How Does Light Rail Transit Impact Housing Affordability?

Student name: Mengqi (Archie) Chen

Email address: mengqic3@illinois.edu

Phone number: (860) 990-7681

School program: Master of Urban Planning, University of Illinois at Urbana-Champaign

Course title: UP 460 Transportation/Land Use Policy

Date originally submitted: 2020-12-18

Subject: How Does Light Rail Transit Impact Housing Affordability?

Abstract

Transit-oriented development (TOD) has gained popularity as a means of redressing a series of urban problems. TOD has, however, also been criticized for potentially posing risks for neighborhood equity and affordability. In order to research the validity of these concerns, the author utilizes a multiple linear regression (MLR) analysis that investigates whether and the extent to which the presence of Light Rail Transit (LRT) station areas, especially those in the core city of a county, is associated with housing cost burdens. The analysis is conducted for the neighborhood change between 2000 and 2010 in 2 counties—Multnomah County and Los Angeles County that built light rail transit systems since the 1980s. Results show no statistically significant evidence of more housing cost burdens around LRT station areas in both counties, even for those that are located in their core central cities, Portland and Los Angeles respectively. Hence, the study confirms the positive benefits of transit-oriented developments without more risks posing danger to equity and affordability. On the other hand, the author recognizes the limitations of the MLR analysis and recommends alternative modeling methods such as the spatial autoregressive model for future research on this topic. The author also emphasizes the necessity of taking into consideration the effect of other unobserved factors. One of the major unobserved factors the author mentions is policies that promote affordable housing. The author believes if connecting with adequate affordable housing policies, equity can be further ensured in neighborhoods under transit-oriented developments.

Keywords: housing cost burdens, neighborhood change, light rail transit (LRT), transit-oriented development (TOD)

Introduction

Transit-oriented development (TOD) has gained popularity as a means of redressing a number of urban problems, including traffic congestion, air population, incessant sprawl, and so forth. Usually, a scalable and quantitative method that considers all of the main factors of TODs, such as (1) physical and functional connection to the transit system, (2) mix of moderate to high intensity and diverse land uses, and (3) pedestrian- and cycle-friendly design, is used to define TOD. TOD has, however, also been criticized for potentially posing risks for neighborhood equity. One possible consequence is resulting in gentrification, a process that necessarily involves displacement due to neighborhood upgrading. New younger, whiter, high-education, or wealthier residents replace low-income, working-class, and minority residents to a certain degree (Marcuse, 1985). Grounded in longitudinal data analysis, Baker and Lee (2019) found no evidence of widespread gentrification around Light Rail Transit (LRT) station areas that are often in the form of transit-oriented developments in the 14 US cities studied (Table 1).

	All Tracts (Set 1)	Gentrifiable Tracts (Set 2)
Gentrification and TOD	San Francisco	San Francisco, Cleveland
Countergentrification and TOD	Cleveland, Portland	Portland
Gentrification and counter- or no TOD	Denver, Sacramento, Dallas	Denver, St. Louis, Buffalo
Neighborhood decline	San Diego, Buffalo, Los Angeles, Pittsburgh, Salt Lake City	San Diego, Los Angeles, Baltimore

Table 1. Summary of Baker and Lee's findings (2019)

Based on Baker and Lee's assumption, neighborhood upgrading without displacement is not accounted as gentrification, despite the fact that the neighborhood changes derived from transit-oriented development could impose challenges on housing affordability on the old and new residents. The rising property and land values experienced in some communities can make it difficult for middle- and low-income families–the very households that need access to transit the most–to afford to live within the new and existing communities around transit stations (Haughey

& Sherriff, 2010). One may argue that an increase in land prices is inevitable when neighborhoods are upgrading and becoming more attractive following the market mechanism.

To further respond to the concerns with TOD impacts on housing affordability, I thus ask: whether and the extent to which the presence of Light Rail Transit (LRT) station areas, especially those in the core city of a county, is associated with housing cost burdens. I focus on examining the percentage of housing cost-burdened households in measuring affordability challenges. To answer this question, I utilize multiple linear regression (MLR) to investigate the global connection between LRT station areas and neighborhood change longitudinally in 2 counties, Multnomah County and Los Angeles County that built light rail systems since the 1980s. Inspired by Baker and Lee's research, I choose Multnomah and Los Angeles Counties, because the core city each of them has respectively, Portland and Los Angeles, is found with different results in gentrification. In Portland, the presence of LRT station areas increased transit access without causing displacement. In Los Angeles, on the other hand, the presence of LRT station areas increased transit access while at the same time causing displacement. I think it is worth further studying whether the LRT impacts on housing affordability would differ (or not) in these two areas as well. Understanding whether or not and how much LRT station areas are associated with housing cost burdens will help urban planners develop a better plan for the communities around station areas. Some policy tools may need to go hand in hand with TODs to ensure equity and affordability.

Methodology

To measure housing affordability, there are usually three ways. The first approach is the housing cost to household income ratio in which a measure of housing costs (e.g. rent, mortgage payment, etc.) is compared to a measure of household income. In the U.S., the ratio is usually determined as 30 percent of household income. The second one is the income after housing cost or 'residual income' approach, "which seeks to define the household income remaining after housing costs are deducted and then comparing this to a low-income benchmark of some kind (Pettit et al., 2017, p. 235)." The third measure is the Housing Affordability Index (HAI) (National Association of Realtors, n.d.). For my study, I use the most common one, the housing cost to household income ratio at 30 percent.

To prepare data for the multiple linear regression (MLR) analysis, I first identify a list of 11 explanatory variables to 1 response variable that I use for the data of Multnomah and Los Angeles Counties respectively. Table 2 lists all the variables in the model describing the LRT' station areas' changing effects on housing cost burdens between 2000 and 2010.

Table 2. Description of variables

Variable	Description	Data Source
Change response variable 2000-2010		
Change in % housing cost-burdened households	Change in the percent of households that pay 30 percent or more of their income on housing costs between 2000 and 2010	US Census 2000; ACS 5-Year 2008-2012
11 Explanatory variables		
1. Key spatial variable		
LRT station buffer	Dummy variable: 1 if a census tract's centroid is inside a half-mile station buffer by 2010; otherwise 0	US Census TIGERFiles; TriMet Geospatial Data; Metro Developer's Site
Core city dummy	Dummy variable: 1 if a census tract's centroid is inside the core central city (i.e. Portland, Los Angeles) by 2010; otherwise 0	US Census TIGERFiles; TriMet Geospatial Data; Metro Developer's Site
LRT and core city interaction	Interaction terms: LRT station buffer x each of CBD dummy	US Census TIGERFiles; TriMet Geospatial Data; Metro Developer's Site

Variable	Description	Data Source
2. Change variable of socioeconomics 2000-2010		
Change in log median household income (MHI)	Change in natural log of MHI	LTDB
Change in % non-Hispanic white	Change in percent of the population who are non-Hispanic white	LTDB
Change in % bachelor and above	Change in percent of population age 25+ with a college degree and more	LTDB
Change in poverty rate	Change in percent of population's poverty status determined below poverty threshold	LTDB
Change in unemployment rate	Change in percent of unemployed civilians	LTDB
3. Change variable of housing 2000-2010		
Change in % renter occupied	Change in percent of renter-occupied housing units	LTDB
Change in log median rent	Change in natural log of median gross rent	LTDB
Change in log median home price	Change in natural log of median housing value	LTDB

Note: (1) LRT = light rail transit; LTDB = Brown University's Longitudinal Tract Database. (2) For Multnomah County, the core city dummy tests if one tract's centroid is in Portland; for Los Angeles County, it tests if one tract's centroid is in Los Angeles.

The explanatory variables chosen here are based on the study by Baker and Lee (2019) in order to make mine comparable to theirs. It aims to examine whether the affordability crisis has been worsened in Portland and Los Angeles since the opening of LRT stations there by 2010, even if the displacement did not necessarily happen widely. For Multnomah County, the study subject of LRT is MAX Light Rail; for Los Angeles County, the study subject is Metro Rail. My study is at the neighborhood level and census tracts serve as the basic unit of analysis because they most closely approximate the size of neighborhoods. The core city dummy variable, as well as its interaction term with the LRT station buffer, are the key variables that will show how the presence of LRT stations affect neighborhood changes of housing cost burdens. The LRT station dummy variable indicates whether or not a census tract's centroid is within a half-mile, straight-line radius of a light rail station. The core city dummy variable indicates whether or not a census tract is within a given core central city. For Multnomah County, it is Portland, while for Los Angeles County it is Los Angeles. The interaction term is a test of the additivity of the effects of the LRT station area and core city. If the interaction is significant, it can be understood that housing cost burdens occur differently in neighborhoods depending on whether they are in the core central city and station areas. With an interaction term, the coefficient of the LRT station buffer dummy will become a part of the baseline for a station neighborhood's cost-burdened household percentage. The interaction term is designed to capture city-specific station effects on neighborhood change. Hence, the sum of coefficients of the core city dummy and the interaction term reveals the extent of LRT impacts on station neighborhoods in the city. These dummy variables are calculated through spatial joins between (1) centroids of census tracts and 0.5-mile buffers of LRT stations and (2) centroids of census tracts and core central cities both through a "within" relationship.

Most of the census tract-level socioeconomic variables come from Brow University's Longitudinal Tract Database (LTDB), except for key geographic variables and the percent of housing cost-burdened households that derives from the U.S. Census Bureau 2000 and ACS 5-Year 2008-2014 survey. To convert count data from 2000 into 2010 boundaries, I use the weight information provided by the LTDB and then calculated the percentage of housing cost-burdened households. With non-spatial data ready by joins from 2000 Census and 2010 ACS datasets through GEOID of 2010, I then convert them into GIS data and joined with another dataset containing key dummy variables. The three key geographic variables, including station buffer, core city dummy variables, and their interaction term, on the other hand, are generated based on the shapefiles. In ArcGIS, I created a Euclidean buffer radius of a half-mile around every light rail station and determined (1) whether the centroid of a census tract is inside one or more station buffers, and (2) whether the centroid of a census tract is inside the core central city of each of the counties. The process described above is the same for both counties' data.

Results

The housing affordability crisis is a county-wide issue for both Multnomah and Los Angeles Counties, and it is worse for the latter. Over 2008-2012, over 42 percent of households in Multnomah County were under cost burdens, and it gets to 50 percent for Los Angeles County. The percentages of both counties increased at a rate of around 40 percent from 2000 to 2010.[1] Figures 1 and 2 show that the affordability crisis is more severe in the core central cities of Portland and Los Angles where the majority of LRT stations are located. However, this spatial clustering of cost-burdened households around LRT station areas does not necessarily mean the presence of station areas is relevant to the exacerbation of the affordability crisis over 2000-2010. That is why a statistical modeling analysis that can control other factors is required.

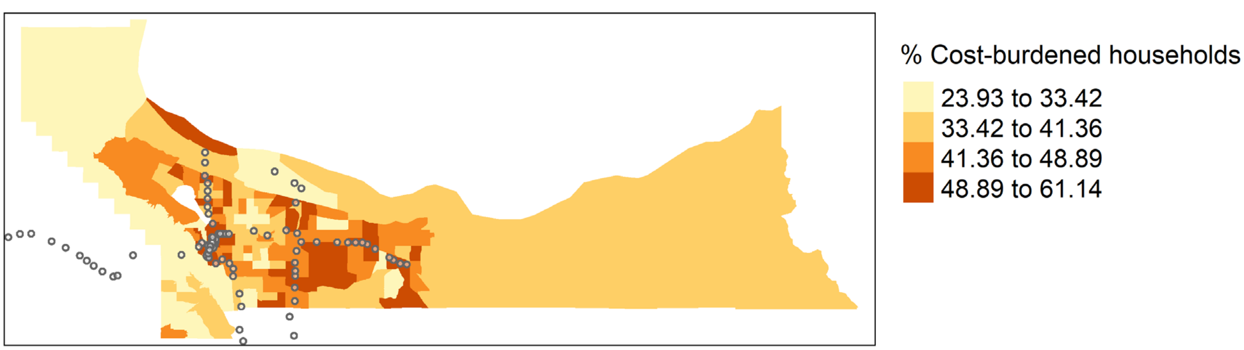

Figure 1. Distribution of % housing cost-burdened households in Multnomah County 2010
Note: White dots represent LRT station buffers.

---

[1] Sources: U.S. Census Bureau Decennial Census 2000, ACS 5-Year 2008-2012.

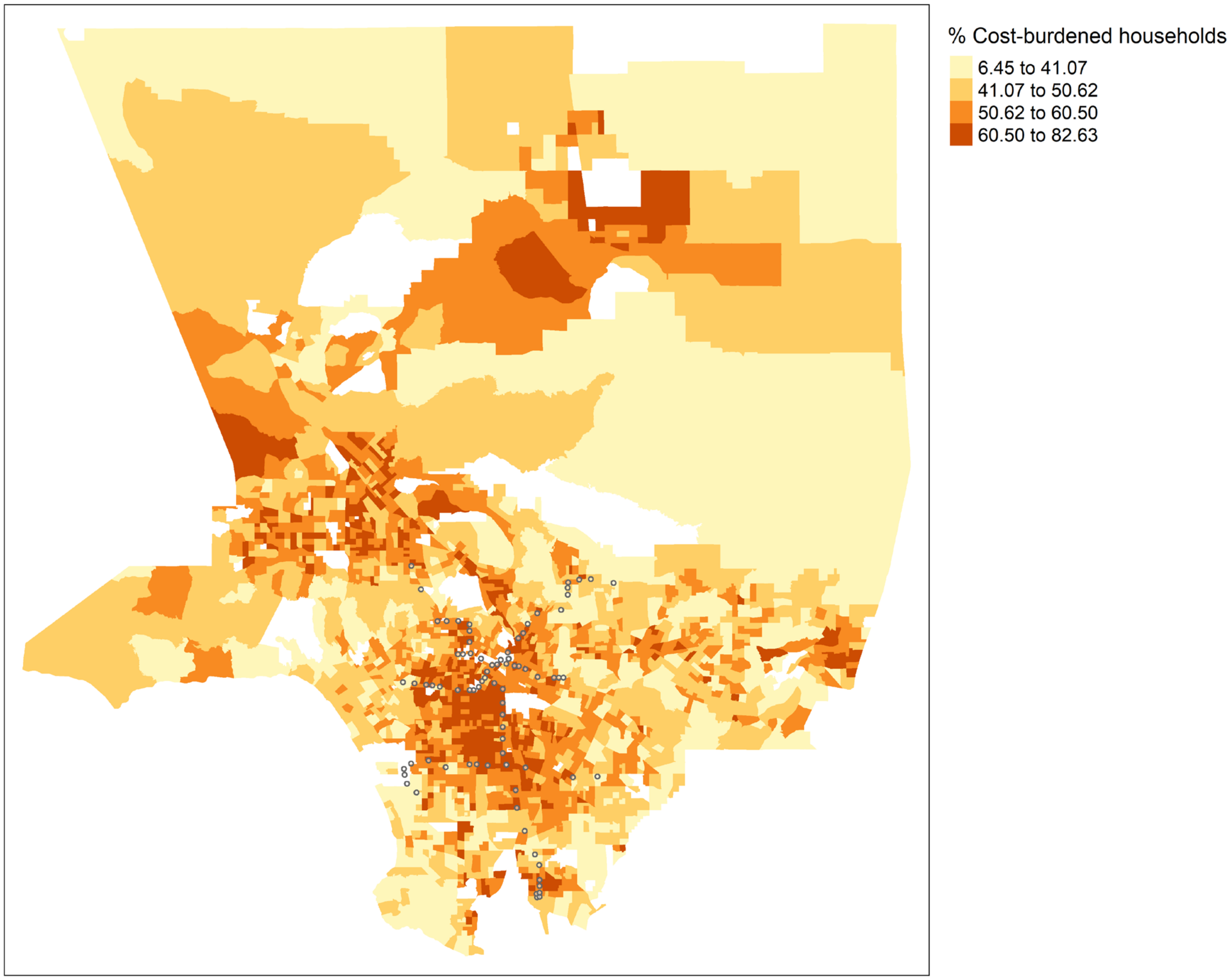

Figure 2. Distribution of % housing cost-burdened households in Los Angeles County 2010

Note: White dots represent LRT station buffers.

Because socioeconomic information is the control variable, I limit my discussion to the coefficients of the LRT station buffer dummy, core city dummy, and the interaction terms between these two dummy variables. Before getting to the modeling results, it is necessary to test if the models meet some key assumptions since linear regression is a parametric model. Table A in Appendix summarizes the diagnosis results for the models of the counties' data and both models have two same assumptions unmet, global statistics, and kurtosis. That is to say, my model is not a linear combination of explanatory variables and the data sets are heavy-tailed relative to a normal distribution due to their high kurtosis. In short, the coefficients estimated by my models could be biased and the estimates could be unreliable too. Moreover, the adjusted R-squared values for both models are also lower than 35 percent, suggesting their capability to explain the variance of the response variable is relatively limited. With an understanding of the limitation of my model, I think it is still important to explore the statistical results because to a certain extent the results can still reflect some levels of relationships.

Table 3 below summarizes the MLR results for both counties on the change of the percentage of burdened households between 2000 and 2010. If holding other variables fixed:

1. **Multnomah County**
   - For non-station neighborhoods that are not located in Portland, between 2000 and 2010, the percentage of burdened households increased 11.8 percent on average.
   - For station neighborhoods that are not located in Portland, between 2000 and 2010, the percentage of burdened households increased 10.3 percent on average.
   - Portland's station neighborhoods experienced a 2.2% relative increase in cost-burdened households compared to non-station neighborhoods, but the coefficients are not statistically significant.
2. **Los Angeles County**
   - For non-station neighborhoods that are not located in Los Angeles, between 2000 and 2010, the percentage of burdened households increased 7.8 percent on average.
   - For station neighborhoods that are not located in Los Angeles, between 2000 and 2010, the percentage of burdened households increased 7 percent on average.
   - Los Angeles's station neighborhoods experienced a 4% relative increase in cost-burdened households compared to non-station neighborhoods, but the coefficients are not statistically significant as well except for the core city dummy.

Table 3. Summary of MLR results

Variable	Multnomah County		Los Angeles County	
	Coefficients	P-Value	Coefficients	P-Value
(Intercept)	1.18E-01 ***	0.00E+00	7.83E-02 ***	0.00E+00
Key Varibale				
Within LRT station buffer	-1.50E-02	6.98E-01	-8.02E-03	4.47E-01
Within core city	-1.90E-02	2.02E-01	2.64E-02 ***	0.00E+00
Within LRT station buffer * within core city	4.11E-02	3.07E-01	1.29E-02	3.18E-01
LRT impact = Coeff. with core city + Coeff. within core city * within LRT station buffer	2.21E-02		3.93E-02	
Change Variable				
Log median household income	-1.82E-01 ***	4.90E-06	-1.73E-01 ***	0.00E+00
% Non-Hispanic white	-1.63E-01 **	4.89E-02	-1.24E-01 ***	2.00E-06
% Bachelor and above	-2.18E-05 ***	9.77E-03	1.00E-07	9.67E-01
Poverty rate	1.48E-04	8.40E-01	1.63E-03 ***	0.00E+00
Unemployment rate	4.02E-05	3.29E-01	6.69E-05 ***	9.00E-07
% Renter occupied	-1.22E-01	2.48E-01	-3.37E-02	2.00E-01
Log median rent	2.10E-02	5.89E-01	2.92E-02 ***	5.70E-06
Log median home value	8.37E-02 **	1.94E-02	3.25E-02 ***	1.10E-06
Number of observatoins:	170		2231	
Adjusted R-squared:	0.3481		0.2615	

Note: *p < .10, **p < .05, ***p < .01.

3. Summary[2]

In general, if a neighborhood is not located in the core central city, the presence of a Light Rail station area there can slow down the increase in the percentage of burdened households. If a neighborhood is located in the core central city, the presence of a Light Rail station area can exacerbate the problem, and the situation would be worse in Los Angeles in comparison to Portland. However, there is no statistically significant evidence of more hosing cost burdens around LRT station areas in both counties.

[2] Please note that I am aware that most of key variables' coefficients are not statistically significant.

Discussion

Based on the results of my study as well as Baker and Lee's, no statistically significant evidence of widespread gentrification and housing cost burdens at the neighborhoods around LRT station areas in the studied counties and cities is found. It confirms the positive benefits of transit-oriented developments (TODs) without more risks posing danger to equity and affordability. Figure 3 shows that Americans lower their housing costs by moving to suburbs whereas their transportation costs increase largely. TODs can help concur with this trade-off. If connecting with adequate affordable housing policies, equity can be further ensured. Hence, the general public does not need to be highly skeptical and concerned about the downsides of the TODs. These findings can contribute to advocating transit-oriented developments like a light rail transit station area.

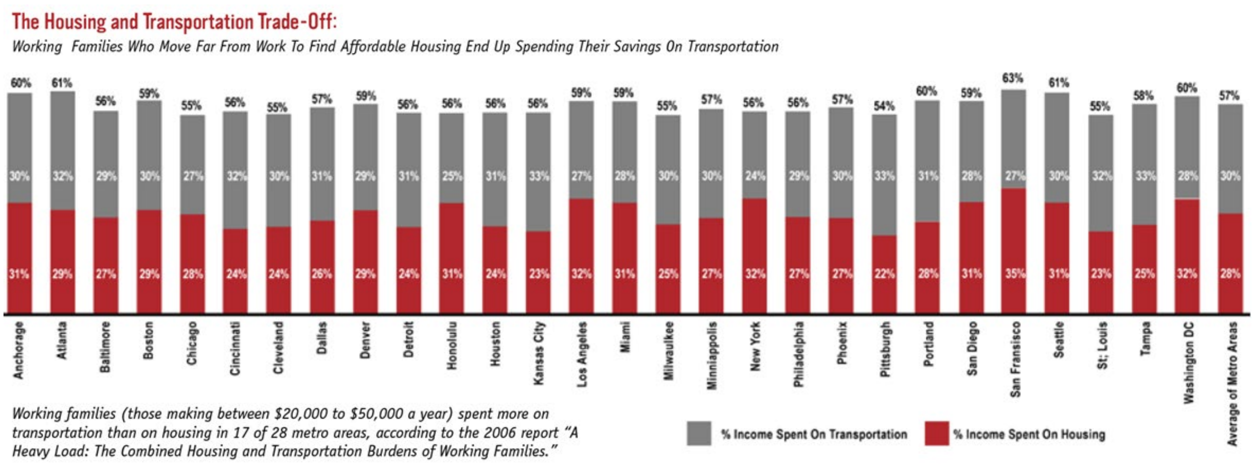

Figure 3. The Housing and Transportation Trade-Off

Sources: Center for TOD, 2009, p. 4.

However, my studies have some limitations that may influence the validity of the results. Figures 1 and 2 above show that over 2008-2012 the spatial clustering of more cost-burdened households exists around LRT station areas especially for those that are in the core central cities. I also conduct a test through the global Moran's I statistics and confirm that this positive spatial autocorrelation is statistically significant for the change in the percentage of cost-burdened households between 2000 and 2010 (Appendix Figure A). Therefore, a modeling approach that takes into account the possibility of spatial autocorrelation of the response variable and/or the spatially varying parameters of explanatory variables, which is a response to the locality and heterogeneity of the housing market, is required. Such alternative modeling methods may produce better results than the multiple linear regression. For example, the spatial autoregressive model (SAR) used by Baker and Lee's study can be an alternative, since it accounts for spatial dependence that potentially exists among the census tract characteristics. It is highly possible that the response variable in a given census tract is correlated to the values of the same variable in adjacent census tracts and the SAR model would capture this effect (Baker & Lee, 2019, pp. 38–39).

Another alternative can be geographically weighted regression (GWR) since it takes into consideration spatial heterogeneity, but it seems to be arbitrary and bold to assume each census tract varies fundamentally from each other (Tomal, 2020). Sometimes the influences of selected explanatory variables on one place can be combined with its neighbor(s), whereas the GWR model is conducted at the individual level only. Given this comparison, the SAR model would be a better alternative for the analysis of my research question. Besides the spatial component, such a high dimensional data that has a number of variables may be beyond the explanatory capacity of linear models. In addition, there may be multicollinearity between variables that undermine the explanatory power of the key variables. Future research needs to test the correlation between variables first. Also, linear regression models can only represent linear relationships that may not fit for the association I would like to study between the presence of LRT stations and housing cost burdens. Another approach to consider is to use machine learning methods like random forests. With many decision trees present breaking down variables, I can also learn about which variables are necessary to include and which of them are not. When a better modeling method is selected, I can further conduct similar studies on other cities that have LRT station areas and continue examining whether the presence of a station area can widely cause more housing cost burdens to the residents.

Another factor to adjust is the measure of housing affordability. The cost-to-income ratio may overestimate housing cost burdens. For example, 30 percent or even 50 percent for one earning $1 million annually is less a burden than moderate- and low-income persons. Residual income approach may be better in expensive housing markets like Portland and Los Angeles. I can redefine the measure of housing affordability by calculating the gap between actual reported median monthly housing costs and the residual income available for housing costs of a census tract. Residual income available for housing costs is the subtraction of estimated monthly non-housing costs of self-sufficiency in the county from low-income median monthly household income (AMI). By HUD definitions, a 2-adult household without children earning between 50 and 80 percent of AMI is considered a "low-income household". For future study, low-income households can be defined as earning 80 percent or less of the AMI.

Overall, linear models are far from perfect as it assumes unobserved factors—those that are not included in the model—do not affect the variance of the response variable. One example of the unobserved factors is transportation cost. Although housing prices and rents could rise higher around station neighborhoods, residents have the benefit of fewer transportation costs and hence would have more money to spend on housing without rising housing cost burdens. Another major unobserved factor is policies that promote affordable housing. If the studied counties or cities have proper affordable housing policies, LRT impacts to increase housing cost burdens could be weakened. For instance, there is a growing consensus that communities that provide housing for a mix of incomes produce better economic, social, and environmental outcomes for all residents (TOD, 2007, p. 3). Both my studied counties' station neighborhoods may provide housing for a mix of incomes, which makes it possible for people of all incomes to access a wide variety of jobs as well as opportunities and be able to afford housing and services.

State of California law provides an adjustable density bonus[3] in exchange for including affordable housing units in new construction, and local agencies may build upon and augment these incentives if they so choose. Los Angeles County in particular establishes the Green and Blue Line TOD Ordinance and it applies in a number of transit-oriented districts. This ordinance allows a density bonus of up to 50 percent to be allowed if at least one-third of the units are provided for low-income households or half of the units are reserved for qualifying senior citizens (for a minimum of 30 years) (Los Angeles County Metropolitan Transportation Authority, n.d.). These incentives can encourage developers to provide affordable housing in exchange for density, and it makes sense to build around transit stations because low-income populations benefit the most from transit access and use it at greater rates than the general population. The City of Portland, on the other hand, also has a policy for inclusionary housing. Portland requires that all residential buildings proposing 20 or more units provide a percentage of the new units at rents affordable to households at 80 percent of the median family income or below (City of Portland, Oregon, n.d.). However, it is challenging to quantify the effects of these policies in a statistical model. If future research is conducted, one should first identify the release and operation date of each policy and then figure out ways to quantify them. One potential

[3] Under the Density Bonus Ordinance, one may be able to build more residential units on one's property than what is allowed by code.

approach could be a dummy variable capturing whether a certain policy is adopted or not for given years. Then, a panel data analysis seems more appropriate than a cross-sectional one. As discussed above, many improvements can be made based on this study. Although this study certainly has limitations, it provides a necessary component to understanding transit's role in the housing affordability crisis.

References

Note: please see Table 2 for data sources.

Baker, D. M., & Lee, B. (2019). How Does Light Rail Transit (LRT) Impact Gentrification? Evidence from Fourteen US Urbanized Areas. *Journal of Planning Education and Research*, *39*(1), 35–49. https://doi.org/10.1177/0739456X17713619

Center for TOD. (2009). TOD201 Mixed income housing near transit: Increasing affordability with location efficiency.

City of Portland, Oregon. (n.d.). *Inclusionary Housing*. Portland.Gov. Retrieved December 18, 2020, from https://www.portland.gov/phb/inclusionary-housing

Haughey, R., & Sherriff, R. (2010). *Challenges and Policy Options for Creating and Preserving Affordable Housing near Transit and in Other Location-Efficient Areas*. National Housing Conference.

Los Angeles County Metropolitan Transportation Authority. (n.d.). *Incentives & Bonuses*. LA Metro. Retrieved December 18, 2020, from https://www.metro.net/projects/tod-toolkit/incentives-bonuses/

Marcuse, P. (1985). *Gentrification, Abandonment, and Displacement: Connections, Causes, and Policy Responses in New York City*. *28*, 195–240.

National Association of Realtors. (n.d.). *Housing Affordability Index*. www.nar.realtor. Retrieved October 14, 2020, from https://www.nar.realtor/research-and-statistics/housing-statistics/housing-affordability-index

Pettit, C., Tice, A., & Randolph, B. (2017). Using an Online Spatial Analytics Workbench for Understanding Housing Affordability in Sydney. In P. Thakuriah, N. Tilahun, & M.

Zellner (Eds.), *Seeing Cities Through Big Data* (pp. 233–255). Springer International Publishing. https://doi.org/10.1007/978-3-319-40902-3_14

Tomal, M. (2020). Modelling Housing Rents Using Spatial Autoregressive Geographically Weighted Regression: A Case Study in Cracow, Poland. *ISPRS International Journal of Geo-Information*, *9*(6). https://doi.org/10.3390/ijgi9060346

Appendix

Table A. Summary of MLR model diagnosis results

Requirement	Multnomah County			Los Angeles County		
	Value	P-value	Decision	Value	P-value	Decision
Global Stat	1.08E+01	2.94E-02	Not satisfied	4.60E+01	2.43E-09	Not satisfied
Skewness	2.77E+00	9.61E-02	Acceptable	6.35E-02	8.01E-01	Acceptable
Kurtosis	7.95E+00	4.81E-03	Not satisfied	4.29E+01	5.76E-11	Not satisfied
Link Function	5.80E-03	9.39E-01	Acceptable	1.94E-02	8.89E-01	Acceptable
Heteroscedasticity	3.92E-02	8.43E-01	Acceptable	3.05E+00	8.09E-02	Acceptable

Figure A. Summary result for the global Moral's I test on the change in percent of housing cost-burdened households of Multnomah County

```
        Moran I test under randomisation

data:  multn0010_sf10.3$p_burden_df
weights: w.mul.df

Moran I statistic standard deviate = 4.4108, p-value = 5.15e-06
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
       0.18254921       -0.00591716        0.00182572
```