers over the stations and rail lines. An important observations can be made from this map. Recalling that the centers are defined as contiguous employment that is denser than the surrounding area, the first observation is that station location choice does not look to be a problem with regard to sample selection. That is, none of the Green line is in a center, most of the Expo, Blue, and Gold line stations are located outside of centers in relatively less dense areas. Only the Red Line has a majority of the stations in centers.

With the geography laid out, the data used to inform the analysis consists of Census employment and population data and employment data from the National Employment Time Series (NETS) data. The Census data is used because it the most commonly used data and the Census tract the most commonly used unit of analysis. The Census data offer the best systematically available, high-quality data. The problem is that the data have been traditionally available only decennially. And, the Census tract is established to approximate roughly a neighborhood of about 4,000 residents. These are based on population and often not well-suited for employment data. Urban areas that may be heavily dominated by employment may have few residents. As such, the tract areas used to accumulate the requisite number of residents may include many different types of employment and employment densities. Moreover, the station locations are generally geocoded, but the Census tracts are areas. Being within one mile of the station can mean many things: from any portion of a tract, from the centroid, such that the whole tract is within a mile, etc. Over a large sample, it is hoped that these sorts of noise are not influential, but in smaller samples like as is the case with the small number of stations in metropolitan systems, the spatial units can pose a challenge.

The second source of data we use in the analysis is the NETS data. In contrast to the Census data, the NETS data are point specific. Establishments are geocoded for location and data concerning number of employees and industrial classification are included. The geocoded establishments means that a different spatial unit can be used. We use hexes that are one-quarter square miles. This are small units and uniform throughout the region, but we lose the ability to readily incorporate other data, like Census demographics. The lagged dependent variable is our control for these variables.

4 Results

4.1 Typical DiD

The first set of results follow a "standard" approach to difference-in-differences, the model laid out in Equation (2). The data include the 6,804 hexes – each of a uniform area of one-quarter square mile – that are within four miles from an existing Metro station. The employment data in the hexes is constructed from the underlying NETS establishment data. Table 1 reports mixed results. The first observation from this specification and these models is that much of the varia-

Table 1: Specification from Equation (2)

Model	1	2	3
d^{2005}	0.094	0.094	0.098
	(1.90)	(1.90)	(2.13)
d^{2009}	0.085	0.085	0.089
	(1.72)	(1.72)	(1.93)
$d^{station}$	0.777	0.450	0.447
	(13.51)	(6.17)	(6.40)
$d^{newStation}$	− 0.226	− 0.22	− 0.15
	(2.08)	9	7
$dist^{station}$	−	− 0.20	− 0.08
	−	2	2
City Fixed-Effects	*Yes*	(7.20)	(2.74)
Center Fixed-Effects	*No*	*No*	*Yes*
r^2	0.183	0.189	0.298

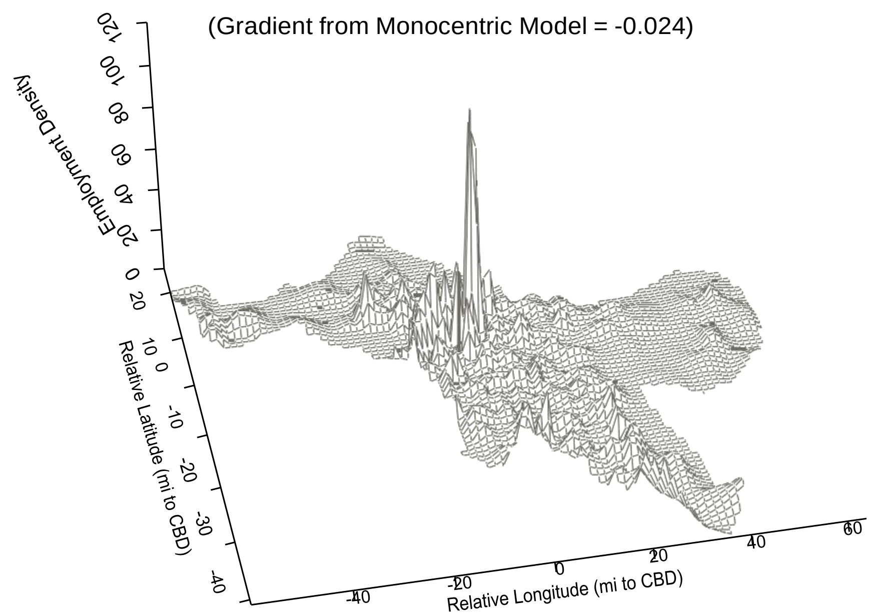
tion in employment density is unexplained for – even with city fixed-effects (not reported in the table), at most 19 percent of the variation is explained. With the center fixed-effects (not reported in the table), the power improves to 29 percent. Because we suspect that centers ought to be employment amenities on balance, we expected that stations and new stations would be positive and that the employment density would decline with distance from the stations. Surprisingly, the estimated impacts of new stations is negative and significant. It is possible that new stations have negative externalities such as crime, pollution, and traffic associated with them. But these are generally associated with residences and impacts to house prices (Bowes and Ihlanfeldt 2001). If stations were generally associated with negative spillovers, it would likely be manifest in both the new and old stations. In Table 1, locations the ever get a station are significantly and positively related to employment density. The fact that the distance variable was negative was consistent the our priors, but inconsistent with the parameter on new stations.

4.2 Persistence

The second set of results leverages off previous work the persistence in the Los Angeles MSA. Redfearn (2009b) demonstrated a high degree of stability in the spatial orientation of employment and population. The low power of the previous models in Table 1 were likely a function of the sparse set of covariates that govern firm location choice. Those models had only stations and the city fixed-effects to guide them.

Figures 5 shows the employment density in the more populous Census tracts in the Los

Figure 5. Employment Density in L.A. Basin Urban Tracts: 2000



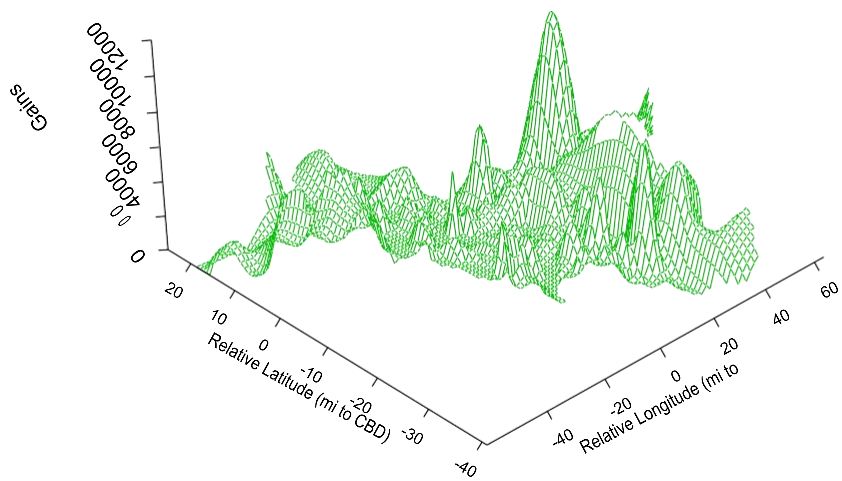
Angeles metropolitan area.[4] To orient yourself, look to the spike in employment density in the center. Though the vast majority of employment in the metropolitan area is outside the CBD, clearly Los Angeles has a center. Outside the downtown area of Los Angeles, it is also clear that a simply monocentric model fails to capture the multinodal nature of employment in the region. Oxnard is the local peak to the west of the surface. Riverside and San Bernardino are the local peaks to the eastern edge of the map. The foreground traces the Pacific Ocean, from Oxnard through Dana Point in south Orange County.

Figures 6 and 7 shows the respective gains and losses in net employment using the same geography and orientation. Figure 6 is a hard map to read, but clearly makes its point. That is,

[4]Following Redfearn (2007), a so-called 'compact set of [Census] tracts' are used in these surface maps. He does this because of the mapping. The five-county metropolitan statistical area is huge spanning the Pacific Ocean along Ventura, Los Angeles, and Orange Counties across Riverside and San Bernardino Counties to the eastern border of California. The 'compact' set of Census tracts are contiguous and have sufficient employment to allow for the nonparametric surface maps to be built.

Figure 6. Gains in Employment in L.A. Urban Tracts: 1980-2000

(Total Gains = 1.4 Million)



that the big gains in employment during the 1980-2000 period were at the periphery, especially to the western half the region, the core employment areas lagged. But where there is this broad trend, there are many exceptions – multiple peaks that reflect local fundamentals and a complex set of histories. The reason these figures are included here are to motivate the use of these local peaks as units of analysis and controls in the DiD regressions.

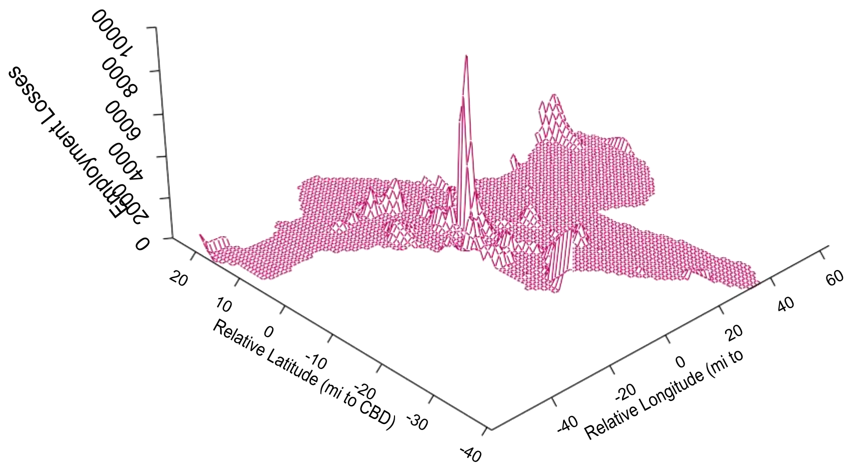
Figure 7 may make the use of centers more compelling. Figure 7 maps net losses in employment over the 1980-2000 period. The spike in the center of the map is not exactly the CBD and reflects of concentration in aerospace and the end of the Cold War. In fact, the contraction of defense and aerospace employment can explain most of the employment loss. Much of the remaining loss is associated with the loss of employment at some older larger scale manufacturing.

Keeping in mind the requirement of the parallel trend assumption in Equation (1). These two figures suggest that employment growth and decline were neither random nor parallel in the areas around the stations. The maps make estimating how bias might manifest itself difficult. But, it is clear that imposing the parallel trend is problematic.

Figures 8 and 9 echo the same dynamics in population as does those in employment. Figure 8 maps the population density around the Los Angeles Basin. Note that the orientation has been rotated 180 degrees to make lower population densities in the periphery more clear. Now, Oxnard

Figure 7. Employment Losses in L.A. Urban Tracts: 1980-2000

(Total Losses = -0.778 Million)



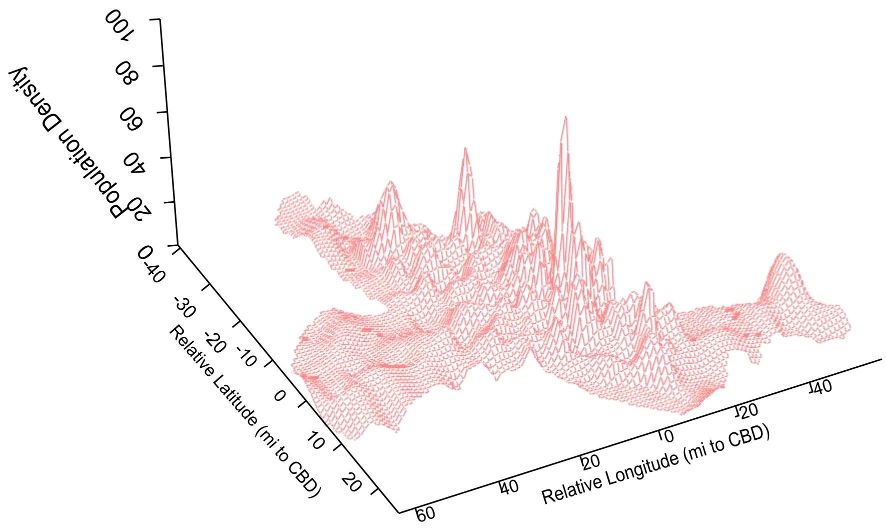
is the prominent peak at the very right (the West) of the map; San Bernardino and Riverside are the the left (East). The map is more varied than for the employment surfaces. But, like the employment surface, there is no simple gradient. Using distance from the CBD as a control for general trends in density look to be poor.

Figure 9 shows the net population gain (again reoriented from the Pacific Ocean looking up to the north, looking to the right as East). The marked growth of population in the periphery is clear, but so too are the multiple nodes of growth throughout the region. The nodes in the population and employment maps roughly correspond to the subcenters mapped above in Figure 3.

The figures suggest a) that use of the CBD as the sole control for geography is incomplete and b) that both population and employment growth was neither random nor parallel. By showing changes in employment and population in levels, the maps made clear that suburbanization was underway. But, the change-figures mask how much of both population and employment was oriented similarly despite the growth of each. The CBD remained the CBD in 1980 as it was is 2000, and as it is today. All of the major highway systems during the period were largely unchanged. Certainly, some lanes have been added, but the backbone of the system had largely been established by 1980. We will make use of this persistence to control better the cofactors that

Figure 8. Population Density in L.A. Basin Urban Tracts: 2000

(Gradient from Monocentric Model = -0.029)



contribute to changes in employment and population over time.

Table 2 reports the regression of population today and lagged population. These regressions are based on Census population data from 1980, 1990, and 2000. The table shows a remarkable

Table 2: Employment Density = f(past Employment Density)

Dep. Var.	empDens00	empDens90	empDens00	empDens00
constant	0.503	1.692	1.51	0.291
	(8.36)	(22.42)	(18.37)	(4.770)
empDens90	0.863	–	–	0.705
	(179.6)			(54.20)
empDens80	–	0.823	0.844	0.221
		(106.99)	(88.49)	(17.59)
r^2	0.924	0.883	0.774	0.934

explanatory power using no covariates beyond what was there a decade before. Against this backdrop of stability, the remaining variation will be examined against station location and the sites of new stations.

Table 3 reports the analog to population in Table 2. The results are that much more stable. The explanatory power of population density lagged by 10 or 20 years is 90 percent or more. Indeed

Figure 9. Change in Population in L.A. Urban Tracts: 1980-2000

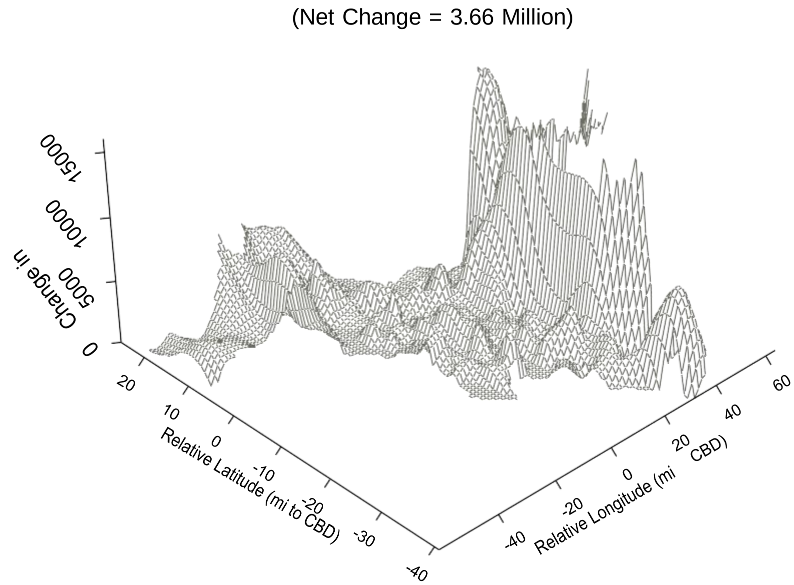


Table 3: **Population Density = f(past Population Density)**

Dep. Var.	popDens00	popDens90	popDens00	popDens00
constant	0.326	− 0.659	− 0.256	0.504
	(4.18)	(5.88)	(1.68)	(6.18)
popDens90	1.054	–	–	1.1526
	(268.3)			(77.20)
popDens80	–	1.236	1.294	− 0.130
		(179.11)	(137.90)	(6.82)
r^2	0.967	0.932	0.891	0.969

almost 97 percent of the population density in 2000 can be explained by population density in 1990.

The spatial distribution of economic activity is discussed at length in Redfearn (2009b). For our purposes, the two difference from the original set of models reported in Table 1 address two earlier concerns. The first is that the employment and population density are determined via a complex process that makes modeling them explicitly difficult. Though the persistence in these table suggests otherwise, it would be a mistake to assume nothing changes. Rather, it is more likely that places that are already dense are more likely to get more. Tall buildings do not spring up in residential neighborhoods; they are built where other tall buildings exist. The centers are an accumulation of marginal change. The first concern we had about low explanatory power in Table 1 will be addressed by the lagged employment density.

The second concern was the more significant problem of parallel trends required to retrieve the underlying impact of new passenger rail stations. Here, we use the centers, rather than other units of analysis to control for differential trends in growth. The two changes from the models in Table 1 are not unrelated. The differential trends that we saw in the surface mapping are likely to reinforce the same persistence. That is, where treatment and controls are likely to trend differently, they are likely to be in and out of centers. Of course, the underlying fundamentals vary across centers, but systematically centers are places that employment change is inherent. To turn a residential neighborhood into an employment center would require a hostile approval process. To add the same number of additional employees to an existing neighborhood is likely to be far more feasible.

The final set of results in Table 3 incorporate the discussion above and the model summarized in Equation (4). These results also reflect an improved geographic unit, the hexes comprised of the NETS establishment data. These small, consistent geographies should help mitigate noise from the earlier work on Census tracts. The results are generally as expected with regard to

Table 4: **Emp. Density = f(past Emp. Density, Centers, Stations)**

Sample	All	Red Line	Gold Line
constant	23.41	83.74	37.60
	(0.00)	(0.05)	(0.04)
empDens00	0.837	0.910	0.828
	(140.7)	(66.57)	(59.11)
$d^{newSta05}$	-9.692	$-$	-74.705
	(0.20)	$-$	(1.14)
$d^{newSta00}$	66.475	-46.365	$-$
	(1.92)	(0.49)	$-$
$d^{inCenter}$	151.41	-215.80	262.66
	(3.52)	(1.43)	(2.83)
$d^{inCenter} * d^{newSta00}$	279.01	379.78	30.01
	(2.69)	(2.14)	(0.11)
r^2	0.912	0.940	0.878

the explanatory power. The lagged employment data from 2000 to 2009 yield an r-squared of about 90 percent. Using the full sample, being in a center is highly significant and positive as is being a hex with a new station. But most interestingly is the interaction of being in a center and getting a new station which suggests that the stations that most benefit from new stations are those that are better able to exploit the opportunity to redevelop and densify in the area around the stations.

Not surprisingly, the results vary when the same model is applied to different stations along two different Metro lines. The Red Line is older and is largely located in centers. For this subsample, the interaction alone significantly explains the change in employment density from the lagged density. For the Gold Line, which has fewer stations in centers and has many newer stations, the interaction is not significant. This may be a function of the newness of the stations; redevelopment in California is famously slow, but particularly slow in infill neighborhoods. The centers along the Gold Line are highly significant and add positively to employment density.

5 Conclusion & Extensions

The motivation for this paper is to understand the impact of new passenger rail station on land use intensity. We focus primarily on employment and find that new stations add statistically significant employment in the surrounding areas. It is, of course, not an unconditional finding. While the average population impact of new stations was positive and significant, the results varied by subsample. In both, stations were relevant and contributed to employment density, but varied by the effect of new versus existing stations. This hints at the notion that stations may induce new jobs at longer time horizons. The Red Line, the more mature of the two subsamples, showed a marked contribution of new station – where those "new" stations were a decade old by the time we measured their impacts.

The Gold Line stations, by contrast were added to the system in 2003 (north/east to Pasadena) and 2009 (south/east to Atlantic). So, it may be that more time is needed to see investment manifest itself. The 2009 stations were delivered at a time when no one was developing real estate and has only recently seen a return. Alternatively, the Gold Line experience to the north hints at regulation. The growth along the Gold Line varies, but is most pronounced in the centers. A station like Allen Street, which is not and sits among mostly residential neighborhoods has seen no material change in a decade. The topic of regulation is one that requires more attention.

While the motivation to start the research was straightforward, the process and the evolution of the empirics has not been. A “standard” version of differences-in-difference (DiD) led to counterintuitive results that led to a longer exploration of what is required of DiD. The process yield two markedly different empirical executions of the same basic notion. In both, treatment and control groups are required; time variation to identify the true treatment effect is required; and covariates are required. The first approach (Equation (2 and Table 1) yielded poor explanatory power and a negative spillover of new stations. A story can be told about new crime, congestion, and pollution that might make new stations exhibit some negative effects. But, the other stations – the existing stations – were significant and positive contributors to surrounding employment with all the same issues.

As a result of this, we revisited the basic execution and used employment centers as controls and we exploited the persistent nature of employment to better control other elements of firm location and expansion/contraction choices. This approach seemed far more compelling as to the underlying dynamics, especially land use patterns which were likely to be consistent over time. In doing so, this approach yielded a significant role for centers and new stations in yielding new employment density.

The contribution of this research is twofold. The first is the finding that new stations have added significantly along some line. The second finding may be more compelling. Empirics in urban settings often does not match the theoretical requirements of tools like difference-in-differences. This needs additional testing, but at this preliminary point, the treatment of the CBD as a meaningful control broadly for density seems overly simple. Furthermore, the results suggest that centers are useful in understanding urban form and the evolution of the spatial distribution of employment and population. The case of passenger rail stations is but one example of other settings where centers may prove useful as well.

References

Abadie, A. (2005): "Semiparametric difference-in-differences estimators," *The Review of Economic Studies*, 72(1), 1–19.

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

Altshuler, A. A., and D. E. Luberoff (2003): *Mega-projects: The changing politics of urban public investment*. Brookings Institution Press.

Baum-Snow, N., and M. E. Kahn (2000): "The effects of new public projects to expand urban rail transit," *Journal of Public Economics*, 77(2), 241–263.

Baum-Snow, N., M. E. Kahn, and R. Voith (2005): "Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000 [with Comment]," *Brookings-Wharton Papers on Urban Affairs*, pp. 147–206.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How much should we trust differences-in-differences estimates?," *The Quarterly Journal of Economics*, 119(1), 249–275.

Billings, S. B. (2011): "Estimating the value of a new transit option," *Regional Science and Urban Economics*, 41(6), 525–536.

Bogart, W. T., and W. Ferry (1999): "Employment Centres in Greater Cleveland: Evidence of Evolution in a Formerly Monocentric City," *Urban Studies*, 36(12), 2099–2110.

Bowes, D. R., and K. R. Ihlanfeldt (2001): "Identifying the Impacts of Rail Transit Stations on Residential Property Values," *Journal of Urban Economics*, 50, 1–25.

Card, D., and A. B. Krueger (1994): "Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania," *The American Economic Review*, 84(4), 772–84.

Craig, S. G., and P. T. Ng (2001): "Using Quantile Smoothing Splines to Identify Employment Subcenters in a Multicentric Urban Area," *Journal of Urban Economics*, 49(1), 100–120.

Donald, S. G., and K. Lang (2007): "Inference with difference-in-differences and other panel data," *The Review of Economics and Statistics*, 89(2), 221–233.

Gibbons, S., and S. Machin (2008): “Valuing school quality, better transport, and lower crime: evidence from house prices,” *Oxford Review of Economic Policy*, 24(1), 99–119.

Giuliano, G., and C. L. Redfearn (2007): “Employment Concentrations in Los Angeles, 1980-2000,” *Environment and Planning A*, 39(12), 2935–2957.

Giuliano, G., and K. A. Small (1991): “Subcenters in the Los Angeles Region,” *Regional Science and Urban Economics*, 21(2), 163–182.

Imbens, G. W. (2004): “Nonparametric estimation of average treatment effects under exogeneity: A review,” *Review of Economics and statistics*, 86(1), 4–29.

Kahn, M. E. (2007): “Gentrification Trends in New Transit-Oriented Communities: Evidence from 14 Cities That Expanded and Built Rail Transit Systems,” *Real Estate Economics*, 35(2), 155–182.

Lee, B. (2007): ““Edge” or “Edgeless” Cities? Urban Spatial Structure in US Metropolitan Areas, 1980 to 2000,” *Journal of Regional Science*, 47, 479–515.

McMillen, D. P. (2001): “Nonparametric Employment Subcenter Identification,” *Journal of Urban Economics*, 50(3), 448–473.

McMillen, D. P. (2010): “Issues in spatial data analysis,” *Journal of Regional Science*, 50(1), 119–141.

McMillen, D. P., and J. McDonald (2004): “Reaction of house prices to a new rapid transit line: Chicago’s midway line, 1983–1999,” *Real Estate Economics*, 32(3), 463–486.

McMillen, D. P., and S. C. Smith (2003): “The Number of Subcenters in Large Urban Areas,” *Journal of Urban Economics*, 53(3), 321–338.

Mills, E. S. (1967): “An Aggregative Model of Resource Allocation in a Metropolitan Area,” *American Economic Review*, 57, 197–210.

Muth, R. F. (1969): *Cities and Housing*. University of Chicago Press: Chicago, IL.

Redfearn, C. L. (2007): “The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area,” *Journal of Urban Economics*, 61(3), 519–541.

——— (2009a): "How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets," *Regional Science and Urban Economics*, 39(3), 297–306.

——— (2009b): "Persistence in urban form: the long-run durability of employment centers in metropolitan areas," *Regional Science and Urban Economics*, 39(2), 224–232.

Suzuki, H., R. Cervero, and K. Iuchi (2013): *Transforming cities with transit: Transit and land-use integration for sustainable urban development*. World Bank-free PDF.