ABSTRACT

Title of Dissertation:	HOUSING VALUE AND LIGHT RAIL TRANSIT CONSTRUCTION: EVIDENCE FROM THREE ESSAYS
	Qiong Peng, Doctor of Philosophy, 2020
Dissertation directed by:	Professor Gerrit-Jan Knaap, Department of Urban Studies and Planning

In three essays, this dissertation explores what's the determinants of multifamily rents and whether an anticipated investment in light rail transit influences multifamily rents and single family housing prices in the rail transit pre-service period.

In the first essay, I applied a multilevel linear model approach to account for the multifamily housing hierarchical data structure, and assessed the effects of service provision and management on multifamily rents. The findings show that pet allowance, availability of a short-term lease, and storage service increase rents significantly, while general renovations and availability of services for those with disabilities do not increase rents.

The second essay empirically tests whether light rail transit in the pre-service period impacts multifamily housing rent in the transit corridor. Two approaches, a first-difference method and a difference-in-difference method, are used to test the

research question. The results indicate that the rents of two-bedroom, three-bedroom, and four-bedroom units within a half mile from planned light rail stops have significantly increased from 2015 to 2018 compared with the rent of units in other areas in Montgomery County.

The third essay examines the temporal and spatial variation of the effect of the Purple Line on single-family home prices during the rail line pre-service period. The results show that the housing market saw a premium in 2012, the year the Purple Line project progressed into the preliminary engineering phase. The results also show that the effect of the new light rail transit line is distributed unevenly across the catchment areas of newly built stations and established stations.

HOUSING VALUE AND LIGHT RAIL TRANSIT CONSTRUCTION: EVIDENCE FROM THREE ESSAYS

by

Qiong Peng

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Gerrit-Jan Knaap, Chair
Professor Casey J. Dawkins
Professor Paul Elhorst
Dr. Andrew McMillan
Professor Ingmar R. Prucha

Acknowledgements

First and foremost, I would like thank my advisor Gerrit-Jan Knaap for guiding me through a long, enjoyable, and fulfilling journey of Ph.D. study. I deeply appreciate his training and supports. He trained me to conduct meaningful and rigorous research and generously supported me in many ways. He provided timely, detailed, and insightful comments on the writing of this thesis. I could not finish my dissertation in five years without his wonderful guiding.

I would like to express my deep gratitude to the rest of my dissertation committee: Casey J. Dawkins, Paul Elhorst, Andrew McMillan, and Ingmar R. Prucha. This dissertation benefits a great deal from their comments. Casey has known me since I came to UMD five years ago, and he has provided me lots of valuable help over the past five years. Ingmar and Paul taught me econometric methods and gave me ingenious suggestions, which greatly improved the quality of my work and open my eyes.

I am also deeply grateful to Marie Howland who provided me long-time and strong supports. She is the all-weather supporter. Many thanks to Chengri Ding for his nice service on my comprehensive exam committee. I would like to thank Thomas Tippett from the Office of Montgomery County Executive, Jean M. Appleton from the Maryland Department of Planning, and Chao Liu for their work in helping me access key data. Many thanks to Linda C. Macri and Alyssa Wolice for their editing suggestions. I also had the great pleasure of working with Chao Liu, Nicholas Finio, Kerry Li Fang, Marccus D Hendricks, Hyunjoo Eom, Jinyhup Kim, Scott

Dempwolf, Andrew McMillan, Morgan Wortham, and Jiemin Wu. Thanks again to Thomas Tippett and Nicholas Finio. My conversations with them helped shape the tables featured in Appendix I.

I would like to acknowledge research support provided by the National Center for Smart Growth Research and Education and the U.S. Economic Development Administration, University Centers Program.

Gerrit and Nick were co-authors on Essay #2: *Do multifamily unit rents increase in response to planned light rail construction*? As described in the letter included in Appendix II, I made substantial contributions to the coauthored work included in Essay #2. Any errors or omissions are my own.

Last but not least, I am deeply indebted to my family. Without their relentless support, I could not have completed this research and would not have the energy to explore my potential.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIC	Akaike information criterion
LRT	Light rail transit
MLE	Maximum likelihood estimation
MLM	Multilevel linear model
OLS	Ordinary least square
REIT	Real estate investment trust
REML	Restricted maximum likelihood

1. Introduction

Urban areas around the world are confronting crowding, housing unaffordability, congestion, pollution, and other challenges. Light rail transit[1] is a popular public investment used to address congestion, reduce greenhouse gas emissions, and revitalize decayed urban centers, but the cost of construction and maintenance of light rail transit is substantial. Therefore, as a public policy consideration, light rail transit projects often garner attention from researchers and practitioners in the fields of transportation, housing, real estate, urban planning, economics, and the environment. The relationship of housing values and light rail transit is a hot topic in discussions of urban development and sustainability because housing availability and affordability are major components of urban life, and housing values are impacted by light rail transit projects. More specifically, examinations of property value increments and multifamily rent changes resulting from a light rail transit project present important implications for property taxation adjustments, benefit-cost analysis of transit investment, and assessments of the affordability of multifamily housing near a given transit project.

Despite the extensive volume of existing research regarding the influence of light rail transit on housing values, the debate is far from settled. First, more research about the impacts of light rail transit projects in their pre-service period is needed. A

[1] Urban rail transit is an all-encompassing term for various types of local rail systems providing passenger service within and around urban or suburban areas (Wikipedia). Urban rail transit infrastructure investment includes light rail transit, rapid transit, monorail, commuter rail, etc. (Wikipedia). Light rail transit is one of them and is popular given the advantages it has.

light rail transit project typically takes multiple years to progress from the planning phase to the operation phase. The influence of a rail transit project on housing value may emerge when the rail transit project is in its pre-service period. As such, investigation into the timing of housing value added by rail transit investments draws important implications for benefit-cost analysis of rail transit investment and property taxation.

Second, more research about multifamily housing is needed. Multifamily housing plays a vital role in the U.S. real estate market and multifamily housing accommodated 17.8 million multifamily renter households (14.5% of total households) in the United States in 2013[2]. The multifamily housing is a common option for low- and medium-income households. Distinct from single-family housing, multifamily housing has unique characteristics; for example, multifamily housing provides services for tenants and employs formal property management. Therefore, it is important to measure the effects of service and management on multifamily rents in order to optimize management methods, set rents, and attract potential tenants. Furthermore, the multifamily housing data structure is hierarchical, and failure to account for this hierarchical data structure will result in incorrect inferences when conducting hedonic analysis. Nevertheless, there is little research on the effects of service and management on rents, and of these studies, still fewer account for the hierarchical data structure.

[2] The number of households in multifamily rental housing comes from Multifamilybz.com. The definition of multifamily housing for this calculation is a structure with five or more units that is renter-occupied.

Third, a large percentage of multifamily housing units are located near light rail transit, but there is little to no existing research on the influence of light rail transit in the pre-service period on multifamily rents. This dearth of research is mainly due to the lack of available multifamily housing data. There is no concrete evidence that points to whether or not light rail transit projects in the pre-service period influence multifamily rents. Similarly, it is unclear whether the influence of light rail transit construction on multifamily rent varies across unit types.

Light rail transit remains popular but it represents a large-scale infrastructure investment. Even more, the timeline for rail transit projects often spans multiple years from the planning phase to operation. These underlying factors demonstrate why it is necessary to assess how light rail transit in its pre-service period influences both single-family housing value and multifamily rents. This dissertation seeks to fill the current research gaps by examining the aforementioned three challenges through three empirical essays. The first essay examines the determinants of multifamily rents – in particular, service and management variables – given the hierarchical data structure being taken into account. The second essay centers on the ongoing construction of the Purple Line in the Washington, D.C. metropolitan area as a case study, coupled with a new data source derived from a high-quality multifamily landlord rental survey. This study empirically tests whether light rail transit in the pre-service period impacts multifamily housing rent in the transit corridor. The third essay uses the ongoing construction of the Purple Line as a case study to investigate the timing of value added by rail transit investments and the spatial variability of the impact.

This dissertation is organized as follows. Chapter 2 focuses on the determinants of multifamily rents, specifically, the implicit values of service and management variables. Chapter 3 examines if light rail transit construction impacts multifamily rents. Chapter 4 sheds light on the temporal and spatial pattern of any such impacts of light rail transit construction on single-family home values. Chapter 5 replicates key findings and discusses some topics that are related to this dissertation but not explored sufficiently in existing literature.

2. Investigating the Effects of Service and Management on Multifamily Rents: A Multilevel Linear Model Approach

Introduction

Multifamily housing[3] plays a vital role in the U.S. real estate market. As a primary rental housing type, multifamily housing accommodated 17.8 million multifamily renter households (14.5% of total households) in the United States in 2013[4]. Meanwhile, multifamily housing is a common option for low- and medium-income households. Distinct from single-family housing, multifamily housing provides services and management characteristics. Services could include offerings such as pet allowance, maid service, storage service, special services for those with disabilities, or others. Management characteristics could include an option for a short-term lease, renovations, etc. As a focus of multifamily housing research and practice, measuring the effects of service and management on multifamily rents has important implications for optimizing management, setting rents, and attracting potential tenants. However, there is little research on the effects of service and management on rents.

The purpose of this paper is to appropriately estimate the market value (implicit

---

[3] The definition of a multifamily house varies by organization The standard industry definition of multifamily housing is a structure with five or more units. According to this definition, multifamily housing is also generally considered to be renter-occupied housing, while owner-occupied condominiums are0 usually not considered to be multifamily housing units even though they may be located in multifamily structures.
[4] The number of households in multifamily rental housing comes from Multifamilybz.com. The definition of multifamily housing for this calculation is a structure with five or more units that is renter-occupied.

price) of the effects of service and management within a hedonic framework. That is, I set transaction rent as the dependent variable, include indicators of service and management, structure characteristics, amenities, and locational characteristics, and conduct multilevel linear regression analyses. Using a recent and comprehensive micro-level dataset of multifamily housing units in Montgomery County, Maryland, I have conducted multilevel linear regressions to examine the effects of service and management on multifamily rents. My findings suggest the following: (1) Some service and management offerings – such as pet allowance, availability of short-term lease options, and storage service – increase rents significantly. Other service and management features – such as renovations and availability of service for those with disability – do not increase rents. (2) The multilevel linear model is an appropriate approach to account for both hierarchical residential location decisions and the hierarchical data structure. More specifically, the random coefficient model outperforms the random intercept model in terms of data fitness.

This study addresses the following research questions: (1) Does availability of service and management impact multifamily rents? (2) Should a multilevel linear model be used in multifamily hedonic analysis? If so, should one consider the random coefficient model or the random intercept model? This paper contributes to the literature in two aspects: one is to estimate the effects of service and management; another is to put forth a methodology.

This paper is organized as follows. The next section reviews existing literature assessing both the effects of service and management on multifamily rents and the application of a multilevel linear model on housing and real estate. After introducing the dataset and modeling approaches, I compare the results of the traditional Ordinary Least Square (OLS) model, the random intercept model, and the random coefficient model and, in particular, explain the results of the random coefficient model. The final section replicates the key findings and offers implications for the housing and real estate fields.

Literature review

In the last three decades, there has been a growing body of research focused on assessing multifamily housing characteristics and locational characteristics on multifamily rents (for reviews, see Jud, Benjamin, and Sirmans (1996) and Zietz (2003)). These studies use a hedonic approach to estimate implicit prices of complex attributes, as well as locational characteristics (Allen, Springer, & Waller, 1995; Asabere & Huffman, 1996; Babawale, Koleoso, & Otegbulu, 2012; Bogdon & Ling, 1998; Frew & Jud, 2003; Gipe, 1976; Guntermann & Norrbin, 1987; Hoch & Waddell, 1993; Hoesli, Thion, & Watkins, 1997; Jaffe & Bussa, 1977; Jud & Winkler, 1991; Kain & Quigley, 1970; Lin & Cheng, 2016; Marks, 1984; McMillan & Lee, 2017; Miller, 1982; Ogur, 1973; Portnov, Genkin, & Barzilay, 2009; Shear, 1983; Shenkel, 1975; Sirmans, Sirmans, & Benjamin, 1989; Smith & Knoll, 1988; Wilson & Frew, 2007). However, there are few existing studies that assess the effects of offered service and management on rents, despite the fact that service operated and management play a critical role in practice. The lack of related studies can be attributed to the absence of data related to micro-level service and management variables in multifamily housing analysis. The following examples serve as exceptions. Sirmans and Sirmans (1992) use professorial designations as a proxy for true quality of service provided by the property management and find that the holding of designations has a positive effect on monthly rent. Security service and management – such as gated access restrictions and 24-hour security – enhances rental rates and rents in both garden-style apartments (WG Hardin III & Cheng, 2003) and high-rise apartments (Benjamin, Sirmans, & Zietz, 1997). Multifamily properties owned and managed by real estate investment trusts (REIT) have higher effective rents than non-REIT-owned properties (William Hardin III, Hill, & Hopper, 2009). Regarding the relationship between property management and rent, Benjamin, Chinloy, and Hardin (2007) examine Atlanta apartments and find interesting results. Their findings show that larger-scale owners and local property managers earn higher effective rents (Benjamin et al., 2007). Pet allowance was found to have a positive effect on rent (WG Hardin III & Cheng, 2003; Sirmans et al., 1989). The effects of short-term renovations on apartment rents are small and non-significant (Mejia & Potter, 2015). The availability of maid service has no significant effect on rent (Sirmans et al., 1989). Age restrictions have a positive effect on the price of condominiums (Guntermann and Norrbin, 1987). It is worth noting that local apartment markets are segmented by unit type (Wolverton, Hardin, & Cheng, 1999). Wolverton et al. (1999) use the Chow-test and the Tiao-Goldberger test to confirm the existence of segregation of apartment markets. As such, it is not enough to simply run a hedonic regression for the aggregate dataset, and it is necessary to conduct a separate hedonic analysis for unit-type submarkets.

Regarding approaches applied within the hedonic framework, OLS is used widely and broadly throughout the existing housing literature. Most of the OLS models use semi-log or double-log specifications because heteroscedasticity is substantially reduced or eliminated by semi- or double-log models (Clapp & Salavei, 2010). To account for the hierarchical data structure – the data structure type most commonly referenced in the existing housing literature – there is a growing body of research that uses multilevel linear models. But, these studies focus on single-family housing (Brunauer, Lang, & Umlauf, 2013; Chasco & Gallo, 2013; Djurdjevic, Eugster, & Haase, 2008; Gelfand, Banerjee, Sirmans, Tu, & Ong, 2007; Giuliano, Gordon, Pan, & Park, 2010; Glaesener & Caruso, 2015; Goodman & Thibodeau, 1998; Jones & Bullen, 1993; Orford, 2002; Shin, Saginor, & Van Zandt, 2011; Treg, 2010; Uyar & Brown, 2007). It is surprising that there are still few studies that apply the multilevel linear model (MLM) for multifamily hedonic analysis, given the fact that dozens or even hundreds of units can be nested within a complex. The decision to ignore hierarchical data structures may underestimate standard errors and lead to false inferences (Krull & MacKinnon, 2001). Furthermore, current literature does not convey much about the kind of multilevel linear model that should be applied for multifamily hedonic analysis if a multilevel linear model is implemented. Specifically, should the random intercept model or the random coefficient model be used?

It is likely that rents for apartments that are geographically clustered together are correlated, since the apartments share the same location benefits and urban amenities. To account for spatial effects within the hedonic framework, spatial econometrics models could be applied. Anselin and Lozano-Gracia (2009) refer to these models as spatial hedonic models. By adding spatial lags of the dependent variable, spatial lags of independent variables, or spatial lags of the error term – or a combination of these variables (Anselin, 2013; Arbia, 2006; Elhorst, 2003) – spatial hedonic models offer more explanatory power than classic hedonic models. To estimate the spatial models, the maximum likelihood estimation (Lee, 2004), two stage least square (2SLS), generalized method of moment (GMM) (Kelejian & Prucha, 1998, 1999, 2004), or the Bayesian approach (LeSage & Pace, 2009) can be conducted. In the spatial panel context, Elhorst (Elhorst, 2003, 2014a, 2014b, 2014c) and Baltagi (2008) provide comprehensive applications. However, researchers have reached little consensus regarding which hierarchical spatial model should be used.. Even more, few software packages can be applied to estimate the hierarchical spatial models.

Overall, the gaps in assessing the effects of service and management on multifamily rents within a hedonic framework stem from the following: (1) Only a few studies examine the effects of service and management on multifamily rents, despite the fact that service and management are critical factors in optimizing management and setting rents; (2) Most multifamily hedonic analyses ignore the hierarchical data structure, which may result in false inferences; (3) There is limited information available to determine which specification of the multilevel linear model should be used in a multifamily housing context – the random coefficient model or the random intercept model.

This study aims to fill the gaps in the literature by (1) constructing an up-to-date comprehensive unit-level multifamily dataset with transaction rent, indicators of service and management, amenities, structure characteristics, and locational characteristics; and (2) using an appropriate approach – MLM – to assess the effects of service and management on multifamily rents. MLM can account for the hierarchical data structure, the most commonly existing data structure in the multifamily housing context.

Data

My empirical dataset is compiled based on four data sources. The first data source is the Montgomery County rental housing survey conducted by the Department of Housing and Community Affairs (DHCA) in 2018. I have created an interactive map for the survey data. Here is the web address: https://rpubs.com/xqpeng/RentalSurveyAnalysis. The survey includes 79.5% of 930 complexes throughout Montgomery County in the state of Maryland. The rental housing survey has a hierarchical structure with two levels. The lower level is the unit level, which has three variables – unit rent, number of bedrooms, and a dummy variable indicating if a unit is occupied by a property company employee. The unit rent is a transaction rent for which the payment is made to property managers or landlords. I exclude observations that are vacant because the rents of these observations are not market-driven. The number of bedrooms indicates how many bedrooms are in a unit. Employees of property companies may get a discount rent price if they live in property owned or managed by their companies. The dummy variable indicating if a unit is occupied by an employee is used to control this situation. The survey dataset's higher level represents the complex level, which features information regarding service and management, amenities, and structure characteristics. Using longitudinal and latitudinal information for complexes, I geocode the observed complexes to get locational characteristics because accessibility is a key determinant of property values and rent, as implied by the adage “Location, location, location.” To do this, I extract shapefiles of shopping centers, public schools, light rail transit stations, bus stops, highway routes, highway exits, and crime information in Montgomery County from the second data source, Data Montgomery. By spatially combining the location of multifamily complexes with shapefiles of locational characteristics of communities, my dataset establishes locational characteristics for multifamily complexes. To indicate the performance of public schools, I acquired access to the 2013 Maryland State Education Indicators, the third data source. The education indicators offer information on average math and reading scores for fifth-graders for each public elementary school in Montgomery County. The fourth data source is Open Data DC. In order to control for each complex’s distance from the central business district (CBD) of Washington, D.C., I extract the shapefile CBD of Washington, D.C. from Open Data DC and compute the distances.

After I cleaned the raw data by dropping duplicated entries, vacant units, and outliers, our final dataset included 73,094 units from 740 complexes. Figure 2.1 shows the locations of multifamily complexes in this study. Information about the observations is grouped into five categories: unit-level attributes, service and management information, complex structure characteristics, complex amenities, and locational characteristics. Table 2.1 lists variables included in the analysis as they correspond with each category. For each observation, variables associated with unit-level attributes include transaction rent, number of bedrooms, and a dummy variable indicating if the unit is employee-occupied. It should be noted that I have a unit-level service and management variable indicating if a unit is employee-occupied. Complex-level service and management variables contain the following: a dummy variable indicating if the complex allows pets in the building; a dummy variable indicating if renovations had been made in the prior 12 months; a dummy variable indicating whether the complex offers short-term leasing options; a dummy variable indicating if the complex offers a storage service; and a dummy variable indicating if the complex provides special services for those with disabilities.

The complex-level structure attributes include: a continuous variable indicating the age of the structure; a category variable indicating the structure type for the complex (for which the base group is “townhouse”); and the total number of units for the complex. The complex-level amenities include: a dummy variable representing if the complex has a gym center; a dummy variable indicating if the complex provides a common laundry room; a dummy variable indicating if the complex has a parking lot or garage; and a dummy variable indicating if the complex has a swimming pool. Variables for locational characteristics include school quality, which uses average reading scores for fifth-graders for each school to indicate school quality; crime rate, which describes the correlation of the number of properties to violent crimes reported to law enforcement agencies per 100,000 total population at the census tract level in 2017; a dummy variable indicating if a complex is within one mile from its nearest light rail station; the distance between the complex and the nearest light rail transit station; the distance from highways measured in two bands up to 0.5 mile; distances from highway exits to complexes, measured in 0.25 mile bands, up to two miles; the distance between the complex and the nearest shopping center; and the distance between the complex and the CBD of Washington, D.C. All distances were measured in Euclidean terms.

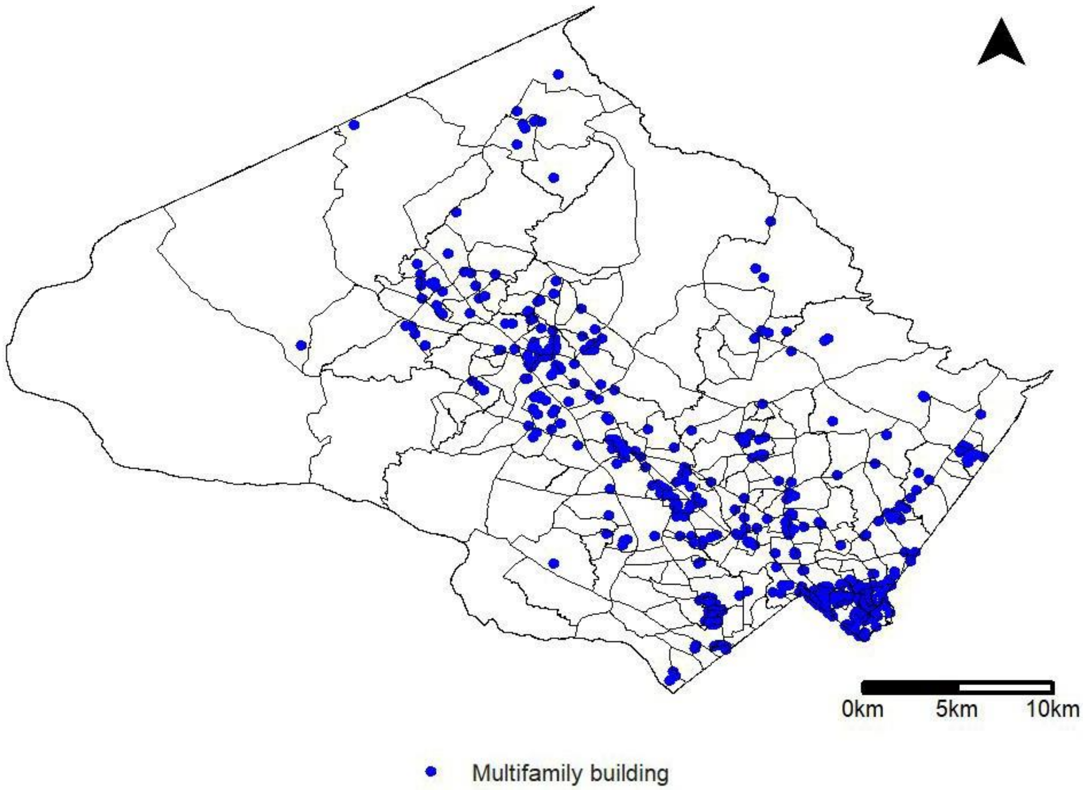

Figure 2.1: Location of multifamily complexes in the study

Table 2.1: Definition of variables

Category	Variable Name	Definition
1. Unit-level attributes	Rent	Monthly rent of unit
	Bedrooms	Number of bedrooms of a unit
	EmployeeOccupied	If a unit is occupied, by property employee, assign 1. Otherwise,0.
2.Complex-level characteristics 2.1 Service and management	**Pets**	If a complex allows pets, assign 1. Otherwise, 0.
	ShortLease	If a complex provides a short-term lease contract (fewer than 12 months) for tenants, assign 1. Otherwise, 0.
	Renovation	If a complex has implemented a renovation in the last 12 months, assign 1. Otherwise, 0.
	Storage	If a complex provides storage service, assign 1. Otherwise, 0.

	Disability	If a complex provides service for those with a disability, such as by offering specially designed rooms for those with a disability, assign 1. Otherwise,0.
2.2 Complex amenities	Gym	If there is a gym center, assign 1. Otherwise, 0.
	Laundry	If there is a laundry room, assign 1. Otherwise, 0.
	Parking	If a complex provides a parking lot or garage, assign 1. Otherwise, 0.
	Swimming	If a complex has a swimming pool, assign 1. Otherwise, 0.
2.3 Structure attributes	Age	Age of building structure
	Structure	Structure type of building. If it is a garden apartment with 1-4 stories, assign 1. If it is a midrise apartment with 5-8 stories, assign 2. If it is a high-rise apartment with 9+ stories, assign 3. If it is a townhouse, assign 0.
	ComplexSize	Number of units in a complex
2.4 Locational characteristics	School quality	Use average fifth-grade reading score per school as indicator of the quality of elementary schools
	Crime	Crime rate at the census tract level (2017)
	Metro_1 mile	If a complex is within one mile from a light rail transit station, assign 1. Otherwise, 0.
	DisMetro	Distance between a complex and its nearest light rail transit station.
	Bus_0.25 mile	<0.25 mile from bus stop (dummy).
	Highway_0.25mile	<0.25 mile from highway (dummy).
	Highway_0.5 mile	0.25-0.5 mile from highway (dummy).
	Exit_0.25 mile	<0.25 mile from highway exit (dummy).
	Exit_0.5 mile	0.25-0.5 mile from highway exit (dummy).
	Exit_0.75 mile	0.5-0.75 mile from highway exit (dummy).
	Exit_1 mile	0.75-1 mile from highway exit (dummy).
	Exit_1.25 mile	1-1.25 miles from highway exit (dummy).
	Exit_1.5 mile	1.25-1.5 miles from highway exit (dummy).
	Exit_1.75 mile	1.5-1.75 miles from highway exit (dummy).
	Exit_2 mile	1.75-2 miles from highway exit (dummy).
	DisShopping	Distance between complex and its nearest shopping center.

	DisCBD	Distance between complex and the CBD of Washington, D.C.

Note: Names in bold are service and management variables.

The dataset I construct presents four merits and one possible weakness compared with previous multifamily housing studies. First, the dataset includes unit-level attributes for multifamily housing, which was seldom reported in previous studies. The unit-level data usually contains more detailed information, and therefore, the estimation of implicit price could be more accurate based on that data. Second, our dataset has an appropriate indicator for the market value of rental units, which is the transaction rent. As the indicator of the market value of rental units, transaction rent is better than its alternatives, such as asking rent, appraised rent, and assessment rent. Such alternative indicators of rent are not often market-determined and, therefore, are subject to inconsistencies inherent to non-market valuation techniques. Third, the dataset includes detailed service and management information, which makes the dataset unique to the literature. Both researchers and practitioners are consistently interested in the effects of various services and management factors on multifamily rents but are not able to access these variables. Fourth, the dataset is based on the latest, population-scale multifamily surveys and has a large sample size. The dataset has a possible weakness, however. There are only two unit-level explanatory variables: the number of bedrooms and a dummy variable indicating if a unit is occupied by a property company employee. This may lead us to an omitted-variable bias problem. For example, the dataset is missing the size of the unit in terms of square footage. While our dataset offers information on the number of bedrooms for each unit, this may not be an efficient proxy for square footage; but, I believe it is an acceptable proxy for unit size.

The summary statistics for each variable are reported in Table 2.2. The average transaction rent is $1,573 per month. The average number of bedrooms is 1.57, while 0.8% of units are employee-occupied. Regarding complex-level service and management characteristics, 77.5% of units are nested in complexes that allow pets in their complexes; 46.2% of units are nested in complexes that offer short-term lease options for tenants; 8.2% of units are nested in complexes that have made observable renovations in the prior 12 months; 40.2% of units are nested in complexes that provide storage services; and 13.4% of units are nested in complexes that offer service for those with disabilities. Regarding structure attributes, the average age of the complex structures is 37.6 years. The average complex size is 308 units. Furthermore, complex-level amenities include 60.6% of units nested in complexes that have gym centers, while 55.5% of units are nested in complexes that have laundry rooms, 88.3% of units are nested in complexes that have parking lots or parking garages, and 64.8% of units are nested in complexes that have swimming pools. With regards to locational characteristics, we can see that the average fifth-grade reading score is 91.7 (out of 100), while the average crime rate is 9,751 incidents per 100,000 people a year. The average distance to the nearest rail transit station is 1.984 miles while the average distance to the nearest shopping center is 0.353 miles. The average distance to the CBD of Washington, D.C. is 12.15 miles.

Table 2.2: Summary statistics

Sub-category	Variable Name	Mean	Std.Dev.	Min	Max
1. Unit-level attributes	Rent	$1573	$458.41	$50	$3673

	Bedrooms	1.57	0.729	0	5
	EmployeeOccupied	0.008	0.089	0	1
2. Complex-level characteristics 2.1 Service and management	**Pets**	0.775	0.417	0	1
	ShortLease	0.462	0.498	0	1
	Renovation	0.082	0.274	0	1
	Storage	0.402	0.490	0	1
	Disability	0.134	0.340	0	1
2.2 Complex amenities	Gym	0.606	0.488	0	1
	Laundry	0.555	0.496	0	1
	Parking	0.883	0.321	0	1
	Swimming	0.648	0.477	0	1
2.3 Structure attributes	Age	37.6	21.005	1	133
	Structure	Category variable Number of units in garden apartment with 1-4 stories: 43145 Number of units in midrise apartment with 5-8 stories: 3066 Number of units in high-rise apartment with 9+ stories: 24176 Number of units townhouse: 2707			
	ComplexSize	308	220.2	1	1067
2.4 Locational characteristics	School quality	91.7	6.102	73.0	100.0
	Crime	9751	12028	953	66161
	Metro_1 mile	0.317	0.465	0	1
	DisMetro (in mile)	1.9842	2.098	0.021	13.768
	Bus_0.25 mile	0.579	0.493	0	1
	Highway_0.25mile	0.057	0.231	0	1
	Highway_0.5 mile	0.0868	0.281	0	1

	Exit_0.25 mile	0.0304	0.171	0	1
	Exit_0.5 mile	0.0974	0.296	0	1
	Exit_0.75 mile	0.0691	0.253	0	1
	Exit_1 mile	0.0732	0.260	0	1
	Exit_1.25 mile	0.166	0.372	0	1
	Exit_1.5 mile	0.118	0.322	0	1
	Exit_1.75 mile	0.178	0.382	0	1
	Exit_2 mile	0.0639	0.244	0	1
	DisShopping (in mile)	0.3536	0.283	0.013	4.489
	DisCBD (in mile)	12.15	5.467	5.19	30.73

Note: Variables in bold are service and management variables. 73094 units in 740 complexes.

Econometric Methods

Merits of the multilevel linear model

As a valuation technique, hedonic price modeling has been used broadly and has been noted to generate fruitful results in housing literature since Lancaster (1966) and Rosen (1974) began to explore the determinants of housing prices. Most hedonic models cited in the housing literature are one-level OLS regression models; however, residential location decisions are inherently hierarchical. The search process for a house begins with choosing a town or city to live in, followed by a neighborhood, and finally, a house, given the neighborhood and the town (Quigley, 1985). This search process could happen within a multifamily context. Moreover, it is reasonable to assume that individual units in the same complex will have correlated responses on rent since they share the same location, amenities, services, and management. In other words, the multifamily dataset is hierarchical in structure – individual units are nested within complexes. The one-level OLS assumes that all observations are independent. It is apparent that the one-level OLS based on multifamily hierarchical dataset violates the assumption, and the one-level OLS will yield biased results. To account for the hierarchical search process and multifamily hierarchical data structure, it is more appropriate to implement a multilevel linear model (MLM) (also known as a hierarchical linear model, or a mixed model) in the multifamily hedonic analysis. The MLM approach has the following benefits compared with a one-level OLS when the dataset is hierarchical.

1.) One-level OLS will underestimate the standard error, and thus overestimate test statistics and the statistical significance of the parameters; this can result in spuriously significant effects (Krull & MacKinnon, 2001). Column 1 of Table 3 reports the results of the one-level OLS regression. We can see that nearly all coefficients of explanatory variables are statistically significant, which differs from the results of the MLMs in Column 2 and Column 3. This is spuriously significant.

2.) The MLM approach allows parameters for the model to vary across space. Thus, both spatial correlation and heterogeneity can be taken into account within the MLM framework (Djurdjevic et al., 2008).

3.) The MLM specification can help in examining the variability of the coefficient across groups, and can be useful in examining cross-level interaction.

The MLMs in our case have two levels[5] . Level 1 is the unit level while Level 2 is the complex level. It is worth noting that conventional spatial hedonic model cannot be applied here because hundreds of units nests within one complex and weight matrix based on geographical location would be singular. MLM can take many forms depending on which predictors are included at each level, and whether the model accounts for a random intercept, or a random coefficient, or both. Let's start with the simplest MLM model: no explanatory variables at any level. Raudenbush and Bryk (2002) refer to it as the "unconditional model."

Unconditional model

The first level of the unconditional two-level model is:

$$y_{ij} = \beta_j^0 + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2) \qquad (1)$$

where the ij subscript refers to the ith individual unit in the jth complex; y_{ij} is the natural log of rent of unit i in complex j. β_j^0 is the varying intercept across complex j. ε_{ij} is a stochastic error term following a normal distribution with zero mean and variance of σ^2.

The second level of the two-level model assumes the coefficient β_{0j} can be expressed as:

$$\beta_j^0 = \gamma^0 + \mu_j, \mu_j \sim N(0, \tau^2) \qquad (2)$$

[5] Three-level MLM could be an alternative: Level 1 (unit level), Level 2(complex level), and Level 3 (neighborhood level). Level 3 could include a few explanatory variables. Two of them are school quality and census tract crime rate. The school district boundaries do not entirely contain boundaries of certain census tracts, which may complicate the analysis when I applied a three-level MLM. For simplicity and feasibility, this study utilizes a two-level MLM.

where γ^0 indicates average intercept across complexes; μ_j is a complex- specific effect on the intercept. μ_j is assumed to have a variance of τ^2 and uncorrelated with ε_{ij} . After substituting Equation (2) into Equation (1), I have a combined equation:

$$y_{ij} = \gamma^0 + \mu_j + \varepsilon_{ij}, Cov(\varepsilon_{ij}, \mu_j) = 0 \quad (3)$$

while the simple (intercept-only) multilevel model does not include independent variables in the regression, it includes important information regarding how variations in y_{ij} are partitioned between variance among the individual units and variance among the complexes. Regarding the estimation method for MLMs, the maximum likelihood estimation (MLE), the restricted maximum likelihood (REML), or the Bayesian can be used in place of standard OLS, because the random errors for the combined equation (3) are neither independent nor have a constant variance. MLE is an iterative methodology in which the algorithm searches for parameter values that will maximize the likelihood of the observed data (Raudenbush & Bryk, 2002). REML goes a step further and corrects the estimate of the variance by taking an appropriate degree of freedom into account, which has proven more accurate than MLE in estimating variance parameters (Kreft & De Leeuw, 1998). In this study, I apply REML; the difference in value for MLE and REML estimates becomes very small as the number of Level 2 clusters increases (Snijders & Bosker, 1999), as is the case in this study (816 complexes). To fit multilevel linear models in this study, I utilize an R library, nlme (Team, 2013), and apply REML.

The “intraclass correlation coefficient” (Raudenbush and Bryk, 2002) is a measure of the proportion of variations in the outcome variable that are attributable to differences at the group level in the MLM literature. The intraclass correlation coefficient (ICC) is computed as follows:

$$ICC = \tau^2/(\tau^2 + \sigma^2) \tag{4}$$

For this case, ICC is 0.594 suggesting 59.4% of the variance in rents due to differences at the complex level. This further justifies the idea that the multilevel linear model approach should be applied in this study.

Random intercept model

After determining that the multilevel linear model should be used, it is reasonable to include variables for both levels. The random intercept model allows the intercept to vary across complexes, and includes all the explanatory variables[6], but it makes the coefficients of explanatory variables fixed. Then, at the first level

$$y_{ij} = \beta_j^0 + \sum_{k=1} \beta^k x_{ij}^k + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2) \tag{5}$$

where x_{ij}^k is the kth independent variables at the first level; β^k is the coefficient of x_{ij}^k. In this study, I only have two explanatory variables at the unit level: the number of bedrooms and a dummy variable indicating if a unit is occupied by a property company employee. This means k=2 in this case. Notice that β^k does not vary across complex j. At the second level

$$\beta_j^0 = \gamma^0 + \sum_{l=1} \omega^{0l} z_j^{0l} + \mu_j, \mu_j \sim N(0, \tau^2) \tag{6}$$

where *l* indicates the number of complex-level independent variables; ω^{0l} is the

[6] I tried to include an explanatory indicating distant to open space, and an explanatory variable indicating distance to elementary school in the model. However, VIF values of these explanatory variables are larger than 10, which suggests it may have multicollinearity problem if I include them into the model. As such, I do not include these two explanatory variables into the models.

coefficient indicating the fixed effect of the *l*th independent variable, z_j^{0l} . After substituting Equation (6) into Equation (5), the random intercept model in combined form is:

$$y_{ij} = \gamma^0 + \sum_{k=1} \beta^k x_{ij}^k + \sum_{l=1} \omega^{0l} z_j^{0l} + (\mu_j + \varepsilon_{ij}), Cov(\varepsilon_{ij}, \mu_j) = 0 \qquad (7)$$

The random components of the model have two parts: ε_{ij} and μ_j ; others are fixed components.

Random coefficient model

There is no reason that the coefficients of the first-level independent variables must remain constant. For example, there may exist a unique *Bedrooms* effect on rents for complex j in this case. Likewise, there may exist a unique Employee-Occupied effect for complex j. The random coefficient model considers that the coefficients of unit-level independent variables – as well as the intercept – can vary across complexes. At the first level,

$$y_{ij} = \beta_j^0 + \sum_{k=1} \beta_j^k x_{ij}^k + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2) \qquad (8)$$

where β_j^k is the random coefficient of unit-level independent variables, x_{ij}^k . At the second level, the random intercept (β_{0j}) and the random coefficients (β_j^k) can be expressed as the following, respectively:

$$\beta_j^0 = \gamma^0 + \sum_{l=1} \omega^{0l} z_j^{0l} + \mu_j, \mu_j \sim N(0, \tau^2) \qquad (9)$$

$$\beta_j^k = \gamma^k + \mu_j^k, \mu_j^k \sim N(0, \tau_k^2) \qquad (10)$$

After substituting Equation (9) and Equation (10) into Equation (8), I have the combined form of the random coefficient model:

$$y_{ij} = \gamma^0 + \sum_{k=1} \gamma^k x_{ij}^k + \sum_{l=1} \omega^{0l} z_j^{0l} + (\sum_{k=1} \mu_j^k x_{ij}^k + \mu_j + \varepsilon_{ij}), Cov(\varepsilon_{ij}, \mu_j^k) = 0 \qquad (11)$$

The random components of the model have three parts: $\sum_{k=1} \mu_j^k x_{ij}^k$, ε_{ij} and μ_j; others are fixed components.

Empirical Results

I started with reporting one-level OLS hedonic regression results (see Column 1 in Table 2.3). Nearly all the coefficients are statistically significant at 99% and most of them have expected signs. However, the one-level OLS regression could lead to a substantial underestimation of the standard errors when the data structure is hierarchical; as such, the results are spuriously significant (Krull and MacKinnon, 2001). Based on the suspect standard error estimation, the inference would not be reliable.

To account for the hierarchical data structure, I tried two primary types of multilevel linear models: the random intercept model and the random coefficient model. The only difference between the two kinds of models is that the random coefficient model allows the intercept and the coefficients of the unit-level independent variables to vary across complexes, while the random intercept model only allows the intercept to vary across complexes. I ran Equation (7) and reported results for the random intercept model in Column 2 of Table 2.3. I also ran Equation (11) and reported results for the random coefficient model in Column 3 of Table 2.3. With these two candidate models, the question that arises is: which model fits the dataset better – the random intercept model or the random coefficient model?

I do Hausman test to check if it is necessary to allow unit-level variables, variables of Bedrooms and EmployeeOccupied in this essay, vary across complexes.

A Hausman test testifies if a random estimator should be used and would be efficient. Regarding variable Bedrooms, the statistic value of Hausman test is 0.833, which cannot reject the null hypothesis. This means that both a random estimator and a fixed estimator are consist, but the random estimator is more efficient. So I should do random estimator for variable Bedrooms. Regarding variable EmployeeOccupied, the statistic value of Hausman test is 0.737, which cannot reject the null hypothesis. This means that both a random estimator and a fixed estimator are consist, and the random estimator is more efficient. So I should do random estimator for variable EmployeeOccupied as well. In short words, the Hausman tests suggest that I should choose the random coefficient model.

The AIC (-53007) of the random coefficient model is smaller than that of the random intercept model (-48535), which indicates that the random coefficient model fits the dataset better than the random intercept model. The likelihood ratio statistic (4482.2) justifies the conclusion that the random coefficient model is a better fit than the random intercept model. Additionally, the conditional R^2 of the random coefficient model is 84.6% – larger than that of the random intercept model (76.8%) – which further affirms that the random coefficient model fits the dataset better than the random intercept model[7].

[7] R^2 is frequently used indicator for comparing models in terms of model fitness. However, Nakagawa and Schielzeth (2013) suggest that using R^2 from traditional OLS linear model for MLMs is misleading and should not be used. There are multiple ideas how to compute R^2 for MLMs, for example, pseudo-R2, marginal R2, and conditional R2, but there is no consensus on it. In this study, I also use marginal R2 and conditional R2 to compare MLMs in terms of modeling fitting. Marginal R2 describes the proportion of variance explained by the fixed factor(s) alone, while conditional R2 describes the proportion of variance explained by both the fixed and random factors (Johnson, 2014; Nakagawa & Schielzeth, 2013). It is worth noting that these marginal R^2 and conditional R^2 are not comparable with R^2 in traditional OLS context.

Next, I report the results of the random coefficient model (Column 3 of Table 2.3) in detail. The natural logarithm of unit rent is the dependent variable. The multilevel linear model accounts for a respectable 84.6% of the variation in log-rents. The results show that the number of bedrooms positively and significantly impacts unit rent. These findings echo those of previous studies (Babawale et al., 2012; Guntermann and Norrbin, 1987; I.Hoch and Waddell, 1993; Sirmans et al., 1989; B.Wilson and J.Frew, 2007). More specifically, with the addition of one bedroom, unit rent increases by 18.42% on average in this study.

Next, Let's turn to the service and management variables of greatest interest in this study. For the unit-level service and management variable a dummy variable indicating that a unit is property employee-occupied is used, the results show that rent is 9.63% less than a unit that is not employee-occupied. As for the effects of complex-level service and management, pet allowance increases multifamily rent by a magnitude of 5.63%, which echoes previous studies (WG Hardin III & Cheng, 2003; Sirmans et al., 1989). Permitting pets in a rental property has several benefits for property-owners. This offering attracts more tenants interested in renting property, and landlords have a fantastic opportunity to make more money by charging pet rent. Also, pet owners may be more likely to renew their lease and they may tend to be responsible tenants. The results also show that offering short lease options is statistically significant and positively associated with rent. This means that the average rent of a complex that offers short-term leasing contracts is higher than the average rent of a complex that does not provide short-term leasing contracts for tenants. The study also shows that short-term renovating complexes does not increase rents significantly although the renovation can be seen by tenants, which echoes the findings of Mejia and Potter (2015). The observable renovations referenced here include new paint, new carpet, yard landscape, etc. in the prior 12 month. Landlords understand that these renovations are necessary repairs or replacements that make their property nicer than the cheapest generic rental units available to the public. As such, landlords make the renovations but may not increase rents. It is worth noting that we do not has information of how much expenditure and magnitude of renovations the complexes spend. Major renovation could have positive effects on rents and further studies are needed. This study also shows that storage service offerings increase multifamily rent significantly, to the magnitude of 8.57%. Regarding availability of disability service, the effect is negatively significant. I am not aware of any previous studies that investigate the effects of employee occupied, short-term leasing contract options and storage service offerings on multifamily rent. Note that the results may suffer omitted variable bias problem although the conditional R-squared of the random coefficient model is more than 0.8. For example, whether a unit has wood floor, walking-in closet, balcony, etc. I tried to include more explanatory variables affecting rents into the regression. However, I am not able to access that kind of data.

Moving to complex characteristics, multifamily rents significantly decline as the structure ages. This result is consistent with the literature (Allen et al., 1995; J.Frew and Jud, 2003; I.Hoch and Waddell, 1993; G.D.Jud and Winkler, 1991; J.F.Kain and Quigley, 1970; J.J.Lin and C.Cheng, 2016; Sirmans et al., 1989). In our case, each 1% increase in age reduces rent by 0.039%. With regards to the effect of size – as the complex size increases, the rent of the units, on average, increases significantly. Regarding complex-level amenities, having a swimming pool increases rent significantly; this is consistent with the results put forth by previous studies (Guntermann and Norrbin, 1987; I.Hoch and Waddell, 1993; G.D.Jud and Winkler, 1991; Sirmans et al., 1989; B.Wilson and J.Frew, 2007). Regarding the presence of laundry rooms, the estimated coefficient is positive but not statistically significant. Some tenants may favor the common-area laundry room while others do not. The presence of a common-area laundry room implies the lack of in-unit laundry machines, which is a disamenity. The preference dichotomy makes the effect of the laundry room statistically insignificant in this study. The coefficient of parking availability is not statistically significant. In this study, 88.3% units have parking lot or garage. For those that do not have parking lot or garage, they could have no difficulty to park their vehicles along roads, which is significant different from Eastern Asian cities and Western European cities.

Regarding locational characteristics, units in close proximity to light rail transit stations and shopping centers have higher rents than units located further from these urban amenities. For example, units that are within one mile from their nearest light rail transit stations are expected to see a rent increase of 13.15% than other units. Furthermore, I observe that the closer proximity to light rail stations results in higher rent; however, proximity to bus stops does not increase unit rent significantly. The results show that two coefficients on the dummy variables for the distance bands from highways have negative signs as hypothesized, though none are even close to being significant. Thus, disamenities such as noise – represented by distance from highways

– appear to have no significant effect on multifamily rents. Seo et al (2014) have similar results regarding the effects of proximity to highway routes on property value. Accessibility effects, on the other hand, are very significant at the 0.05 level as complexes are located upwards of 0.75 miles away but fewer than two miles away from highway exits. Units located in higher-performing public elementary school districts have higher rents than units that are not. This finding echoes the results of J.F.Kain and Quigley, 1970. The crime rate of a neighborhood is also an important factor in determining rents. The coefficient of crime rate is negative and statistically significant, which means that safer neighborhoods equate to higher unit rent. This result is consistent with the results of a study put forth by J.F.Kain and Quigley (1970). In addition to the estimates of fixed effects, the estimates of random effects are also listed in Table 2.3.

In summary, the random coefficient model results show that controlling unit, complex, neighborhood, and locational characteristics, the effects of service and management – such as availability of short-term leasing options, pet allowance, and storage service offerings – are significant in increasing rents. Conversely, employee occupancy and services for those with disabilities decrease rents significantly. Short-term renovations do not increase rent significantly.

Table 2.3: Regression Results

	Model 1 One level standard linear hedonic model (OLS)		Model 2 Random intercept model		Model 3 Random coefficient model	
Fixed eff.	Coef.	Std.Error	Coef.	Std.Error	Coef.	Std.Error
Level 1 (unit level)						
Intercept	8.5959***	0.0995	7.9334***	0.8262	7.6866***	0.8212
Bedrooms	0.1822 *** *	0.0013	0.1806***	0.0010	0.1842***	0.0040

EmployeeOccupied	-0.0496***	0.0104	-0.0763***	0.0071	-0.0963***	0.0243
Level 2 (complex level)						
Pets	-0.0041	0.0026	0.0450*	0.0198	0.0441*	0.0196
ShortLease	0.0944***	0.0021	0.0635**	0.0229	0.0664**	0.0224
Renovation	0.0118**	0.0036	0.0031	0.0322	-0.0039	0.0317
Storage	0.0988 ***	0.0023	0.0843***	0.0227	0.0857***	0.0223
Disability	-0.0555***	0.0028	-0.0685*	0.0321	-0.0692*	0.0313
Gym	0.0278***	0.0027	0.0637*	0.0312	0.0338	0.0304
Laundry	-0.0270***	0.0022	0.0152	0.0186	0.0029	0.0184
Parking	0.0257***	0.0033	-0.0310	0.0215	-0.0329	0.0215
Swimming	0.1020***	0.0027	0.1320***	0.0304	0.1407***	0.0296
Log(Age)	-0.0662***	0.0014	-0.0535***	0.0150	-0.0390**	0.0146
ComplexSize	0.0366***	0.0013	0.0164*	0.0075	0.0218**	0.0075
Log(School)	0.2333***	0.0163	0.2122	0.1360	0.2312^	0.1348
Log(Crime)	-0.0353***	0.0014	-0.0337*	0.0144	-0.0316*	0.0141
Metro_1 mile	0.0696***	0.0042	0.1287**	0.0394	0.1315***	0.0387
Metro_1 mile:DisMetro	-0.1224***	0.0031	-0.1471***	0.0250	-0.1492***	0.0250
Bus_0.25 mile	0.0009	0.0022	-0.0197	0.0206	-0.0164	0.0203
Highway_0.25mile	-0.0866***	0.0097	-0.0811	0.1101	-0.0757	0.1071
Highway_0.5 mile	-0.0575***	0.0073	-0.0286	0.0890	-0.0340	0.0864
Exit_0.25 mile	0.0893***	0.0119	0.0157	0.1375	0.0012	0.1347
Exit_0.5 mile	0.1089***	0.0084	0.0881	0.0978	0.1121	0.0952
Exit_0.75 mile	0.0652***	0.0050	0.0430	0.0535	0.0584	0.0529
Exit_1 mile	0.1411***	0.0048	0.1259**	0.0432	0.1256**	0.0427
Exit_1.25 mile	0.0648***	0.0038	0.0955*	0.0377	0.0900*	0.0374
Exit_1.5 mile	0.1351***	0.0040	0.1576***	0.0354	0.1557***	0.0350
Exit_1.75 mile	0.0537***	0.0034	0.0822*	0.0334	0.0891**	0.0329
Exit_2 mile	0.1195***	0.0045	0.1513***	0.0376	0.1576***	0.0370
Log(DisShopping)	-0.0116***	0.0013	-0.0236^	0.0123	-0.0230^	0.0122
Log(DisCBD)	-0.2226***	0.0042	-0.1479***	0.0370	-0.1408***	0.0367
Control structure type	Yes		Yes		Yes	
Random eff.						
Level 1 (unit level)						
σ^2 (Residual)	——		0.170		0.163	
Level 2 (complex level)						

τ^2 (intercept)	——	0.223	0.225
τ_1^2 (Bedrooms)	——	——	0.079
τ_2^2 (EmployeeOccupied)	——	——	0.318
AIC	5595.3	-48535	-53007
Log likelihood	-2762.6	24303	26545
Likelihood Ratio test	——	——	4482.2 (Model 2 is the base)
Number of Obs.	73094	73094 at unit level; 740 at complex level	73094 at unit level; 740 at complex level
R^2	0.423	Conditional R^2: 0.768	Conditional R^2: 0.846

Notes: logarithm of rent for each unit as dependent variable.

*** p=.001(two-tailed). **p=0.01(two-tailed). *p=0.05(two-tailed). ^p=0.10(two-tailed).

Conclusion

Three main findings stand out in this study. First, the results show that the effects of service and management factors on multifamily rents vary across types of service and management. Pet allowance, availability of short-term leasing options, and storage service availability increase rents significantly, while short-term renovations do not increase rents significantly. Conversely, offering units to property employees and services to those with a disability decrease rents significantly. Second, I show that a multilevel linear model should be applied when conducting multifamily housing hedonic analysis. A hierarchical data structure is commonly used in the multifamily context, where hundreds and, at times, thousands of units are nested within one complex. Ignoring this type of hierarchical data structure and running only a one-level OLS model would result in spuriously significant effects and, thus, false inference. Third, I show that, in terms of model fitness, the random coefficient model is more appropriate than the random intercept model for multifamily hedonic analysis.

Service and management variables are determinants of multifamily rents. Investigation into the effects of service and management variables yields important implications for multifamily housing landlords/managers in optimizing management, setting rents, and providing myriad services. In addition, both the hierarchical residential location decisions and the hierarchical data structure are commonly viewed in the multifamily housing context. The multilevel linear model should thus earn more attention for conducting multifamily hedonic analysis.

3. Do Multifamily Unit Rents Increase in Response to Planned Light Rail Construction?

Introduction

Homebuyers are willing to pay more for a single-family home under the assumption that they will benefit from greater transit access at some future point in time. However, little is known about how renters and rental property owners and managers will respond to the same exogenous development. Renters are generally not concerned with the sale value of their units. In addition, while renters may be willing to spend more on monthly rent if it means greater access to public transit, it is unclear if they would be willing to pay more in expectation of that benefit. This essay is the first paper in the literature to empirically test whether light rail transit (LRT) in the pre-service period impacts multifamily unit rents.

Scholars have not yet empirically examined whether LRT investment impacts multifamily rents. Nevertheless, multifamily housing accounts for a large share of the American real-estate market and LRT is a popular mass transportation investment made in metropolitan areas. The lack of research is mainly due to the dearth of panel data from multifamily rent surveys. Recent work has tentatively pinned some causality of gentrification and displacement on public investment, and in particular, investment in public transportation Zuk et al. (2018). Federal and state governments invest huge sums of dollars into building LRT, and that investment may increase rents. Tenants in multifamily housing could potentially suffer from housing affordability problems even before the LRT is in service. If that is the case, preemptive actions could be taken to mitigate any negative effects.

This study addresses the following research questions: (1) Do multifamily unit rents increase in response to planned light rail construction? (2) If LRT in the pre-service period increases multifamily rents, by what magnitude does it do so?

Using the ongoing construction of the Purple Line in the Washington Metropolitan Area as a case study, coupled with new data derived from high-quality multifamily landlord rental surveys (two-year panel), I applied first-difference methods and difference-in-difference methods. The results indicate that the rents of two-bedroom, three-bedroom, and four-bedroom units within a half mile of planned light rail stops have significantly increased from 2015 to 2018 compared with the rent of units in other areas of the county. I also used parameterized distance decay functions to estimate the distance price decay effect at increasing distances from the stations. The results show that there is no significant distance price decay effect in 2015, but there was a significant effect in 2018, during which time the average slope of distance variable is 7.7%.

I proceed with this study as follows. In the next section, I review the results of previous research. Since there is no research regarding how LRT in its pre-service period impacts multifamily rents, I extend the review to the impact of LRT in the pre-service period on prices of a single-family property. In the next section, I introduce the data and describe the models and methods used. I then take two different approaches to test the effect: the first-difference approach and the difference-in-difference approach. In addition, I use parameterized distance decay regressions to estimate the magnitude of the distance decay effect. The final section reports results. The conclusion section discusses the possible reasons for the results and the policy implications.

Literature review

Many studies have found that light rail transit investments have impacts on the value of residential land, housing prices, and multifamily rents after in station areas both before after transit service begins. However, there has been no study examining whether new LRT projects in the pre-service period impact multifamily rents, perhaps because data on multifamily rents are scarce (Table 3.1). In this section I review, in order, findings on the impact of transit on land and housing prices in the post-service period, the impact of transit on rents in the post-service period, and the impact of transit on land and housing prices in the pre-service period.

Table 3.1: Studies on the impacts of light rail transit in the pre-service and post service period on property prices and rents

Measurement period	Land and housing prices	Housing rents
Pre-service	Agostini and Palmucci (2008) + Boucq and Papon (2008) - or minimal Cao and Lou (2017) + Damm et al. (1980) + Devaux et al. (2017) + Dubé et al. (2018) + Gatzlaff and Smith (1993) - or minimal Henneberry (1998) - or minimal Knaap et al. (2001) + Loomis et al. (2012) - or minimal McDonald and Osuji (1995) + McMillen and McDonald (2004) + Yan et al. (2012) - or minimal	No studies

	Yen et al. (2018)	+	
Post-service	Al-Mosaind et al. (1993)	+	Bollinger et al. (1998)
	Bae et al. (2003)	+	+
	Cervero and Duncan (2002)	+	Hass-Klau et al. (2004)
	Chen et al. (1998)	+	+
	Clower and Weinstein (2002)	+	John (1996)
	Devaux et al. (2017)	+	+
	Diao (2015)	+	Nelson et al.(2015)
	Du and Mulley (2007)	+	+
	Dueker and Bianco (1999)	+	Weinberger (2001)
	Duncan (2008)	+	+
	Duncan (2011)	+	Cervero (1994)
	Hass-Klau et al. (2004)	+	+
	Weinstein and Clower (1999)	+	

Note: + indicates studies that find a positive effect of transit on property value, a "– or minimal" indicates studies that have a negative or minimal effect.

A large number of studies have examined the effects of LRT service on property values and rents in the post service period. These studies have generally found that LRT service increases residential land values or single-family home prices near transit stations. For a recent review, see Knowles and Ferbrache (2016). In a Boston case study, Diao (2015) found that land value increases after accounting for sample selection and spatial autocorrelation. However, the effect of LRT service on residential property values may vary by geography and by land use regulations in those geographies. Mohammad et al. (2013) reviewed examples from the United States and found that eight studies showed an increase in residential sales prices, one study found a decrease in sales prices, and three studies produced results ranging from a negative to positive effect. For example, Chen et al. (1998) found that light rail has both a positive effect and a negative effect on single-family home values. Using San Diego as a case study, Duncan (2011) found a premium value associated with station proximity in the post-service period, which was conditional on permissive zoning regulation. Mohammad et al. (2013) argue that the impact of LRT on land and property values was higher in East Asian and European cities than in North American cities because North American cities are lower density and more auto dependent. Mohammad et a. (2013) found LRT to have a smaller effect on property values than heavy rail systems because light rail has lower average speeds, capacities, and geographical coverage.

Fewer studies have examined the effects of LRT on housing rents in the post service period. John (1996) found apartment rents higher near Metrorail stations in Washington DC. Hass-Klau et al. (2004) found light rail service increases housing rents as well as property values in three cases (Greater Manchester, UK, Rouen, France, and Portland, USA). In a case study of Santa Clara, Weinberger (2001) found the rents of commercial properties within one half mile of a light rail station to be higher than those of other properties. Bollinger et al. (1998) found office rents higher within walking distance of a rail transit station. Furthermore, the geographical scope of the LRT premium on office rents could extend up to 1.85 miles from transit stations (Nelson et al. 2015).

The number of studies on the effects of light rail on land and housing prices in the pre-service period has grown in the last two decades. This growing body of evidence shows that land and/or property values begin to rise before service begins. For example, in a study in Washington County, Oregon, Knaap et al. (2001) find that LRT projects increase the value of vacant land in station areas once the location of the station are announced. In a study of the Midway line in Chicago, McMillen and McDonald (2004) also found that overall property values increased in the vicinity of light rail stations before the line began operation. In a study of Metrorail stations in Washington D.C., Damm et al. (1980) found that property values near Metro stations rose significantly before Metro began operation. McDonald and Osuji (1995) found that residential land values in the vicinity of a Chicago transit line began to increase two years before the line opened. Agostini and Palmucci (2008) analyzed the anticipated capitalization into housing prices resulting from Line 4 of the Santiago, Chile metro system and found that prices rose soon after construction was announced. Dubé et al. (2018) investigated the effects of LRT on housing prices in Dijon, France and found that prices increased when LRT was in the construction phase.

Some studies, however, have found the impacts of light rail investments in the pre-service period on property value to be negative or minimal (Gatzlaff and Smith (1993); Henneberry (1998); Yan et al. (2012); Loomis et al. (2012); Boucq and Papon (2008)). These studies found this could be due to the presence of existing unattractive industrial uses in the transit corridor, or noise and traffic congestion caused by construction. Although a few studies offer mixed evidence, newer findings indicate that LRT in the pre-service period positively impacts single-family housing and property values (Cao and Lou 2017).

*Approaches to capturing the impact of light rail transit in the preservice period***.** The main method used to capture the impact of LRT impact in the literature is the hedonic model. For example, Knaap, Ding, and Hopkins use hedonic models to empirically test the impact of LRT on the value of vacant residential land in Washington County, Oregon (Knaap et al., 2001). Damm et al. utilize a hedonic model to test the impact of rail transit in Washington, D.C. in the preservice period on the value of single-family property (Damm et al., 1980). Hedonic models are also used in Ferguson, Goldberg, and Mark’s Vancouver case (Ferguson, Goldberg, & Mark, 1984), Golub et al.’s Phoenix case (Golub, Guhathakurta, & Sollapuram, 2012), McMillen and McDonald’s Chicago case (McMillen and McDonald, 2004), and Loomis et al.’s Puerto Rico case (Loomis et al., 2012). As hedonic analysis pools observations, it suffers from omitted-variable bias, because it fails to include all explanatory variables (Hill, 2013). The repeat sale approach is utilized to reduce omitted-variable biasness, as explanatory variables are consistent across sales. For example, McMillen and McDonald used this approach to examine the impact of the Midway line during the preservice period on the value of single-family properties in Chicago (McMillen & McDonald, 2004). Gatzlaff and Smith utilized a repeat sale approach in their Miami study, while Dube et. al. performed this approach in spatial context in France (Devaux et al., 2017; Dubé et al., 2018; Gatzlaff & Smith, 1993).

Further questions about this topic need to be addressed in future studies: 1) when in the preservice period can the impacts be captured; 2)where can any such impacts be felt; 3)and are the effects distributed evenly across space in the vicinity of the LRT? Related existing literature has failed to incorporate empirical studies on the impact of LRT in the preservice period on multifamily rents. Previous research is limited in part because it is difficult to access housing rental panel data. This paper aims to fill that gap by utilizing a high response rate, population scale, and up-to-date multifamily housing two-year panel data with a large sample size to test if LRT in the preservice period impacts multifamily rents, and if so, to what magnitude.

Context

The Purple Line LRT system will provide transit service to the Maryland suburbs of Washington, D.C., and is expected to begin operation in 2022 (see Figure 3.1). Spanning 16.2 miles and 21 stations, it will directly link the suburban hubs of Bethesda, Silver Spring, College Park, and New Carrollton, and provide a direct radial connection between the Red, Green, and Orange lines at existing stations of the Washington Metro system. It will also connect with three lines and stations of the Maryland Area Regional Commuter (MARC) rail system, and with one stop on Amtrak's northeast corridor – New Carrollton. The cost of construction is expected to exceed $2.3 billion and create 6,300 jobs over the five-year construction period[8]. Estimated daily ridership by 2040 is 74,000 per day[9].

[8] Source: http://mta.maryland.gov. Governor O'Malley Announces Purple Line Receives Federal Environmental Approval.
[9] Source: http://mta.maryland.gov. Governor O'Malley Announces Purple Line Receives Federal Environmental Approval.

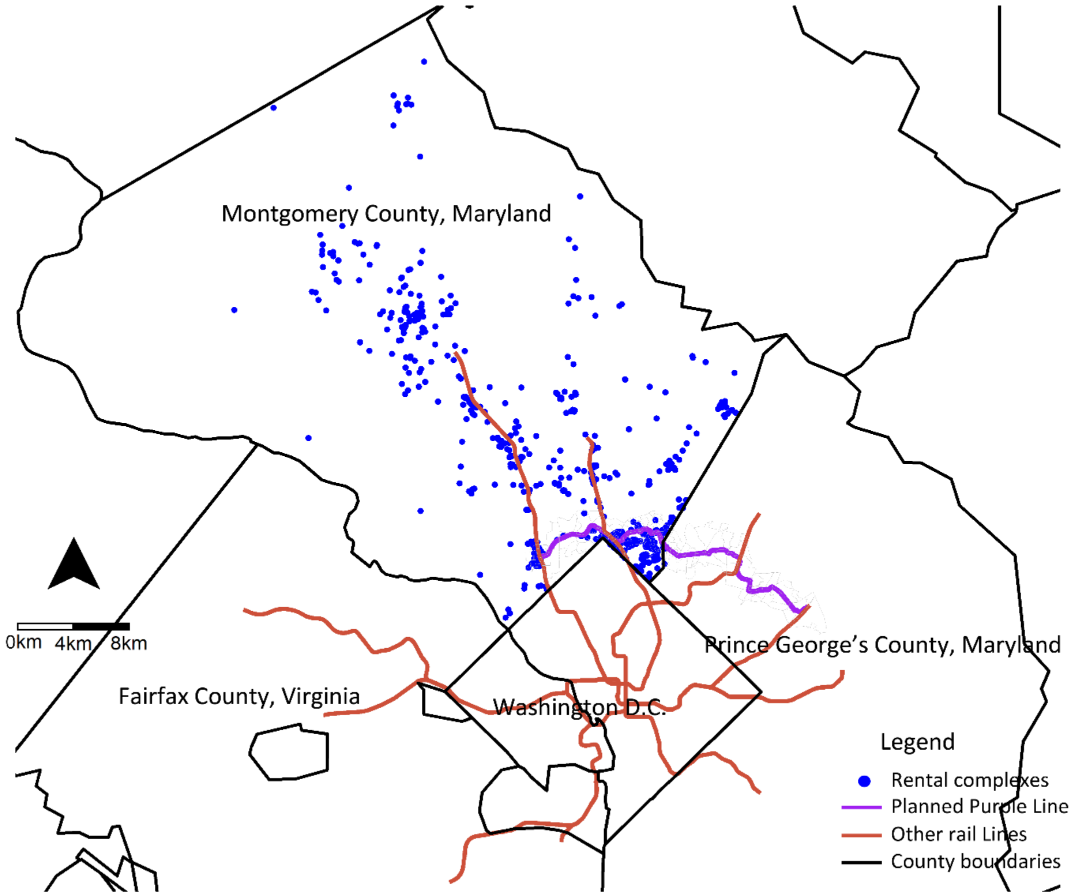

Figure 3.1: Sample location and the Purple Line

The project weathered two decades of political, legal, and financial battles before construction began (see Figure 3.2). The draft environmental impact study was completed in 2008 and the Federal Transit Administration issued a record of decision in 2014[10]. Additional milestones occurred after that decision. In 2016 the Maryland Transit Administration announced the team of private companies that would build, operate, and maintain the Purple Line. In August 2017, $900 million of federal funding was officially granted by the Federal Transit Administration, and

[10] Source: https://www.purplelinemd.com/about-the-project/overview

construction began. In December 2017, and again in March 2019, the U.S. Court of Appeals for the D.C. Circuit dismissed environmental and other lawsuits which sought to block construction. It should be noted that the Purple Line project made significant progress after 2015 and this progress conveys confirmative and promising information regarding the location of planned stations and the time frame during which transit service will begin (see Figure 3.2).

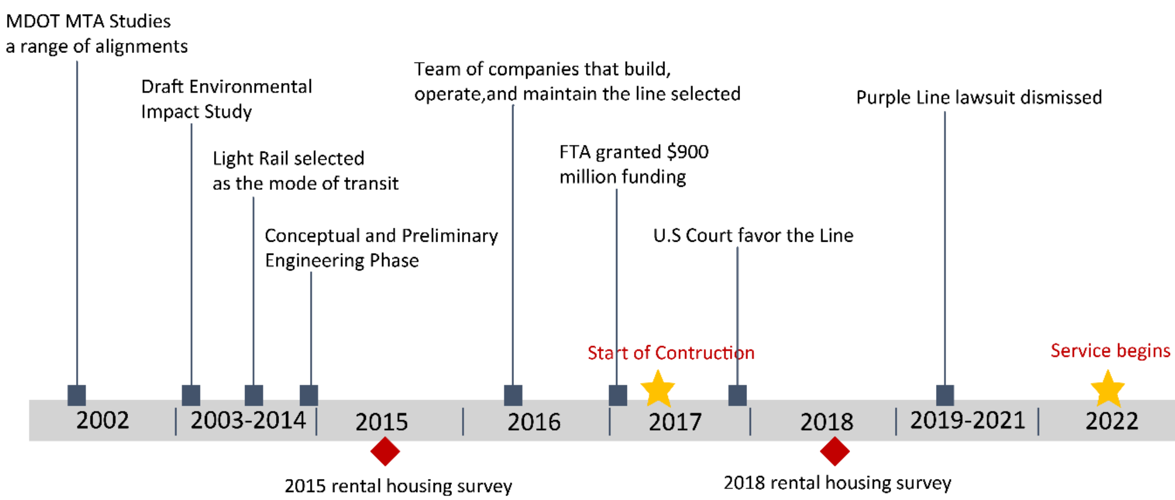

Figure 3.2: Purple Line Planning and Construction Timeline

Although the line will traverse two Maryland counties – Montgomery and Prince George's – I limit my analysis to the 10 stations in Montgomery County, home to more than 1 million residents (U.S. Census Bureau, 2017) and several U.S. federal government agencies. Although the county is one of the most affluent in the United

States[11], tenants in multifamily properties are heavily cost-burdened. In 2015, the county ranked 9th in the nation with a median monthly rent of $1,656[12](see Figure 3.3). The census estimates that more than 49% of renter households spend more than 30% of their incomes on rent (U.S. Census Bureau, 2017).

The population of Montgomery County is diverse, as more than one in three residents are foreign-born (U.S. Census Bureau, 2017), many from Central America and Asia. Socioeconomic diversity is especially pronounced in the Purple Line corridor (see Figure 3.3). In the census tracts along the corridor in Montgomery County, median home values range from just under $400,000 to over $800,000; rents range from less than $1,300 to over $2,300; and median household incomes range from less than $80,000 to over $130,000 (U.S. Census Bureau, 2017). High-income, predominantly white neighborhoods are concentrated in the west side of the corridor while low-income, predominantly minority neighborhoods are concentrated on the east side of the corridor (Figure 3.3).

[11] The estimated Median Income of the county (monthly income $8,243 in 2015) is ranked 15th in the nation in regards to median income. Source:https://stat.montgomerycountymd.gov/stories/s/Rental-Survey/c98k-yku5/

[12] https://stat.montgomerycountymd.gov/stories/s/Rental-Survey/c98k-yku5/

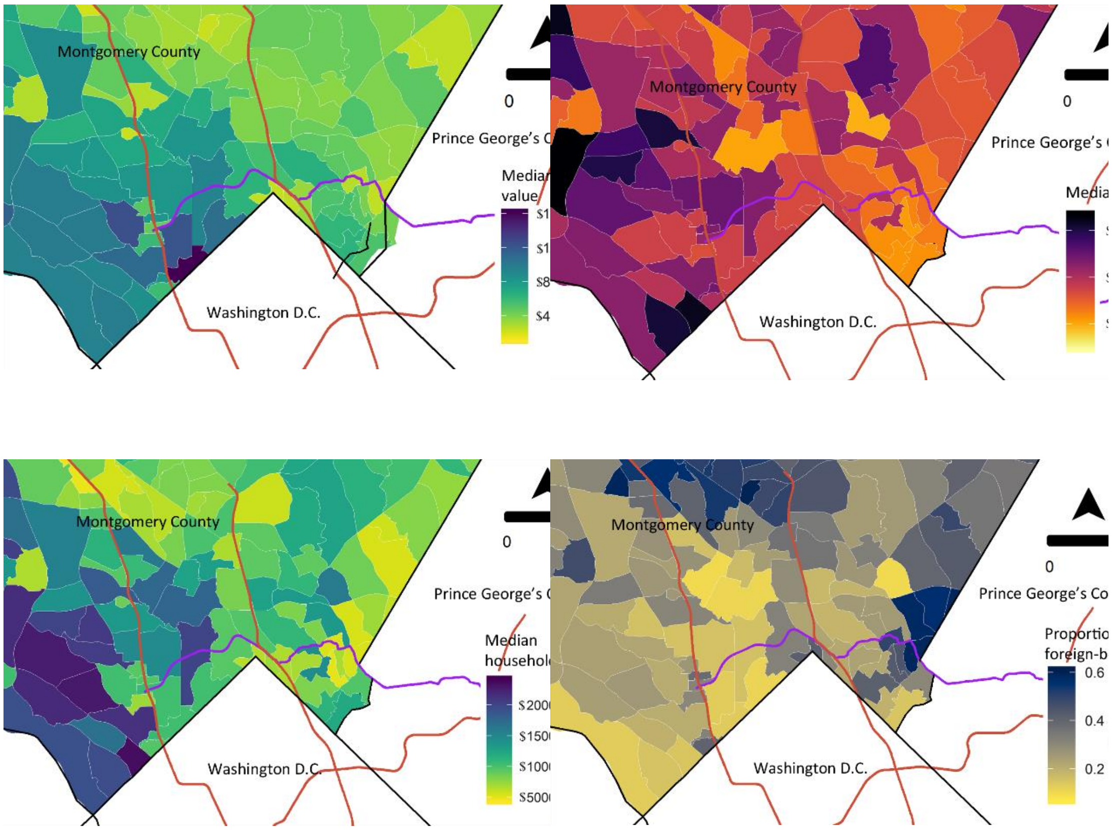

Figure 3.3: Maps of characteristics of demographics and housing market in the rail transit corridor (left top: median home value; right top: median rent; left bottom: median household income; right bottom: proportion of foreign-born residents)

Many expect the Purple Line to increase property values and rents, which has raised concerns about gentrification and displacement, especially on the east side of the corridor. For these reasons, the county has adopted a variety of anti-displacement policies, such as tenant protection laws, inclusionary zoning, and direct acquisition and preservation of affordable units. As such, the effects of the Purple Line on housing prices and rents are of considerable policy relevance.

Data

The prospective rail line crosses two Maryland counties: Montgomery and Prince George's. This essay focuses on Montgomery County, which will host 10 Purple Line stations. My empirical dataset is compiled based on four primary data sources.

This essay's first data source is the Montgomery County Rental Housing Survey for 2015 and 2018, obtained from the Montgomery County Department of Housing and Community Affairs. Montgomery County law requires all landlords of multifamily rental housing units to participate in an annual rental housing survey. I have created an interactive map for the survey data. Here is the web address: https://rpubs.com/xqpeng/RentalSurveyAnalysis. The first source is the Montgomery County Department of Housing and Community Affairs, which by statute requires all owners of multifamily rental property to complete an annual rental housing survey. The earliest survey available for public access is the 2015 rental housing survey, which was conducted and completed in April, 2015, and included 846 of 930 multifamily rental facilities. All facilities in the dataset are multi-unit apartment buildings. In 2018, 923 of 1047 rental facilities were included in the survey. I construct a two-year panel by matching building and unit identifiers across 2015 and 2018. In 2015, construction of the Purple Line had not yet begun, but multiple steps toward implementation had been completed, including the environmental impact statement. By the completion of the 2018 survey, construction had begun. Figure 3.1 shows the location of the multifamily complexes. Each rental housing survey has a hierarchical structure with two levels. The lower level is the unit level, which has two variables – unit rent and the number of bedrooms. Unit rent is a transaction rent for which payment is made to property managers or owners. I excluded observations that are vacant or employee-occupied because the rents of these observations are not market-driven. The higher level of the hierarchy is the complex level, which takes into account the physical characteristics of buildings, amenities, services offered and management attributes.

I also obtained data regarding urban amenities and the Purple Line from Montgomery County Open Data, the second data source. To indicate the quality of public schools, I requested access to the 2013 Maryland State Education Indicators, the third data source. The indicators offer information on average math and reading scores for third-graders and fifth-graders from each public elementary school in Montgomery County. The fourth data source is the American Community Survey, from which I extract neighborhood characteristics, such as the percentage of African American households that comprise the population.

Table 3.2 lists variables included in the analysis. For each observation, the variables contain the following: monthly rent of unit, a dummy variable indicating if an observation is from 2018 survey, a dummy variable indicating whether the complex falls within the half-mile radial catchment area of a Purple Line station; a dummy variable indicating whether the apartment complex is between a half mile and one mile away from a Purple Line station; a continuous variable indicating the age of complex; a dummy variable indicating if a complex has a swimming pool; a dummy variable indicating if a complex allow any pets in; a dummy variable indicating if a complex provide storage service; a dummy variable indicating if a complex provide storage service; a dummy variable indicating if a complex provide any shared laundry room; a dummy variable indicating if a complex has any housing affordability program; a dummy variable indicating if a complex is less than 0.25 mile from a cemetery; a continuous variable indicating public elementary school quality, measured using fifth grade reading scores; ;a continuous variable indicating distance between the complex and the nearest shopping center; a continuous variable indicating foreign born percentage of a neighborhood where a complex locates; a continuous variable indicating median household income of a neighborhood where a complex locates; a continuous variable indicating accessibility level to jobs by walking mode ; a continuous variable indicating accessibility level to jobs by driving mode; and a continuous variable indicating crime rate of a neighborhood where a complex locates.

Table 3.2: Definition of variables

Category	**Variable Name**	**Definition**
Dependent variable	Rent	Monthly rent of unit
Independent variable	Time	If an observation from 2018 survey, assign 1. Otherwise,0.
	Half-mile treated	If a complex is within half mile from planned stations, assign 1. Otherwise, 0.
	One-mile treated	If the distance between a complex and planned stations is larger than half mile but less than one mile, assign 1. Otherwise, 0.
	Age of structure	Age of building structure
	Swimming pool	If a complex has a swimming pool, assign 1. Otherwise, 0.
	Gym	If a complex has a gym, assign 1. Otherwise, 0.

	Pet allowance in building	If a complex allow any pet in, assign 1. Otherwise,0.
	Storage service	If a complex provide storage service, assign 1. Otherwise,0.
	Shared laundry room	If there is a shared laundry room, assign 1. Otherwise, 0.
	Affordability program	If a complex has any housing affordability program, assign 1.Otherwise 0.
	Neighboring cemetery	If a complex is less 0.25 mile from a cemetery, assign 1. Otherwise, 0.
	Shopping center	Distance between a complex and its nearest shopping center.
	5^{th} reading score	Use average fifth grade reading score per school as indicator of performance of elementary school. The score varies from 0 to 100. The larger the score is, the better performance the school has.
	Foreign born percentage	Foreign born percentage at census tract level
	Household income	Median household income at census tract level
	Accessibility score by walking to jobs	An indicator to quantify accessibility of a block to jobs by walking travel mode
	Accessibility score by driving to jobs	An indicator to quantify accessibility of a block to jobs by auto travel mode
	Crime rate	Crime rate at the census tract level (2017)

The descriptive statistics of the variables are shown in Table 3.3. The mean rent is $1,511, while 53.6% observations are from the 2018 survey. 14.2% observations are falling in the half-mile treated group, while 2.8% observations are falling in the one-mile treated group. The mean unit is 38.9 years old, feeds a school with a reading score of 91.4 for 5th grade, and is 1,926 feet from the nearest shopping center. The statistics of other variables can be found in the Table 3.3.

Table 3.3: Descriptive statistics of datasets

Variable Name	Mean	Std.Dev.	Min	Max
Rent	$ 1,511	404	$154	$ 3,470
Time	0.536	0.499	0	1
Half-mile treated	0.142	0.44	0	1
One-mile treated	0.028	0.225	0	1
Age of structure	38.9	20.2	1	133
Swimming pool	0.645	0.478	0	1
Gym	0.586	0.493	0	1
Pet allowance in building	0.769	0.422	0	1
Storage service	0.376	0.484	0	1
Shared laundry rooms	0.564	0.496	0	1
Providing Affordability program	0.594	0.491	0	1
Neighboring cemetery	0.022	0.148	0	1
Shopping center(feet)	1,926	1,515	70	23,706
5^{th} reading score	91.4	6.29	73.0	100.0
Foreign born percentage	37.5	11.1	6.5	61.0
Household income	$81,396	26552	$42,955	$227,778
Accessibility score by walking to jobs	5,968	5,165	33	25,220
Accessibility score by driving to jobs	85,976,173	9,104,000	51,447,026	98,358,892
Crime rate	9,722	11,862	953	66,161

Note: The two-year pool dataset has 137,547 observations in 793 complexes.

Econometric Methods

To test if rents of units in the light rail transit corridor increase significantly, I apply difference-in-difference (DID) approach and report results. To validate the DID results I get, I apply a first difference approach and report the results. To do so, I define two treatment groups: those units within a half-mile and those within one mile of a proposed station. Rental units within a half mile from a planned Purple Line station are included in the half-mile treated group. Rental units more than a half mile away but less than a radius of one mile from the planned Purple Line stations are included in the one-mile treated group (See Figure 3). Rental units elsewhere in the County are used as the control group.

I defined two treated groups: half-mile, and one-mile. Rental units within a half-mile radius from planned Purple Line stations are included in the half-mile treated group. Rental units located more than a half-mile away and less than one mile from the planned Purple Line stations are included in the one-mile treated group (See Figure 3.4). Rental units elsewhere in the county are grouped as the control group.

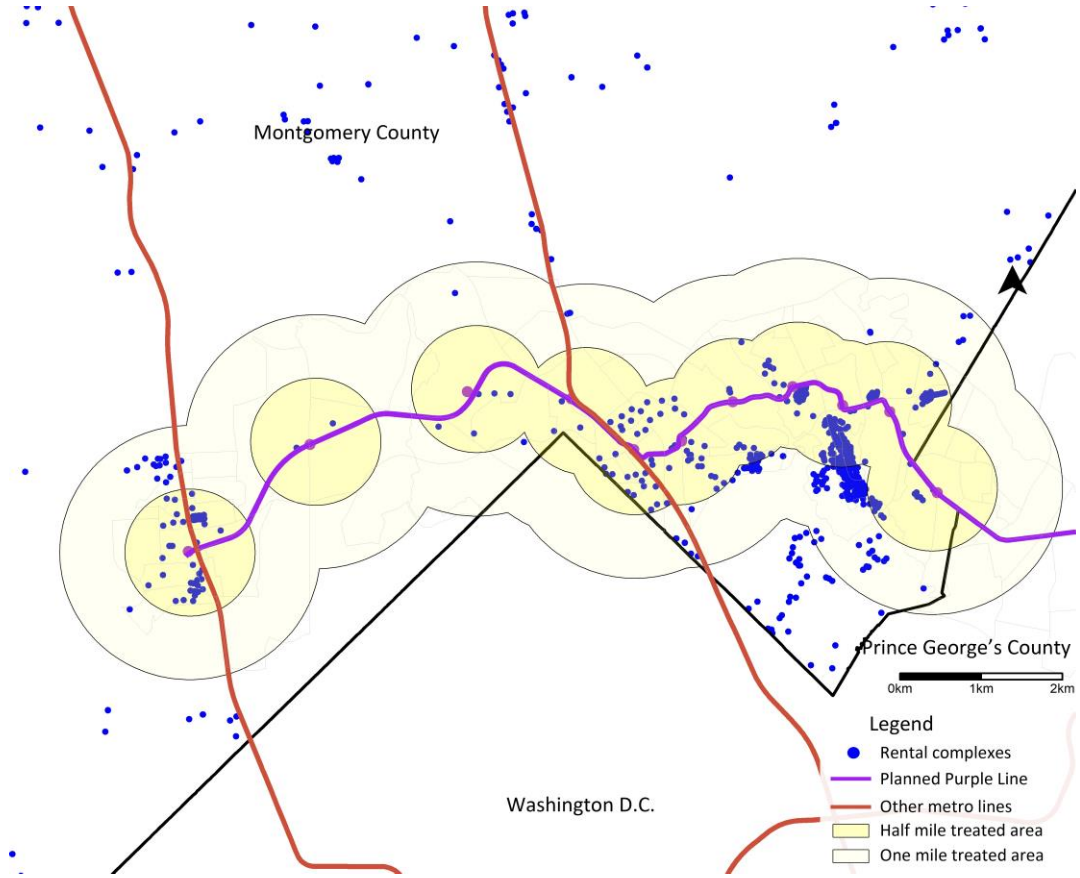

Figure 3.4:The treated groups and the control group

Difference-in-Difference approach. The specification of DID model is:

$$p_i = \alpha + \beta_0 * Time_i + \beta_1 * HalfMile_i + \beta_2 * (Time_i * HalfMile_i) +$$
$$\gamma_1 * OneMile_i + \gamma_2 * (Time_i * OneMile_i) + \omega^T * X_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$$

(1)

where P_i is the rent of the unit: i indicates unit observation[13]. The observation i is either from 2015 survey or 2018 survey. $Time_i$ is a dummy variable which takes the

[13] I use rent as the dependent variable instead of logarithm of rent. As such, we can interpret the coefficients of the treated group variables as: if a unit falls in the treated group, how many rents of the unit increased from 2015 to 2018.

value 0 for observations from 2015 survey. When observations were in 2018 survey, the value is 1. $HalfMile_i$ is a dummy variable which indicates whether unit i is within a half mile from any Purple Line station. $OneMile_i$ is a dummy variable which indicates whether unit i is more than a half mile away and less than one mile from any Purple Line station. ε_i is the error term, which follows the normal distribution. X_i is a column vector of attributes of unit i. α, β_0, β_1, β_2, γ_1, and γ_2 are corresponding coefficients. ω is a coefficient vector for X.

The control group consists of observations which are more than one mile away from Purple Line stations. If the estimate of   is positive and significant, units within the half-mile treated group see a larger rent increase than their counterparts in the control group from 2015 to 2018. If the estimate of   is positive and significant, it means units within one-mile treated group see a larger rent increase than their counterparts in the control group.

There are three frequently mentioned issues regarding the DID approach. First, the DID approach is subject to omitted variable biasness problem. To reduce the problem I may suffer, I include as many control variables I can get into the DID models. To further valid the results of the DID approach, I apply another approach-a first difference approach- in the technical appendix. The second issue is that there is a potential self-selection bias when dividing rental units into the treated groups and the control group based on their distance to the planned rail stations. Table 3.4 examines the balance of characteristics of the rental units in the treated groups and the characteristics of rental units in the control group in 2015. Most of characteristics of units in the treated groups are comparable to units in the control group, which suggests that I may be free of the self-selection bias. The third issue is taking spatial autocorrelation into account. In this analysis, an apartment complex could include dozens or hundreds of number of units. A conventional spatial econometric model cannot be applied because the weight matrix based on unit geographical location is singular. Finally, in order to estimate the rent price gradient with respect to the distance to planned stations, I ran parameterized distance decay regressions using the half-mile treated group.

Table 3.4: Self-selection bias check between units in the treated units in control (2015)

	0-0.5 mile treated group	**0.5-1 mile treated group**	**>1 mile control group**
Variable Name	Mean	Mean	Mean
Rent	\$1,507	\$1,422	\$1,447
Number of bedrooms	1.44	1.4	1.64
Age of structure	47.1	52.3	36.1
Swimming pool	0.466	0.303	0.743
Gym	0.542	0.25	0.612
Pet allowance in building	0.69	0.62	0.804
Storage service	0.365	0.281	0.357
Shared laundry rooms	0.676	0.82	0.512
Providing Affordability program	0.524	0.334	0.617
Neighboring cemetery	0	0	0.031
Shopping center(feet)	1,224	2,770	2,193
5^{th} reading score	93.4	92.4	90.3

Foreign born percentage	32.9	36.3	39.6
Household income	$74,107	$86,430	$84,004
Accessibility score by walking to jobs	9,863	9,946	3,953
Accessibility score by driving to jobs	94,036,150	92,812,090	82,355,301
Crime rate	14,180	13,000	7,709
Number of units	16,511	3,481	43,873

The DID approach I apply in the main text may suffer from omitted variable bias problem, although I include as many control variables as I can get into the DID models. To validate the DID results I get in the main text, I apply a first difference approach and report the first difference results in this technical appendix. The first difference approach is expected to be free of the omitted variable bias problem, if the attributes of units and characteristics of neighborhoods remain unchanged from 2015 to 2018.

First-difference approach. I specify a hedonic price function with unit rents as the dependent variable, as a function of the unit's physical attributes and locational characteristics. For cross section of year t, we have hedonic model (Equation (12)).

$$p_{it} = \alpha_t + \delta_t * HalfMile_i + \lambda_t * OneMile_i + \theta_t * X_{it} + \varepsilon_{it}, \varepsilon_{ij} \sim N(0, \sigma^2) \quad (12)$$

where P_{it} is the rent of unit i in year t; X_{it} is a vector of characteristics of unit i in year t; $HalfMile_i$ is a dummy variable that equals 1 if the unit i is located within a half mile of a planned Purple Line station; $OneMile_i$ is a dummy variable that equals 1 if the unit i is located within one mile of any Purple Line station and more than a half mile far away from any Purple Line station; ε_{it} is the error term. α ,δ_t ,θ , λ_t and are the corresponding coefficients. I allowed the coefficients of $HalfMile_i$ and $OneMile_i$, δ_t and λ_t, to vary across time; other coefficients are constrained to be constant over time.

Consider a first difference of the price equation for the same unit i in year t and in year s, with t>s. With the assumption that characteristics of unit i are constant across the years,

$$p_{it} - p_{is} = \alpha_t - \alpha_s + (\delta_t - \delta_s) * HalfMile_i + (\lambda_t - \lambda_s) * OneMile_i + \varepsilon_{it} - \varepsilon_{is} \quad (13)$$

We have 2-year panel data with t=2018 and s=2015. By controlling the effect of omitted variables that do not change over time, the OLS regression results of the first difference of rent on dummy variables *Half Mile* and *OneMile* are potentially subject to less omitted-variable bias (Equation (2)). ($\delta_t - \delta_s$) and ($\lambda_t - \lambda_s$) are the factors of interest. If the estimate of (($\delta_t - \delta_s$) is positive and significant, it means that units in the half-mile treated group see a larger rent increase than their counterparts in the control group. If the estimate of ($\lambda_t - \lambda_s$) is positive and significant, it means units within one-mile treated group see a larger rent increase than their counterparts in the control group.

The reduction in bias comes at the possible cost of a large decrease in sample size and possible selection bias, as in the repeat-sales method used by McMillen (2003), McMillen and Dombrow (2001), and McMillen and McDonald (2004). In this case, the sample size is reduced from 137,547 to 102,586 after I match 2015 unit-level survey data with 2018 unit-level survey data via unique unit ID and unique apartment complex ID. As a result, 71% units are successfully matched between 2015 and 2018. Some units are not able to be matched because these units' ID in the surveys are muzzy or missing. Given that rental units are clustered in complexes, I note that the standard errors of coefficients should be adjusted following Abadie et al. (2017). To do so, I use first-difference regressions with clustering.

Difference-in-difference approach. In order to verify the results of the first-difference regressions and make use of information of all observations, I next implemented a different-in-difference (DID) approach. The DID approach retains all unit observations, with a total of 144,341. The cost is that omitted variable bias may exist, because I cannot include all the control variables in the regression. In response, I included some critical control variables in the regressions, such as the age of the complex, the distance to the nearest shopping center, and indicators of the quality of public schools. These variables account for the variability of attributes across the treated groups and the control group. The specification of DID model is:

$$P_{it} = \alpha + \beta_0 * Time_t + \beta_1 * HalfMile_i + \beta_2 * (Time_t * HalfMile_i) + \gamma_1 * OneMile_i + \gamma_2 * (Time_t * OneMile_i) + X_i * O + \varepsilon_{it} \tag{14}$$

where P_{it} is the rent of the unit; i indicates unit observation while t indicates time. $Time_t$ is a dummy variable, and for observations based in 2015, the value is zero. For observations based in 2018, the value is 1. $HalfMile_i$ is a dummy variable that indicates whether unit i is within a half mile of any Purple Line station. $OneMile_i$ is a dummy variable which indicates whether unit i is more than a half mile away and less than one mile from any Purple Line station. X_i is a column vector of attributes of unit i.

The control group consists of observations located more than one mile away from Purple Line stations. If the estimate of β_2 is positive and significant, units within the half-mile treated group see a larger rent increase than their counterparts in the control group from 2015 to 2018. If the estimate of γ_2 is positive and significant, it means units within one-mile treated group see a larger rent increase than their counterparts in the control group.

Parameterized distance decay regressions. If units within the half-mile treated group see increases in rent because of the Purple Line project, it is interesting then to examine how the magnitude of that effect varies geographically. Figure 3.5 shows the sample for parameterized distance decay regressions. The parameterized distance decay regressions essentially are hedonic models:

$$\ln(P_i) = \alpha + \rho * d_i + \phi X_i + u_i \tag{15}$$

where $\ln(P_i)$ is the logarithm of rent of unit i, d_i is the distance of unit i from the nearest Purple Line station, X_i is a vector of attributes of unit i and u_i is an error term. α is intercept and ϕ is a vector of coefficients. ρ is the rent price gradient with respect to distance to the nearest Purple Line station, the main interest of the parameterized distance decay models. ρ is expected to be negative, which means that as the distance from a unit to the nearest Purple Line stop increases, its rent decreases at $|\rho|$ magnitude.

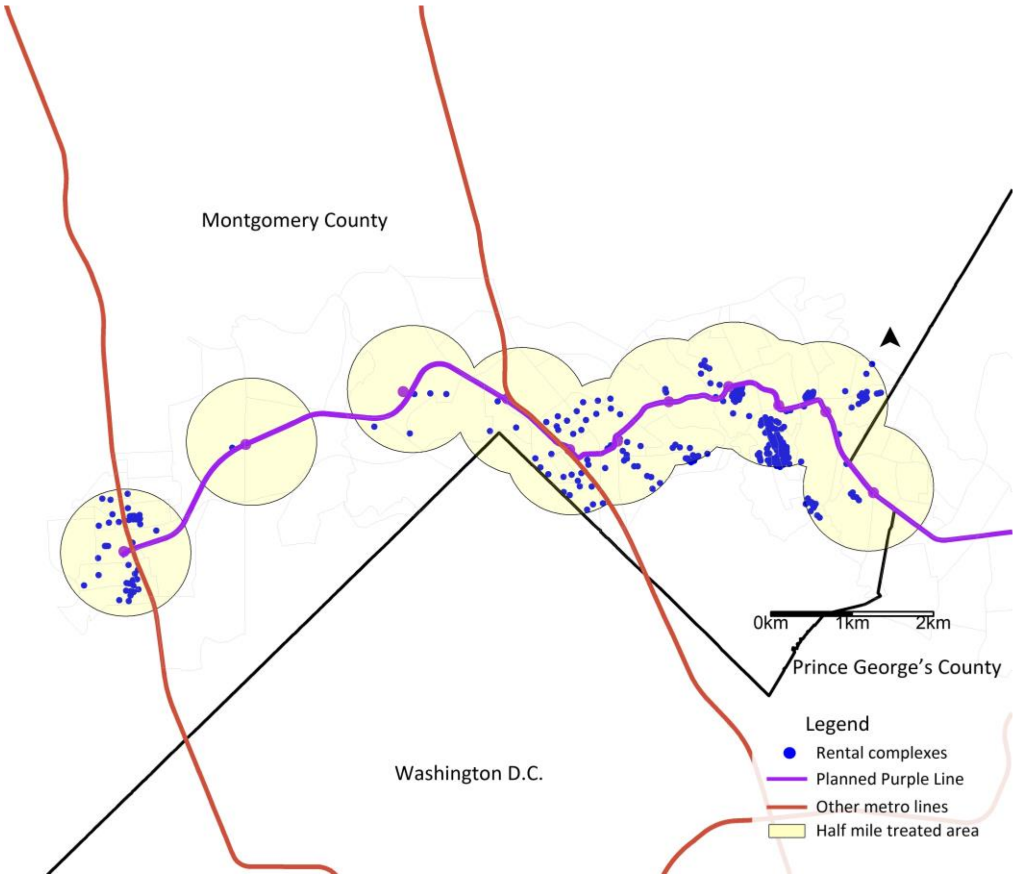

Figure 3.5: Samples falling in the half-mile treated group

Empirical Results

To get you a flavor what the DID results would be, I present average rent comparisons for the half-mile treated group and the control group from 2015 to 2018 (Figure 3.6). As we can see from the top left and top right subfigures, average rents for each bedroom type units in the half-mile treated group and the control group increase from 2015 to 2018. It is noticeable that the average rents of 2-bedroom units, 3-bedroom units, and 4-bedroom units in the half-mile treated group increase more than the their counterparts in the control group, as shown in the bottom right subfigure. The comparison show straightforwardly that rents of 2-bedroom, 3-

bedroom, and 4-bedroom units in the half-mile LRT corridor increase substantially than that in the control group. Of course, only simple comparison is not enough. The rent increase may due to other factors, such as social economic environment, school performance, etc. Next, I apply the DID regressions with a bunch of control variables to control other factors and see if the premium of the LRT still exists.

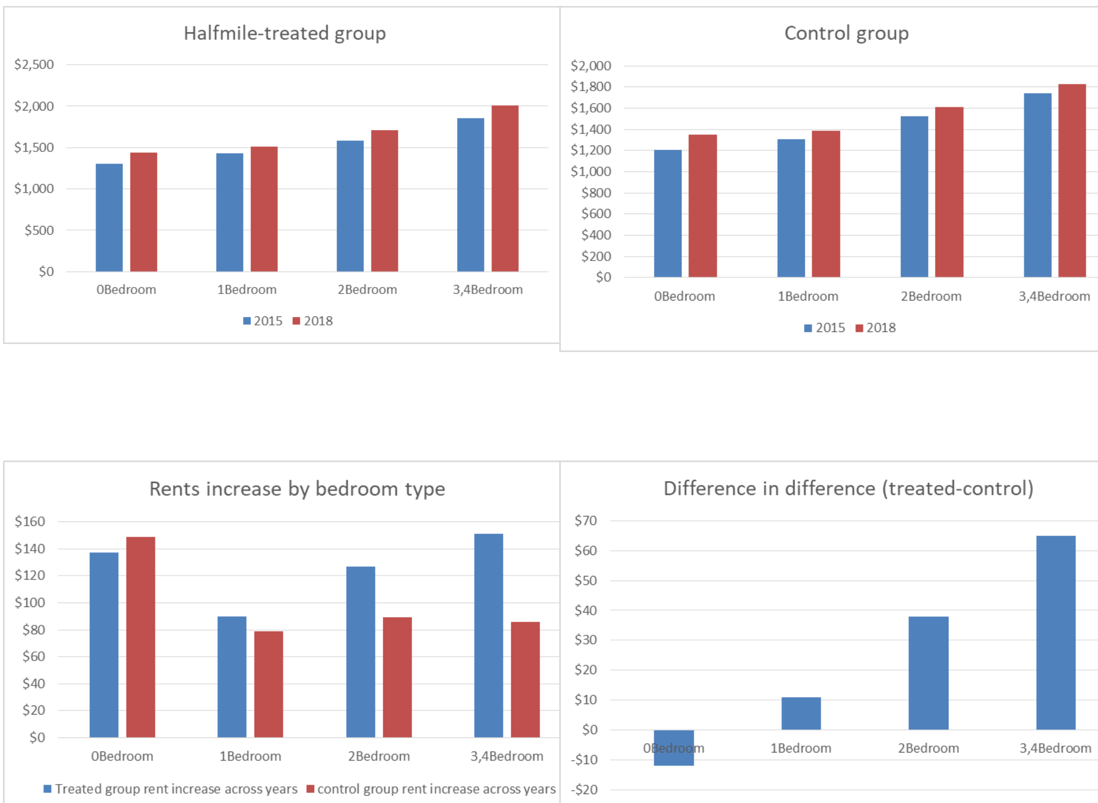

Figure 3.6: average rents comparison between treated group, control group, before, and after event

(left top: half mile-treated group average rent in 2015 and 2018; right top: control group average rent in 2015 and 2018; rents increase by bedroom type for both treated group and control group; right bottom: different in difference by bedroom type)

First, I estimate Equation (14) and report results in Table 3.5. Following Wolverton et al. (1999) who find rental housing markets segmented by number of bedroom units, I analyse rents for four distinct unit-type submarkets (studio, one-bedroom, two-bedroom, and three/four-bedroom). Column (1) to Column (4) report results that correspond to various subsamples. For example, Column (3) presents results based on the subsample of 2-bedroom units. All coefficients of control variables have expected directions and the models have respectful R-squared with values around 0.5. Here, let's focus on the coefficient on (Time*HalfMile) and the coefficient on (Time*OneMile). The coefficients on (Time*HalfMile) for the 2-bedroom, 3-bedroom, and 4-bedroom unit subsamples are positive and significant with the magnitude of $28.654, and $81.817, respectively. These results imply that the rents of 2-bedroom, 3-bedroom, and 4-bedroom units within one half mile of a planned Purple Line station all rose relative to units in the control group. These increases in monthly rent are both statistically significant at 99.9% level.

There are three main findings from the DID regressions. The most important finding is that LRT in the preservice period has positive impact on multifamily rents. Specifically, as shown in Column (3), the results show that the effect for 2-bedroom unit in the half-mile treated group is statistically significant. A unit which is in the half mile transit corridor increased $28.654 more monthly rent than units in other else of the county. For 3-bedroom and 4-bedroom units, the average premium of LRT is $81.817 monthly. The second important finding is that the impact of LRT on rents is found only for units that have more or equal than two bedrooms in the half-mile treated group. Column (1) and Column (2) show that the estimated effects for 0-

bedroom and 1-bedroom units are insignificant. The third finding is that the impact of the LRT on rents is not found for units in the one-mile treated group. The estimated effect of the LRT on 3-bedroom and 4-bedroom units in the one-mile treated group is statistically insignificant.

Table 3.5: Results of difference-in-difference models

	0 Bedroom (1)	1 Bedroom (2)	2 Bedroom (3)	3,4 Bedroom (4)
Intercept(α)	-15793.408*** (1041.979)	-8775.301*** (274.813)	-15969.790*** (220.545)	-10496.545*** (761.824)
Time (β_0)	98.294*** (7.759)	59.952*** (2.907)	73.176*** (2.466)	60.682*** (7.060)
HalfMile (β_1)	2.641 (10.671)	3.084 (4.859)	-6.492 (4.591)	-122.644*** (13.089)
OneMile (γ_1)	49.719*** (14.904)	92.923*** (7.618)	0.618 (8.024)	-87.828*** (26.115)
$Time * HalfMile$ (β_2)	**20.818** (10.888)	**4.682** (5.160)	**28.654***** (5.176)	**81.817***** (14.808)
$Time * OneMile$ (γ_2)	-2.493 (18.242)	-9.625 (9.743)	0.943 (10.264)	42.287 (33.980)
Age	-2.997*** (0.212)	-4.790*** (0.080)	-7.320*** (0.081)	-8.352*** (0.307)
Swimming pool	20.323** (7.695)	132.893*** (3.101)	134.479*** (2.861)	128.853*** (9.025)
Gym	186.381*** (7.400)	110.228*** (3.153)	32.975*** (2.881)	100.311*** (8.833)
Pet allowance in building	-13.641 (7.419)	1.658 (3.163)	60.662*** (2.931)	139.780*** (8.561)
Storage service	114.197*** (6.246)	95.770*** (2.928)	67.611*** (2.690)	63.920*** (8.699)
Shared laundry room	15.278* (7.496)	-63.976*** (2.691)	-28.022*** (2.524)	-26.867*** (8.045)
Affordability program	-31.032*** (7.759)	-158.324*** (2.920)	-107.334*** (2.497)	-207.016*** (7.355)
Neighboring cemetery	-232.126*** (31.577)	-89.624*** (8.622)	-78.939*** (7.209)	-182.943*** (18.901)
Logarithm of distance to shopping center	-0.215 (3.806)	-44.520*** (1.848)	-30.091*** (1.639)	-73.544*** (5.329)

5th reading score	12.055*** (0.596)	4.794*** (0.219)	5.528*** (0.194)	9.062*** (0.553)
Foreign born percentage	5.444*** (0.428)	-1.719*** (0.154)	-1.016*** (0.136)	-7.099*** (0.446)
Log(household income)	56.538** (17.704)	53.474*** (6.470)	197.384*** (5.270)	54.318** (18.854)
Logarithm of accessibility score by walking to jobs	57.091*** (5.030)	104.582*** (1.812)	115.163*** (1.542)	115.595*** (4.432)
Logarithm of accessibility score by driving to jobs	799.732*** (55.679)	496.337*** (14.374)	791.260*** (11.646)	650.811*** (39.242)
Log(crime rate)	-13.941*** (3.874)	-42.020*** (1.808)	-13.911*** (1.831)	-97.671*** (6.593)
R-squared	0.507	0.479	0.502	0.486
Adjusted R-squared	0.506	0.479	0.502	0.485
Number of observations	6,737	57,485	61,836	11,489
Number of apartments	207	674	647	267

The dependent variable is rent for each unit. Standard error is in parentheses
*** p=.001(two-tailed). **p=0.01(two-tailed). *p=0.05(two-tailed).

Second, I ran Equation (12) and observed rental units clustered by complexes. Simply running Equation (12) may underestimate the standard errors in OLS, resulting in incorrect inferences. As such, I solved for Equation (13) and reported the results of the first-difference approach in Table 3.6 using a balanced panel dataset and correct standard errors. The standard errors are corrected by taking the fact that multifamily housing units are clustering by complexes. Ignoring this fact and running OLS would underestimate the standard errors. Column (1) to Column (4) of Table 3.6 report results corresponding to various subsamples. For example, Column (3) presents results based on the subsample of two-bedroom units. The coefficients of the half-mile treated group and one-mile treated group are the main focuses. Breaking down the results into subsamples by unit type, the coefficients of the half-mile treated group for the two-bedroom subsample, the three-bedroom subsample, and the four-bedroom subsample are positive and significant. These results imply that the effect of the Purple Line project is captured via higher rents, specifically rents of two-bedroom units, three-bedroom units and four-bedroom units by the magnitude of $36.666 and $66.300, respectively. With regards to the one-mile treated group, all the coefficients are not significant. The results of first-difference models are similar to the results of the DID models, which verifies the results are convincing.

Table 3.6: Results of the first-difference models

	0 Bedroom (1)	1 Bedroom (2)	2 Bedroom (3)	3,4 Bedroom (4)
Intercept ($\alpha_{2018} - \alpha_{2015}$)	94.751* (43.200)	49.625** (16.302)	53.254*** (6.510)	52.300 *** (15.800)
Half-mile treated group ($\delta_{2018} - \delta_{2015}$)	**-29.234** (44.900)	**0.113** (25.708)	**36.662**** (12.870)	**66.300^** (34.100)
One-mile treated group ($\lambda_{2018} - \lambda_{2015}$)	-40.383 (46.200)	-8.178 (19.104)	3.971 (16.150)	51.700* (24.800)
Sample size	2,559	21,830	22,686	4,063
R-squared	0.0097	0.00008	0.0085	0.0159
Adjusted R-squared	0.0089	0.00004	0.0084	0.0154

The dependent variable is the change of rent for each unit. Standard error is in parentheses and is corrected.
*** p=.001(two-tailed). **p=0.01(two-tailed). *p=0.05(two-tailed). ^p=0.1(two-tailed)

Finally, it is interesting to examine how the magnitude of this effect varies across geography, since the above results show that the effects exist for two-bedroom, three-bedroom, and four-bedroom units in the half-mile treated group. I ran Equation (15) and report the results in Table 3.7 based on various samples and subsamples. Column (1) uses pooled observations from the half-mile treated group in 2015, while

Column (2) uses pooled observations from the half-mile treated group in 2018. Columns (3), (4), (5), and (6) use subsamples of zero-bedroom, one-bedroom, two-bedroom, and three-bedroom units[14]. The main focus is the coefficient for distance to a planned Metro station in miles (ρ). The estimate of ρ for pooled 2015 units is not significantly different from zero. However, the estimate of ρ for pooled 2018 units is negatively significant with the magnitude of -0.077 , controlling attributes of units, its complexes and locational characteristics. This means that proximity to planned Purple Line stations impacted rent in 2018, despite the fact that it did not influence rent in 2015; this further validates the results of the first-difference approach and DID approach. The rent prices gradient is -0.077, which means that rents of units adjacent to the Purple Line stations are, on average, 7.7% higher than those located one mile away. As a reference point, in the case of single-family dwellings, the price gradient in Chicago was 7.4% for households located one mile from the Midway line (McMillen and McDonald, 2004).

The rent price gradients vary from -0.245 (zero-bedroom subsample) to -0.081 (one-bedroom subsample) to -0.104 (two-bedroom subsample) to 0.645 (three-bedroom subsample). Observing zero-bedroom units, one-bedroom units, and two-bedroom units, rents increased with closer proximity to the nearest planned Purple Line station(s). The estimated rent price gradient for three-bedroom units is unusual – it is positive and significant. The most likely explanation of this result is the small sample size and relative rarity of three-bedroom units in apartment buildings; such

[14] I do not report results of four-bedroom unit subsample because the four-bedroom unit subsample sample size is too small and the regression results are not stable.

units may irregularly located across a given space, or clustered together. Another possible cause of this finding is that households in three-bedroom units are likely to have children and, thus, are more likely to rely heavily on private vehicles for commuting instead of transit. As such, they may not value proximity to transit.

Table 3.7: Price gradient results

	Using 2015 dataset	Using 2018 dataset				
	Pool (1)	Pool (2)	0 Bedroom (3)	1 Bedroom (4)	2 Bedroom (5)	3 Bedroom (6)
Variable	Coefficient (Std. Error)	Coefficient (Std. Error)	Coefficient (Std. Error)	Coefficient (Std. Error)	Coefficient (Std. Error)	Coefficient (Std. Error)
Distance to planned metro station in miles (ρ)	**0.012** (0.022)	**-0.077***** (0.020)	**-0.245***** (0.055)	**-0.081*** (0.032)	**-0.104***** (0.029)	**0.645***** (0.089)
Bedrooms	0.164*** (0.003)	0.162*** (0.002)				
Age of structure	-0.001*** (0.000)	-0.003*** (0.000)	-0.001*** (0.000)	-0.004*** (0.000)	-0.002*** (0.000)	0.001* (0.000)
Pets allowance	-0.013** (0.005)	0.010* (0.005)	0.016 (0.015)	0.017* (0.007)	0.020** (0.006)	0.053 (0.029)
Monthly lease	-0.013 (0.007)	-0.047 (0.006)	0.093*** (0.021)	-0.136*** (0.010)	0.128*** (0.009)	-0.000 (0.035)
Affordability program	-0.077*** (0.005)	-0.069*** (0.005)	0.015 (0.020)	-0.160*** (0.009)	0.033*** (0.007)	0.045 (0.024)
Gym center	0.142*** (0.006)	0.101*** (0.006)	0.044* (0.020)	0.129*** (0.009)	0.042*** (0.008)	0.138*** (0.039)
Laundry room	-0.079*** (0.005)	-0.022*** (0.004)	-0.005 (0.013)	-0.061*** (0.007)	0.027*** (0.006)	0.358*** (0.027)
Parking	0.001 (0.008)	0.062*** (0.007)	0.256*** (0.029)	0.057*** (0.012)	0.087*** (0.009)	0.180*** (0.031)
Storage service	0.106*** (0.005)	0.114*** (0.004)	0.112*** (0.012)	0.120*** (0.007)	0.065*** (0.006)	0.057 (0.031)
Swimming pool	0.075*** (0.005)	0.048*** (0.005)	-0.026 (0.019)	0.014 (0.008)	0.070*** (0.006)	-0.108*** (0.026)
distance from shopping center in mile	-0.204*** (0.023)	-0.103*** (0.019)	-0.255*** (0.045)	-0.182*** (0.032)	0.034 (0.027)	-0.166* (0.079)
Log of school quality	1.033*** (0.045)	0.118** (0.038)	-0.334** (0.121)	0.254*** (0.058)	-0.115* (0.051)	-1.745*** (0.312)
Log of crime rate in neighborhood	0.010** (0.003)	-0.057*** (0.002)	0.001 (0.007)	-0.071*** (0.004)	-0.005 (0.004)	-0.071 (0.037)
Proportion of African America	0.156*** (0.027)	-0.043 (0.022)	-0.288*** (0.053)	-0.103** (0.032)	0.035 (0.034)	-0.591*** (0.135)

Intercept (α)	2.356*** (0.202)	7.168*** (0.172)	8.518*** (0.562)	7.046*** (0.266)	7.842*** (0.227)	15.118*** (1.476)
Control structure type	Yes	Yes	Yes	Yes	Yes	Yes
# of observations	15156	18094	1494	8857	6725	959
R2=	0.399	0.401	0.393	0.387	0.474	0.614
Adjusted R-squared	0.399	0.400	0.387	0.386	0.472	0.607

The dependent variable is the natural logarithm of the rent of each unit
*** p=.001(two-tailed). **p=0.01(two-tailed). *p=0.05(two-tailed).

Discussion and conclusion

Based on a two-year panel rental housing dataset, I estimated changes in rents along a new transit corridor using first-difference approaches and difference-in-difference approaches. The results show that the rents of units within a half mile from the nearest planned Purple Line stations increase more than their counterparts in the county, on average. Specifically, two-bedroom units, three-bedroom units, and four-bedroom units see their rents increase more than their counterparts in the control group. I also estimated the rent gradient with respect to the distance to a station using the parameterized distance decay regressions. On average, the rent price gradient is -0.077 in 2018.

I addressed two deeper questions prompted by the results. First, why does the rent of multifamily units within a half mile of the planned stations increase more compared with their counterparts in the control group? This is a particularly interesting question since there is not yet an actual improvement in terms of accessibility via the planned Metro stations. There is no upward shift of the tenant utility function prior to the line opening. A possible explanation for this is that the planned rail transit stations provide an anticipation effect for apartment property value. Plans for the Purple Line convey information about when and where light rail service will be provided, that information leads to expectations of higher apartment property value in the future. And the increased property value needs to be achieve through increasing tenants' rents. In other words, the LRT future premium was capitalized into the current apartment properties, thus increased tenants' paying rents. Turning to the tenants' perspective, why would they are willing to pay higher rents now? I suggest that this can be attributed to nontrivial moving costs. If moving costs were trivial and renters were able to move frequently and without cost, they would likely not be willing to pay more in rents for units that offer access to transit service until sometime in the future. But because moving is not costless, renters might be willing to pay more for units that will provide this benefit in the future. For this reason, landowners may be able to charge more for rents in the preservice period.

But why just 2-bedroom, 3-bedroom, and 4-bedroom units increased rents, while 0-bedroom and 1-bedroom units in the half-mile treated group did not see statistically significant increases in rent? One possible answer is that households in zero-bedroom units and one-bedroom units have higher liquidity than households in units of two bedrooms, three bedrooms, and four bedrooms. Households in zero-bedroom units and one-bedroom units are more likely to be unmarried, and they may have fewer possessions, and be less likely to have children, and therefore, they would face lower moving costs. These households would easily move out of their original units to find alternative accommodations if their landlords raise their rents based on anticipation of the opening of the Purple Line.

Conclusions and policy implications

In this study I have used an unusually detailed data set on multifamily units and rents to explore the impacts of a planned investment in light rail on multi-family rents. Using difference-in-difference and first difference approaches I find that multi-family rents did increase in the preservice period for two-, three-, and four-bedroom units within one-half mile of a planned station. I am not aware of any previous studies that have similar or conflicting results.

I believe that the results have important implications for policy makers in the Purple Line corridor and other places planning similar light rail investments. Some of these policy implications are now fairly well known but others have heretofore not been addressed in the scholarly literature. First, the study adds to the mounting evidence that investments in transit tend to increase housing prices and rents in station areas. Plans for LRT convey information about when and where the light rail service will be provided. That information leads to expectations of higher apartment property value in the future, which was capitalized into the current apartment property value. The apartment property value that capitalized the future LRT service premium rises and the owner of the property achieves the increment through uplifting tenants' rents. In other words, the LRT future premium was capitalized into the current apartment properties, thus push tenants' paying rents up. This lends credibility to the concern that LRT investments have the potential to displace residents that cannot afford higher housing prices and rents. That is, concerns about transit-induced gentrification are valid. Second, transit-induced increases in housing prices and rents can and do occur in the preservice period. This suggests that anti-displacement policies need to be adopted and enacted long before transit service begins. It also means that land value capture strategies must also be in place well ahead of transit service operations. Otherwise, the increases in land value will have already been captured by land owners. Third, I find that rent increases in anticipation of transit service are more likely for multifamily units with more bedrooms than multifamily units with less than two bedrooms. This suggests that families that rent—especially larger low-income families that rent—are most vulnerable to displacement pressures even before transit service begins. Finally, I find that rents increase for apartments located within a half mile of a proposed station but for those not between a half mile and one mile of a proposed station. This suggests that anti displacement and land value capture strategies of the two counties and PLCC should focus first on locations with multiple bedrooms within a half mile of proposed stations.

In sum, I believe that the analysis adds nuance to a growing body of literature on the effects on transportation investments on land and housing markets. It helps to confirm principles of urban economics, real estate development, and urban planning on the relationship between accessibility and property values. More importantly, however, this specific analysis on the effects of anticipated investments in light rail on multifamily rents add information that can lead to more carefully designed land value and anti-displacement policies.

4. When and Where Do Home Values Increase in Response to Planned Light Rail Construction?

Introduction

Property valuation is needed for property taxation. It is critical to accurately measure the changes in property values brought about by public investment in order to design revenue instruments to harness those value increments. Considered a popular public investment to address congestion, reduce greenhouse gas emissions, and revitalize decayed urban centers, the light rail transit (LRT) project – even in its pre-service period – is believed to have increased the values of single-family properties near planned transit stations. Better understanding of this impact and to how far it extends will help tax assessors identify when and where to adjust property tax assessments. In addition, further investigation into the timing of value added by transit investments and spatial variability of the impact could draw implications for benefit-cost analysis of transit investment. Despite the importance of the topic, a limited number of studies have explored this.

Towards this end, I explore in this paper the effects of an anticipated investment in LRT on single family housing prices in Montgomery County, Maryland, during the preservice period. My findings contribute to the growing body of literature that investments in LRT can affect land and housing markets, where conditions are suitable, even before transit service begins. Perhaps more importantly, the findings offer insights into where and when affordable housing policies and land value capture strategies should be implemented to stimulate transit-supportive development and avoid displacement.

This essay is organized as follows. The next section reviews literature on previous research that examines the impact of the pre-service period of LRT on property values. After describing the research methods, I introduce the data and present empirical results. To further confirm the impact arising in the engineering phase and the 1.5-mile catchment area, I conduct a difference-in-difference method and report the results. I conclude this essay with contributions, findings, and policy implications.

<u>Literature review</u>

A large number of studies evaluate the impacts of light rail infrastructure investments on property values, but few focus on impacts before service begins. With few exceptions (Gatzlaff and Smith 1993; Devaux et al. 2017) most of those confirm that light rail investments indeed increase property values even before service begins. To implement policies designed to encourage appropriate development, to use land value capture instruments, and to prevent displacement, action must thus be taken well before transit service begins. Still, findings are mixed on where and when such impacts take place.

Although research has shown that investment in LRT can increase property values even before service begins, the exact timing of this impact remains uncertain. In Chicago's Midway corridor, McMillen and McDonald (2004) found that residential land values increased six years before service began. Knaap et al. (2001), found that light rail investments in the Portland metropolitan area increased vacant land values when the station locations were first announced. In St. Paul, Minnesota's Green Line corridor, Cao and Lou (2017) found that single family housing prices increased after funding for the line was confirmed. In the case of Santiago's Line #4, Agostini and Palmucci (2008) found that apartment prices increased from the moment construction was announced and the basic engineering project was unveiled. Similarly, Yen et al. (2018) found that housing prices began to increase immediately after the announcement of a light rail project in Queensland, Australia. In the Charlotte-Mecklenburg case, peak impact occurred when the transit investment was first announced (Ke and Gkritza, 2019). In addition, Dubé et al. (2018) found that the impact of a LRT investment in Dijon, France occurred at the start of the construction phase. In an examination of the impacts of a light rail investment in Charlotte, North Carolina, however, Yan et al (2012) found that single-family home prices did not increase until operation began.

Many studies have also found, however, that impacts on property values are not uniform throughout the transit corridor. Nearly every study found that the effects on property values decreased with distance from the station. McMillen and McDonald (2004), who examine how these distance-to-station gradients vary over time, found that these gradients also vary across stations. More specifically, they found that price effects occur near stations that have a parking garage or parking lot (McMillen and McDonald, 2004). Dubé et al. (2018) explored the impacts of a conversation of a bus route into a light rail route in Dijon, France and found that price effects concentrated near stations in the center of the city. Devaux et al. (2017) found no aggregate effect of a light rail investment in Lava, Canada, but found that effects differ extensively by station, with the effects greatest effects in places that are densely developed. In their analysis of the Miami metro rail, Gatzlaff and Smith (1993) also found overall price effects to be weak, but that the effects varied widely between stations, with the largest effects in high-income residential areas. Despite these repeated findings of differences in effects, there has been little exploration of why these differences occur between station areas or why these differences matter.

Some of these differences in findings no doubt reflect differences in methods. Hedonic price models, difference-in-difference methods, and repeat-sales approaches are most frequently employed. Hedonic modeling typically involves the estimation of a distance-to-station gradient or a difference in prices among those properties located within a given radius of a station (e.g., Cao and Lou.2017, Mulley and Tsai. 2016). In some cases, the value of those gradients, or discrete premiums, are examined over time (McMillen and McDonald (2004). Difference-in-difference approaches involve observations in a treated group and a control group, and examine whether the gradient, or discrete premium, changes significantly over time (Knaap et al. 2001, Yen et al. 2018). However, when using a hedonic modelling approach and difference-in-difference approach, it is important to include as many control variables in the models to avoid omitted variable bias. Repeat-sales methods include only properties that are sold multiple times in treated and control groups, assume that properties characteristics remain constant aver time, and track the station price gradient over time (e.g. McMillen and McDonald (2004)). The repeat-sales approach may be free from omitted-variable bias, given the assumption that property attributes remain constant over time, but the repeat-sales approach only includes observations that are sold more than two times. The repeated-sale observations thus may not be representative of all properties due to selection bias.

In this study, I intend to fill two gaps in the literature by (1) providing new evidence on if and when price capitalization occurs in the preservice period and by (2) examining when and why impacts differ between station areas. To assure that the results are robust, I use both hedonic and repeat-sales techniques.

Context

The Washington metropolitan area has one of the most extensive metrorail systems in the United States. As shown in Figure 4.1, a new circumferential "Purple" light rail line will feature 17 anticipated new stations and connect four of the existing Metro rail stations. The 16.2-mile line, now under construction, is estimated to serve 74,000, riders per day by 2040[15]. Although the line will traverse two Maryland counties – Montgomery and Prince George's – I limit my analysis to the 10 stations in Montgomery County, the home of over one million residents (U.S. Census Bureau, 2017) and several U.S. federal agencies. Even more, Montgomery County has many innovative affordable housing programs and its political leadership is seriously committed to preventing gentrification and displacement.

The Purple Line project weathered two decades of political, legal, and financial battles. Important milestones include the following[16].

1. General layout: The general layout of the Purple Line addition to the Metro network has been known since 2003;

2. Environmental impact study: On October 20, 2008, a draft environmental

[15] Source: http://mta.maryland.gov. Governor O'Malley Announces Purple Line Receives Federal Environmental Approval.
[16] Source: https://www.purplelinemd.com/about-the-project/overview

impact study was issued;

3. Preliminary engineering phase: The Federal Transportation Administration gave its approval for the Purple Line to enter its preliminary engineering phase at the end of 2011 (Action Committee for Transit, 2020).

4. Start of construction: Construction kicked off in August 2017 (Action Committee for Transit, 2020).

5. Opening: The Purple Line is expected to be in service in 2022 (Action Committee for Transit, 2020).

As in other light rail projects, confidence in whether the Purple Line would ever be built ebbed and waned over time. There is anecdotal evidence that housing prices and rents have started to rise, but it is not known when the market began to anticipate that the line would actually be built and when and where housing prices would start to rise. This is what I explore here.

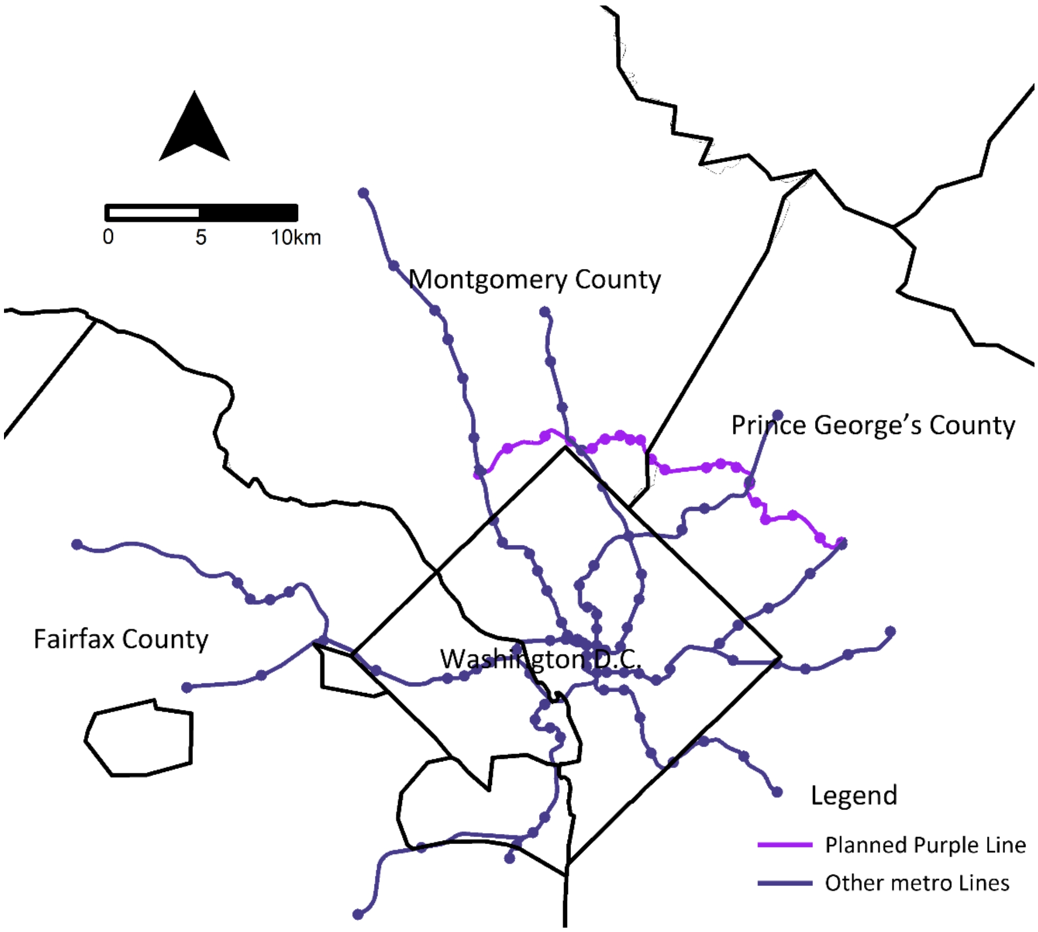

Figure 4.1: Proposed Purple Line and existing Metro rail system

The population of Montgomery County is diverse; more than one in three residents are foreign-born (U.S. Census Bureau, 2017), many from Central America and Asia. Socioeconomic diversity is especially pronounced along the Purple Line corridor (see Figure 4.2). In the census tracts along the corridor in Montgomery County, median home values range from just under $400,000 to over $800,000; owner occupied proportions range from 57% to 100%; poverty rates range from 0.3% to 25.6%; foreign-born proportions range from 6.5% to 69% (U.S. Census Bureau, 2017). High income, predominantly white neighborhoods are concentrated on the west side of the corridor while low income, largely minority neighborhoods are concentrated on the east side of the corridor (Figure 4.2).

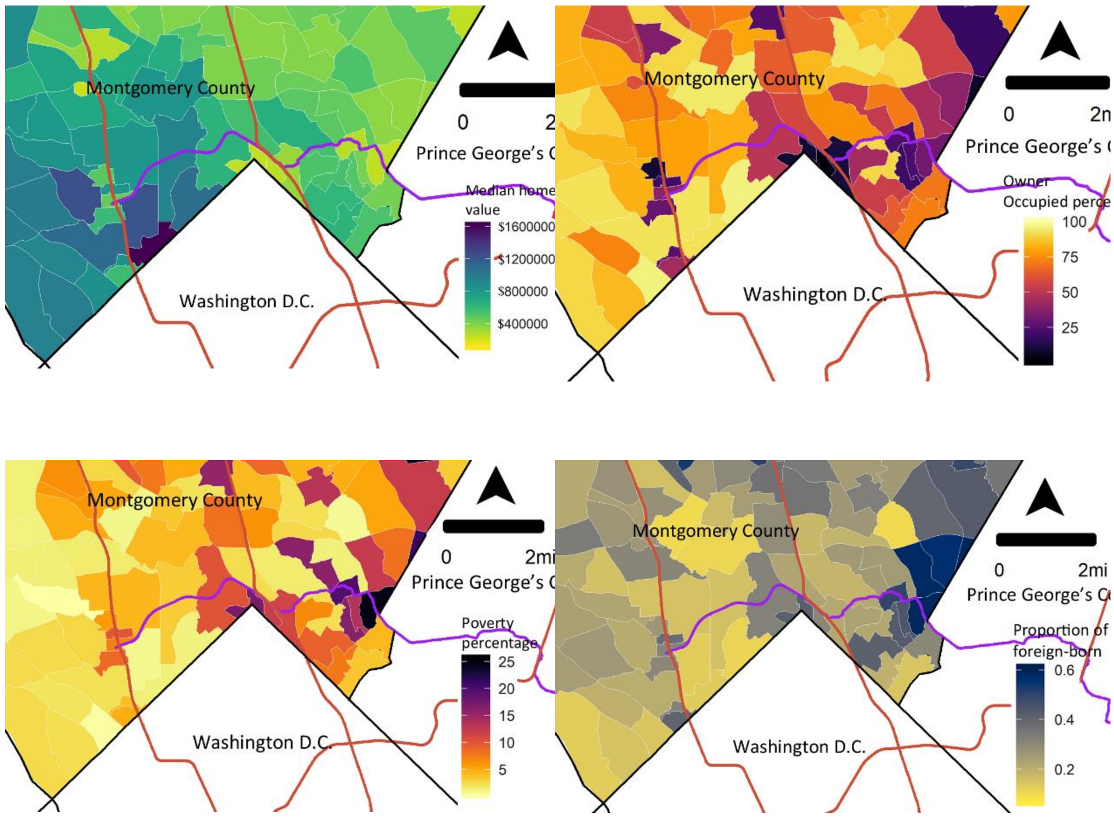

Figure 4.2: Demographics and housing market characteristics in the rail transit corridor (left top: median home value; right top: owner occupied percentage; left bottom: poverty percentage; right bottom: proportion of foreign-born residents)

Given the demographic and housing market conditions in the Purple Line corridor, political leaders in Montgomery County are concerned about gentrification and displacement. The county also has multiple innovative affordable housing programs to prevent this from happening. The question remains, however, when and where should these programs be deployed?

Data

My data come from three primary sources: Maryland Propertyview, Maryland State Indicators Data, and the U.S. Census. From Maryland Propertyview, I collected sales data for the years 2002 through 2017. I have created an interactive map for the data:https://rpubs.com/xqpeng/MassTransitImpactsonLandandPropertyPreservice. Each observation includes information on sales price, sale date, lot size, square footage, built year, story, basement, structure type, tax account ID, longitude, and latitude information. From the location data, I computed the distance from each property to the nearest Purple Line station.

Using the Maryland State Education Indicators, I extracted information on average math and reading scores for fifth graders for each public elementary school in the county. I then assigned these attributes to each observation for the appropriate elementary school. From the 2010 U.S. Census, I collected information on neighborhood characteristics and assigned them to each observation.

For the analysis that follows I selected property transactions that met the following criteria:

a. A property falls in the 1.5 miles catchment area of the Purple Line stations[17].

b. The distance between the property and the closest Purple Line station is less than the distance to other metro stations.

[17] Following some previous studies (McDonald and Osuji, 1995; McMillen and McDonald, 2004; Yan et al, 2012), I select 1.5 miles as the buffer distance of metro stations.

c. The method of conveyance at the time of sale is an arms-length transfer of a single parcel.

After applying these criteria, there remained 15,260 transactions in the full sample and 3,087 observations in the repeat-sales sample. Figure 4.3 shows the location of all individual property transactions in this study. As shown in Figure 4.3, dots in black indicate property sale records that occurred near anticipated new Purple Line stations, while gray dots represent property sale records that took place near established Purple Line stations.

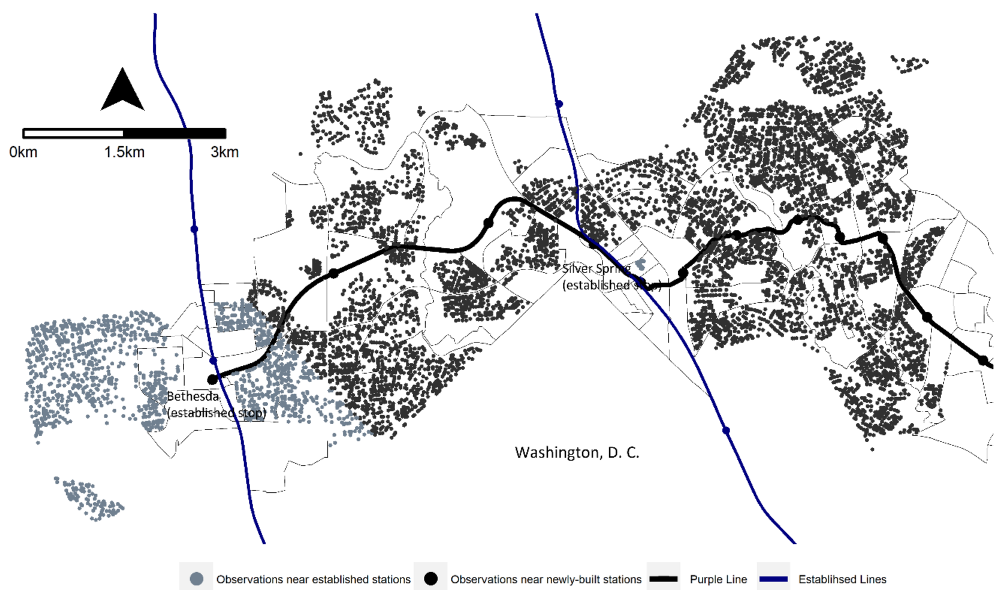

Figure 4.3: Observations near established LRT and anticipated new stations

Table 4.1 lists variables included in the analysis.

Table 4.1: Definition of variables

Variable Name	**Definition**
Price	Sale price of a property
Lot size	Land size of tax lot
Building area	Total square footage of livable space

Age	Age of building structure
More than one story	Dummy variable indicating that a building structure has more than one story.
Brick	Dummy variable indicating that the material of a building structure is brick.
Basement	Dummy variable indicating that a property has a basement.
Grade of structure	Structure grade assigned by the Maryland Department of Planning. Grades vary from X to Y with larger numbers indicate higher quality.
School quality	Average 5th grade reading average score per school. The score varies from 0 to 100. The larger the score is, the better quality.
Proportion Black	Proportion of census tract residents that are Black
Proportion Vacant	Proportion housing units in census tract that is vacant.
Distance to transit station	Distance from property to the nearest Purple Line station

I present descriptive statistics in Table 4.2. Column 1 presents statistics for the full sample and column 2 presents statistics for the repeat sales sample. Note that the mean values for the repeat-sales sample are very close to those of the full sample. This suggests the repeat-sales data are relatively free of sample selection bias.

Table 4.2: Descriptive statistics

Variable	Full Sample	Repeat Sales
Price	$525,372 (260,921) [$30000, $1376000]	$ 499,482 (245,229) [$90000, $ 1373900]
Lot size (acres)	0.184 (0.094) [0.015, 1.796]	0.172 (0.0803) [0.019, 1.020]
Building area (square feet)	1703 (657) [512, 7898]	1675 (665) [562, 5886]
Age	60.9 (19.5) [0, 164]	63.1 (20.1) [0, 160]
More than one story	0.758	0.785

Brick	0.624	0.601
Basement	0.899	0.91
Grade of structure	4.68 (0.852) [2.000, 9.000]	4.65 (0.866) [2.000, 9.000]
School quality	89.2 (5.84) [62.6, 99.00]	89.2 (5.85) [62.6, 99.00]
Proportion Black in census tract	0.173 (0.128) [0.005, 0.481]	0.177 (0.125) [0.005, 0.481]
Proportion vacant housing in census tract	0.0364 (0.0242) [0.000, 0.1914]	0.0397 (0.0278) [0.000, 0.1914]
Distance to transit station (Mile)	0.661 (0.343) [0.033, 1.499]	0.656 (0.342) [0.039, 1.500]
Near anticipated new station	0.846	0.862

Note: The sample average is presented in all cells. For continuous variables, the standard deviation is presented in parentheses, followed by the minimum and maximum in brackets. The full sample has 15,260, and the repeat-sales sample has 3,087 observations.

Econometric Methods

I applied a hedonic price and repeat-sales approach to estimate the time-specific effect of the Purple Line on housing prices. The hedonic approach estimates the slope of the housing price gradient (with respect to distance from the nearest rail station) over time. I used the repeat-sales approach to validate the results of the hedonic approach. Finally, I conducted a difference-in-difference method to check whether there is a significant home price increase on average within the 1.5-mile catchment area of the new LRT construction compared with homes outside the catchment area.

Hedonic method

To estimate the time path of the transit station price gradient, I first applied the hedonic approach. The hedonic equation is designed to account for the temporal variation in the transit gradients flexibly and takes the following form.

$$\ln P_{it} = \alpha_t + \theta_t d_i + \beta' X_i + u_{it} \tag{16}$$

where i indicates individual transaction, t indicates transaction time. P_{it} indicates the sales price of property, while d_i indicates the distance of property i from its nearest Purple Line station. X_i is a set of property i's attributes, while u_{it} is an error term. α_t are coefficients for a set of dummy variables indicating the time period of the sale. The set of α_t constitutes the house price index (2002 quarter 1 is the base group). Similarly, the set of coefficients for the distance from the property to its nearest Purple Line station forms the estimated rail transit station gradient index (θ_t). It is worth noting that θ_t can vary across time t, which allows us to trace the time path of the effect of distance from the nearest Purple Line station on housing prices.

McMillen and McDonald's study shows that the estimated transit station index θ_t could be highly volatile when the index is permitted to vary across quarters (McMillen and McDonald, 2004). To acquire a relatively stable estimated of θ_t, I assumed that the gradients remain constant for several successive years, with breakpoints between some years. Maintaining similar techniques for selecting breakpoints in previous studies (Knaap et al., 2001; McMillen and McDonald, 2004; Dubé et al., 2018) and based on the results of standard regression with random gradients across quarters, I broke the whole time period into two sub-periods: the pre-engineering phase (2002- 2011) and the engineering phase (2012-2017). I allowed the coefficient for distance to the nearest Purple Line station to vary across quarters or two phases, while other coefficients are assumed to keep constant over time.

Repeat-sales method

Next, I applied the repeat-sales method to estimate the price gradient regarding distance to the nearest Purple Line station. The repeat-sales method was proposed by Bailey et al. (1963) and was utilized to construct house price indexes. As it is mentioned by McMillen and McDonald (2004), the repeat-sales method can be derived from Equation (16). If a property is sold more than two times during a sample period, the change in ln P can be expressed as

$$\ln P_{it} - \ln P_{is} = \alpha_t - \alpha_s + d_i(\theta_t - \theta_s) + u_{it} - u_{is} \tag{17}$$

where t is the time of the next sale, s is the time of the previous sale. Equation (17) indicates that the percentage change in the price of a property is a function of an implicit set of time-specific explanatory variables. The time-specific coefficients (θ_s) for the distance to the planned Purple Line station exhibit how the gradient value varies over time. Similarly, the set of coefficients $\alpha_1,..., \alpha_t$ constitutes the housing price index.

Given the assumption that coefficients of explanatory variables and omitted variables are constant over time, the repeat-sales estimates are potentially subject to less bias than standard hedonic estimates. However, the reduction in bias comes at the cost of possible selection bias and a substantial decrease in sample size (McMillen and Dombrow, 2001; McMillen, 2003; McMillen and McDonald, 2004). In this study, the sample size reduces from 15,260 transactions in the full sample to 3,087 of the repeat-sales sample. Fortunately, the repeat-sales estimator in this study seems free of sample selection bias since the repeat-sales sample is comparable to the full sample (see Table 4.2).

Difference-in-difference approach

To check whether there is a significant home price increase on average within the 1.5-mile catchment area of the new LRT construction compared with homes outside the catchment area, I used a difference-in-difference (DID) approach. The equation is:

$$\ln Pi = \alpha + \gamma * Treatedi + \delta * Afteri + \lambda * (Treatedi * Afteri) + \beta jXi + ui \qquad (18)$$

where i indicates individual transaction; P_i is the sales price of property i; $Treated_i$ is a dummy variable that equals 1 if the parcel is located within 1.5 miles of a planned station; $After_i$ is a dummy variable set to 1 if the parcel was sold after 2012; u_i is an error term. X_i is a vector of property characteristics including lot size, building area, story, basement, structure quality, structure material, quality of a school, neighborhood crime rate, etc. The interactive variables ($Treated_i * After_i$) capture the difference in property values between parcels located within 1.5 miles of a planned station and parcels in other areas before and after 2012.

The dependent variable is the log of the single-family home sale price. The regression equation includes a time dummy variable indicating whether the sale happened before or after 2012; a treated effect dummy indicating whether an observation falls in the catchment area of the Purple Line Metro station; the interaction term of time dummy and treated dummy; and other control variables.

Empirical Results

I estimated Equation (16) for observations near anticipated new stations, where the discrete periods are quarters. I do not report the results for each quarter in a table but illustrate the price gradients with respect to distance to the corresponding nearest anticipated station across years and quarters in Figure 4.4. In Figure 4.4, the horizontal axis indicates year and quarter, while the vertical axis indicates the estimated price gradients with respect to distance to the nearest station, θ_t. As shown in Figure 4.4, the price gradient is volatile, but turns negative after 2012 quarter 1, the beginning of the engineering phase. Further, the gradient of the relationship between housing price and distance from anticipated new stations changed significantly in 2012, the year the project entered the engineering phase[18]. In contrast, the gradient on the relationship between housing prices and distance from established Metro stations (see Figure 4.5)

[18] I used difference-in-difference model to identify which milestone date should be selected in terms of model fitting. The results show that the beginning of the pre-engineering phase should be selected.

does not change. Figure 4.5 shows that the price gradient with respect to distance to established stations is less than zero before and after 2012; there is no significant change after 2012. This means that, as my estimates of Equation (15) suggest, the announcement of the Purple Line had no effect on single-family home values near established Metro stations.

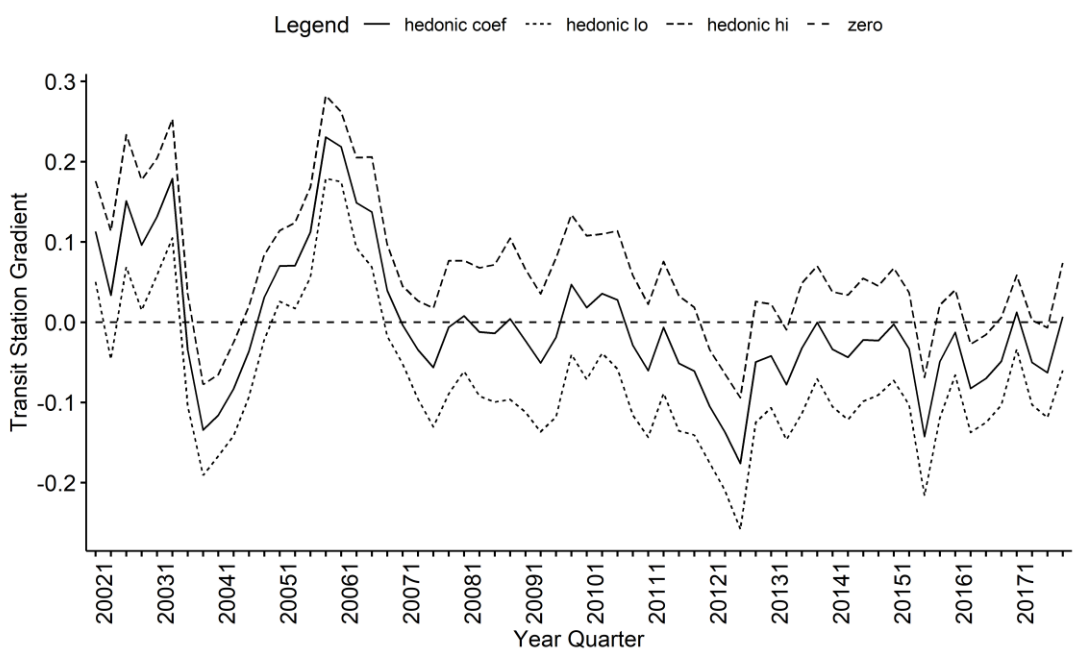

Figure 4.4: Hedonic transit station gradient index for observations near anticipated new stations

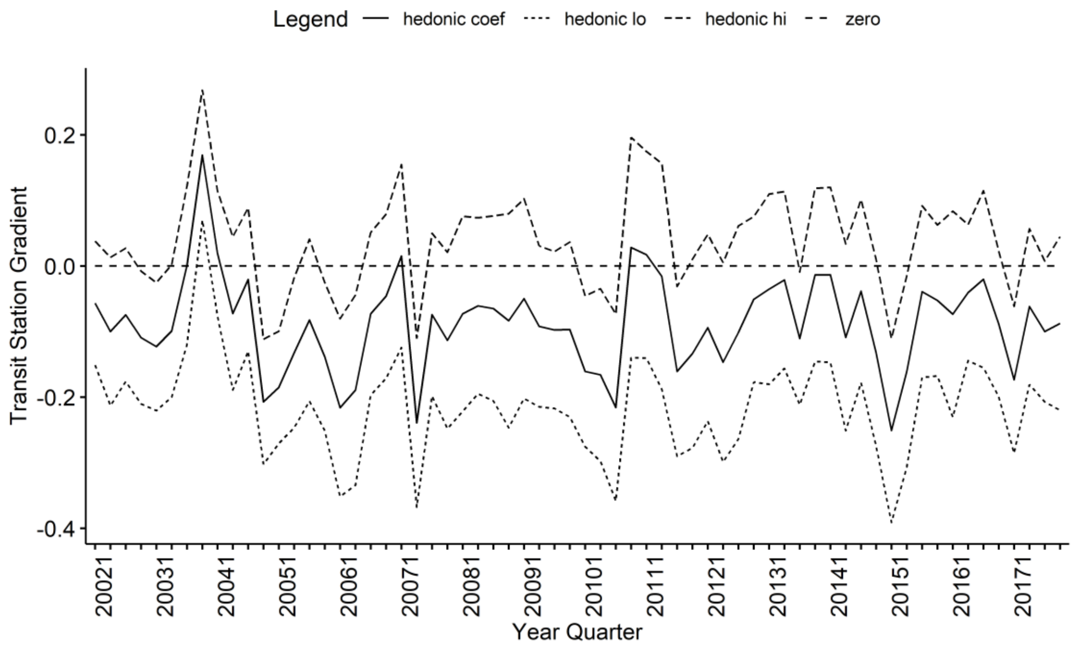

Figure 4.5: Hedonic transit station gradient index for observations near established stations

Since the price gradients regarding distance to both anticipated new stations and established stations are highly volatile across quarters and 2012 represents a milestone year for the Purple Line project[19], I broke the pre-service period into two phases: 2002-2011, the pre-engineering phase, and 2012-2017, the engineering phase. I estimated hedonic Equation (16) for observations near anticipated new stations, where the discrete periods are two phases: the pre-engineering phase and the engineering phase. For each phase, I estimated distinct price gradients. Column 1 in Table 4.3 reports the results of the hedonic models for the observations closer to anticipated new stations. As shown, almost all the coefficients on property and

[19] I have tried multiple milestones and finally I conclude the critical milestone with the beginning year of engineering phase (2012).

neighborhood attributes have expected signs. Specifically, the sale price of a single-family property is positively associated with lot size, building area, presence of basement, structure grade, and performance of its assigned public school. Some characteristics are negatively associated with the sales price of a property, such as whether the house has more than one story, and proportion of the population in a census tract that is African American.

The price gradient with respect to distance from anticipated new stations is 0.03 from 2002 to 2011, however, it becomes -0.058 in the 2012-2017 period. This means that homes close to anticipated new stations were valued, on average, 5.8 percent higher than homes located an additional mile away from the station. The magnitude of the price gradient is less than what McMillen and McDonald (2004) found in the Chicago case, who found a 7.4 percent increase. To examine whether the coefficient of the price gradient from 2002 to 2011 is equal to the coefficient of the price gradient in the 2012-2017 period I used an F statistic test. The F statistic for anticipated new stations is 45.82, which rejects the null hypothesis. This means that the price gradient changed significantly across two periods and suggests that the anticipation of the Purple Line had had a significant impact on home values near anticipated new stations.

These results are illustrated in Figure 4.6. The solid line in Figure 4.6 depicts the estimated price gradients, which vary across years and quarters. The dashed line in Figure 4.6 illustrates the average price gradient, which varies across the two phases. As shown in Figure 4.6, the price gradient with respect to distance to the anticipated new stations is significantly positive before 2012 but significantly negative after 2012 quarter 1. This means that homes nearest anticipated new stations were valued more than homes located further from these station after the Purple Line project entered the engineering phase.

Table 4.3: Regression results

	Hedonic (anticipated new stations) (1)		Hedonic (established stations) (2)		Repeat-sales (anticipated new stations) (3)		Repeat-sales(established stations) (4)	
Variable	Coef.	Std.Error	Coef.	Std.Error	Coef.	Std.Error	Coef.	Std.Error
Constant	8.002***	0.152	8.308***	0.379				
ln lot size	0.107***	0.004	0.110***	0.012				
ln building area	0.399***	0.008	0.343***	0.015				
Age	0.0007***	0.0001	0.0005*	0.0002				
More than one story	-0.005	0.004	-0.031*	0.011				
Brick	-0.002	0.003	0.014*	0.008				
Basement	0.080***	0.006	0.067***	0.011				
Grade of structure	0.254***	0.003	0.138***	0.006				
Ln school quality	0.091***	0.030	0.350***	0.077				
Proportion Black	-0.411***	0.019	-0.216***	0.075				
Proportion Vacant	0.307***	0.086	0.102	0.305				

Distance to transit station, based(2002-2011)	0.030***	0.003	-0.144***	0.004	-0.0589	0.154	-0.0393	0.315
Distance to transit station, based(2012-2017)	-0.058***	0.003	-0.145***	0.006	-0.0949 ^	0.0595	-0.0787	0.178
Control time fixed effect	Yes		Yes		Yes		Yes	
Test for equality of coefficient of distance(F statistic)	45.82 Reject null hypothesis		0.00019 Cannot reject null hypothesis		37.27 Reject null hypothesis		0.012 Cannot reject null hypothesis	
# of observations	12915		2345		2660		427	
R^2=	0.822		0.815		0.714		0.606	

The natural logarithm of the sales price serves as the dependent variable. Note: the regression includes 16 dummy variables indicating the year of sale.
***p=.001(two-tailed). **p=0.01(two-tailed). *p=0.05(two-tailed). ^p=0.1(two-tailed)

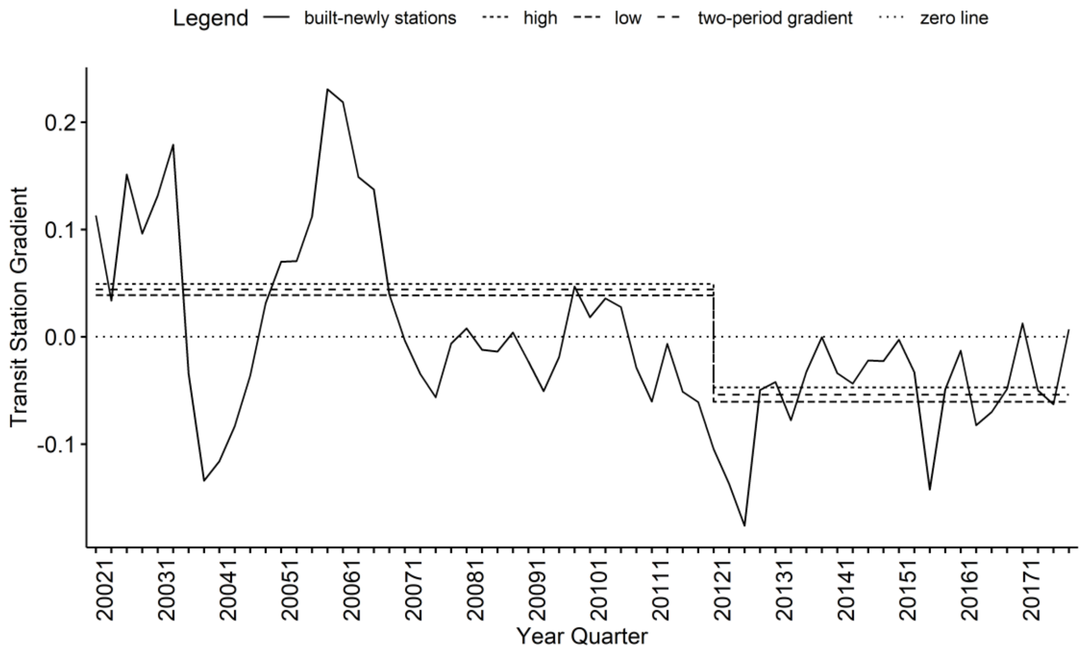

Figure 4.6: Hedonic transit station gradient index of anticipated new stations group across quarters and two periods (before the engineering phase and the engineering phase)

Next, I estimated Equation (16) for observations near established Metro stations and report the results in Column (2) of Table 4.3, where there are once again two discrete periods: the pre-engineering phase and the engineering phase. Again, most of the attributes of the property and characteristics of the neighborhood had the expected effects. The estimated price gradient with respect to distance from existing metro stations is -14.4 percent in the period 2002-2011. The price gradient rises slightly to -14.5 percent in the 2012-2017 period, but this change was significant (see Figure 4.7). To test whether the coefficient of the price gradient with respect to distance from 2002 to 2011 is equal to the coefficient of the price gradient with respect to distance in the 2012-2017 period, I again performed an F test. The F statistic for established stations is 0.00019, which does not reject the null hypothesis. This means that the price gradient did not change significantly across the two periods. This suggests that the anticipation of the Purple Line had no significant impact on home values near established metro stations. The solid line in Figure 4.7 illustrates the estimated price gradients, which vary across years and quarters. The dashed line in Figure 4.7 illustrates the average price gradients, which vary slightly across the two phases. As shown, the dashed line changes little before and after 2012, or at any point in the preservice period.

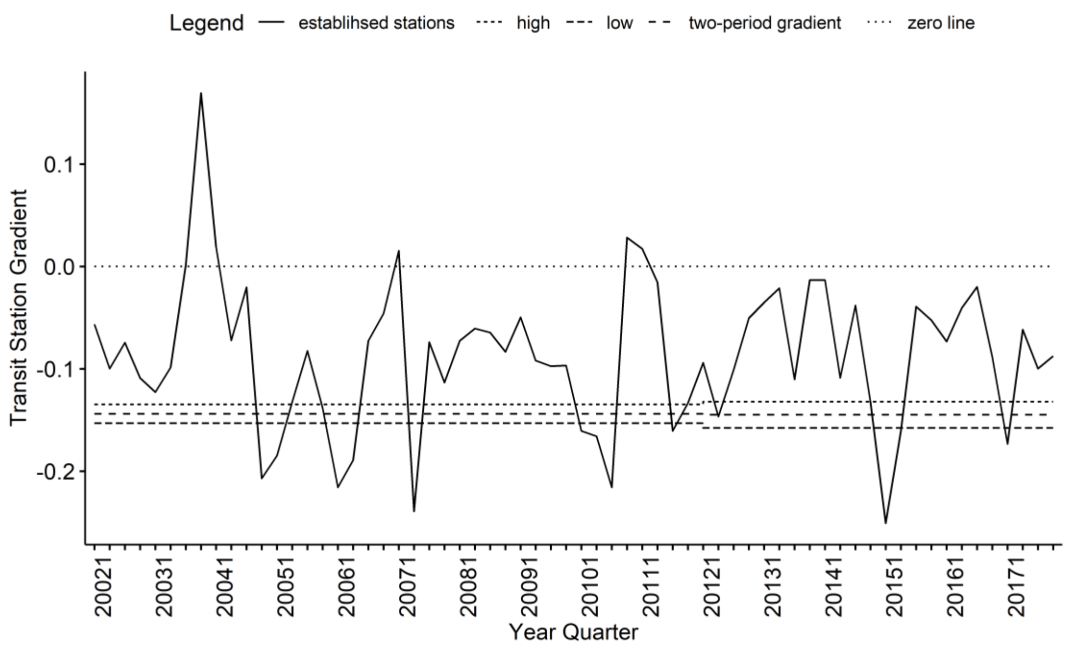

Figure 4.7: Hedonic transit station gradient index of established stations group across quarters and two periods (before the engineering phase and the engineering phase

I next used a repeat-sales approach to validate the aforementioned hedonic results. I used repeat-sales data to estimate Equation (17) for observations located close to anticipated new stations and report results in Column (3) of Table 4.3, where I allowed the price gradient to vary across two phases, the pre-engineering phase and the engineering phase. The price gradient is an insignificant -5.89 percent in the pre-engineering phase, and decreases to a significant -9.49 percent (at the 90 percent level) in the engineering phase. To test whether the coefficient of the price gradient from 2002 to 2011 is equal to the coefficient of the price gradient in the 2012-2017 period I again performed an F test. The F statistic for anticipated new stations is 37.27, which rejects the null hypothesis. This means that the price gradient changed significantly across the two periods. This suggests that the anticipation of the Purple Line had a significant impact on home values near anticipated new stations, and validates the previous hedonic results.

For observations located close to established Metro stations, the repeat-sales results show that there was a decrease of the price gradient before and after 2012, but again the change is insignificant. To test whether the coefficient of the price gradient from 2002 to 2011 is equal to the coefficient of the price gradient in the 2012-2017 period I again performed an F test. The F statistic for established stations is 0.012, which cannot reject the null hypothesis. This means that the price gradient did not change significantly across the two periods. This again suggests that the anticipation of the Purple Line had an insignificant impact on home values near established Metro stations, and again validates the hedonic results. This indicates that the effect of the Purple Line project is not significant after the Purple Line project progressed into the engineering phase for homes near established stations. As evident in Column (4) of Table 4.3, the price gradient is -3.93 percent in the pre-engineering phase and becomes -7.87 percent (not statistically significant) in the engineering phase. This means that the Purple Line project makes proximity to established stations more desirable for single-family residents than before 2012, but not by a significant degree.

To explicitly compare the gradients of the anticipated new station group and the established station group, Figure 4.8 shows that the gradient of the anticipated new transit station group roughly changes from positive to negative and remains negative after 2012, while the gradient of the established station group is less than zero and does not change after 2012. This confirms that the premium of proximity to LRT for properties near anticipated new stations is greater than for properties near established stations after 2012.

The series of coefficients $\alpha\_{1},\ldots,\alpha\_{T}$ in Equation (17) forms the price index of overall housing market. I calculate the price index and plot it in Figure 4.9 to illustrate the Montgomery County single family housing price movement from 2002 to 2017. It shows that the housing price, generally speaking, kept rising after 2012.

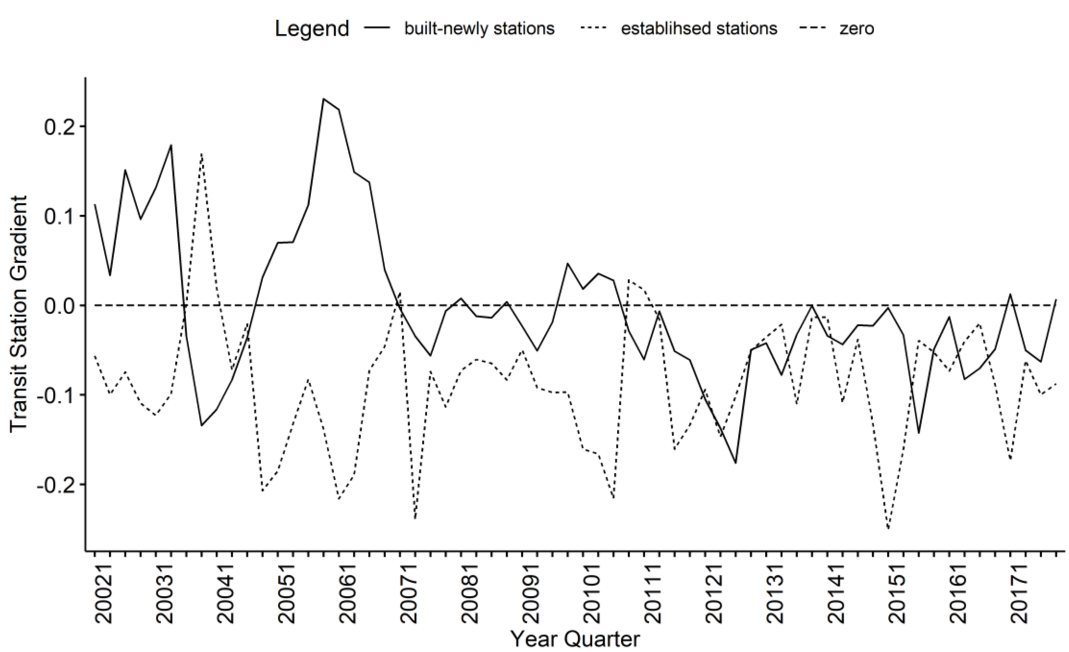

Figure 4.8: Comparison of repeat-sales transit station gradient index between various types of observation

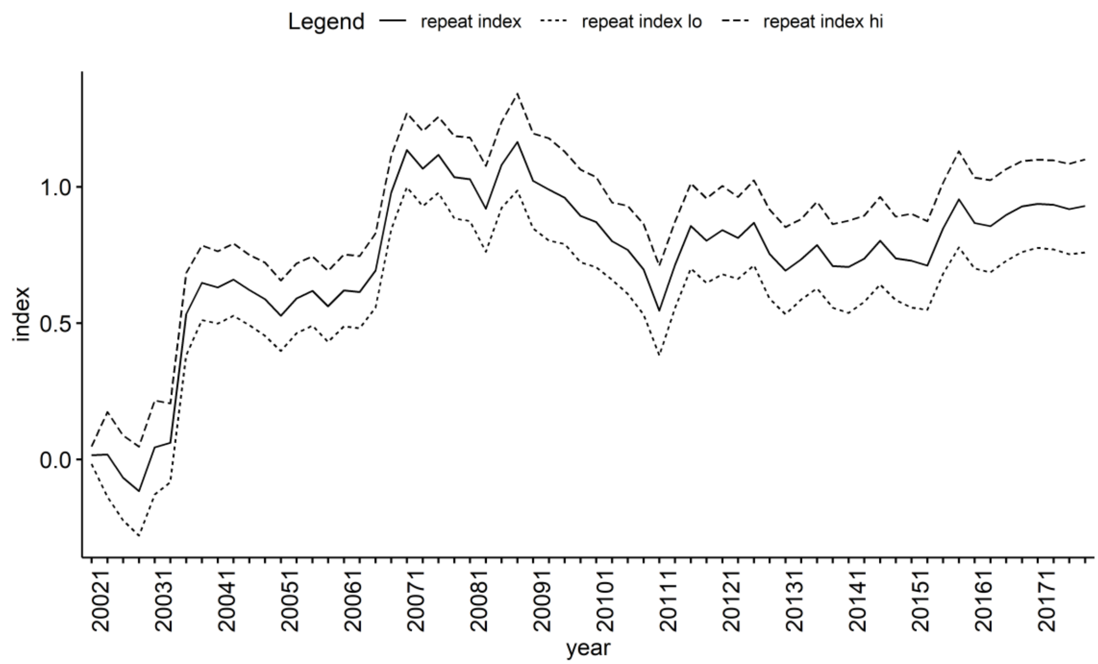

Figure 4.9: Housing price index of the Montgomery County

I utilized Equation (18) to check whether there is, on average, a significant home price increase within the 1.5-mile catchment area of the new LRT construction compared with homes outside the catchment area. The DID model includes control variables indicating property attributes and community characteristics. At the meanwhile, the model includes 15 time-dummy variables. The coefficients of the control variables all have expected signs. Table 4.4 shows the results. The model has a respectable adjusted R-square with the degree of 0.781. The coefficient of the interaction term is positively significant, which means that home prices in the 1.5-mile catchment area increase by more than 10.1% on average than home prices in the control area after 2012. This means that the Purple Line project had an impact on single-family home values after 2012, when the Purple Line progressed into the engineering phase.

Table 4.4: Price effects of proximity before and after the Purple Line progressed into the engineering phase

Variable	Coefficient	Std. Error
Intercept	8.074***	0.021
After	0.600***	0.002
Treated (within one and half mile)	0.139***	0.002
Interaction term of After and Treated	0.101***	0.003
ln lot size	0.084***	0.000
ln building area	0.455***	0.002
Age	0.001***	0.000

More than one story	-0.033***	0.001
Masonry	0.015	0.019
Basement	0.073***	0.001
Grade of structure	0.220***	0.001
Ln school quality	0.016***	0.003
Proportion Black	-0.853***	0.005
Proportion Vacant	0.538***	0.027
# of observations	208,709	
Adjusted R^2=	0.781	

The dependent variable is the natural logarithm of the sales price. The regression include 15 time-dummy variables. For simplicity, I do not present the coefficients of time-dummy variables in the above table.

*** p=.001(two-tailed).**p=0.01(two-tailed). *p=0.05(two-tailed).

Conclusion

I explored in this essay the effects of an investment in LRT on single-family home prices in the preservice period. My findings contribute to the growing body of evidence that investments in light rail begin to affect property values well before service begins. I also confirm the findings of others that housing prices started to rise as the Purple Line entered its preliminary engineering phase. While it is not reasonable to expect that this will occur for every light rail project, there is now growing evidence that this is a reasonable expectation. My findings also confirm that housing price impacts are not uniform throughout the light rail corridor, but vary systematically with distance from anticipated light rail stations – particularly for new stations that are not already part of an existing transit system. These results add strong support for economic theory that suggests that variation in housing prices reflects variation in accessibility. For these reasons, prices fall with distance from light rail stations and prices rise most significantly in locations where accessibility increases significantly. In Montgomery County, this will occur near new light rail stations and not near stations that already offer substantial accessibility via the existing Metro station. There is every reason to expect this to occur in other places under similar conditions.

For policymakers and advocates in Montgomery County, these findings have clear, important, and quite specific policy implications. First, the results suggest that concerns about the potential displacement of low-income residents though transit-induced gentrification are valid. So too, however, is the proposition that value capture strategies offer an opportunity to raise revenues for transit supportive investments in infrastructure. Second, the results suggest that if the county intends to use value capture strategies and implement policies designed to prevent displacement, there is no time to wait. The market has already anticipated the value that the transit-created accessibility will provide and has begun to respond accordingly. This means it is likely that value capture opportunities are already being lost and low-income residents are already being displaced. This of course depends on whether low-income residents are renters or homeowners. Finally, the results suggest that efforts to capture value and/or prevent displacement should be carefully targeted. There is little justification for implementing such policies in Bethesda or Silver Spring where Metro stations already exist. Housing prices in these locations are unlikely to rise significantly because of the Purple Line. Along new stations between Bethesda and Silver Spring and east of Silver Spring, however, accessibility and, hence, housing prices are likely to rise substantially. These are also the places where low-income, minority, and immigrant populations reside in higher concentrations. These are thus the locations where land value capture and anti-displacement strategies are most needed and likely to succeed.

Although the study produces the aforementioned findings, there are issues I need to note. First, the time when the effect of new LRT on single-family property value is captured varies across lines and metropolitan areas. For example, the effect was captured when the Purple Line project progressed into the engineering phase in my study, but the effect was captured when the plan of the LRT was first announced in Knaap et al. (2001). Hence, empirical study for each LRT project is needed for tax assessors to adjust the assessment values of properties located within the vicinity of the LRT project. Second, the effect of newly built stations and the effect of previously established stations on property value differ. It is common for a new LRT system to incorporate both newly built stations and previously established stations that serve existing Metro lines. Of note is that the difference between the effect of newly built stations and the effect of established stations is ignored in the current literature. Ignoring the difference of the effect across types of stations invalidates both benefit-cost analyses of transit investment and property tax assessments. Third, I noticed that it is still necessary to examine property values for the period after the LRT opens. There is no certainty that the kind of anticipatory premium found before the opening will continue. Future research might monitor the development of the value of properties. This evidence indicates how property values near transit stations evolve through the pre-engineering phase, engineering phase, and operating phase.

Given my results, accurate benefit-cost analysis of transit investment should account for when and where the impact of LRT occurs. In addition, knowing the timing of the impact and spatial variability of the impact of LRT will help identify the timing and areas of property tax assessment adjustment. Planners and policymakers may need to shape policies and programs that will help prevent displacement due to possible property tax increases.

5. Conclusion and Discussion

Conclusion

Before I conclude this dissertation, I would like to revisit the key findings. In the first essay, three main findings stand out. First, the essay shows that multilevel linear models are more appropriate than OLS models for multifamily housing rent hedonic analysis. More specifically, random coefficient models are more appropriate than the random intercept model and the OLS model in terms of model fitness and estimation accuracy in multifamily contexts. Second, the essay shows that the effects of service provision and management variables on multifamily rents vary across types of service and management. Pet allowance, availability of short-term leasing options, and storage service availability increase rents significantly, while renovations and availability of disability service do not increase rents. Third, the apartment (multifamily housing) is segregated by submarkets that are disaggregated by unit type. The effects of service provision and management on rents vary across submarkets.

The second essay shows that the rents of units within a half mile from the nearest planned Purple Line stations increase more than their counterparts in the county, on average. This is a noticeable result. To best of my knowledge, this is the first essay in the literature to empirically test whether urban rail transit in preservice period impacts multifamily rents. Furthermore, I find that two-bedroom units, three-bedroom units, and four-bedroom units see their rents increase significantly more than their counterparts in the control group. In contrast, zero-bedroom units and one-bedroom units do not increase statistically. I also estimate the rent gradient with respect to the distance to a station using the parameterized distance decay regressions. On average, the rent price gradient is -0.077 in 2018.

The third essay has two main findings. First, homes located within the vicinity of new LRT stations saw the premium beginning in 2012, when the planned rail transit line progressed into the engineering phase, 10 years before its service operation. Second, the premium varies across geographical spaces. In particular, the magnitude of the effect associated with the newly built stations is significant. In contrast, homes located near previously established stations capture this kind of effect, but the effect is not significant. The reason may be that the established stations already offer urban rail transit service and the marginal improvement of accessibility that the Purple Line would bring could be limited.

<u>Discussion</u>

There are some important topics that are related to the dissertation, but that require more research in the future.

1. It would be warrant to use a multilevel linear model to conduct multifamily rent hedonic analysis. Meanwhile, it would need to account for spatial autocorrelation. Multifamily housing rents are hierarchical and rents are spatially correlated. An apartment property manager or owner sets the costs of rents for his/her property based not only on the property attributes but also in comparison to rents of other nearby properties. It is reasonable and essential to take a hierarchical data structure and spatial autocorrelation into account simultaneously in the future when conducting multifamily housing rent hedonic analysis. The main challenge of this study is identifying which specification of the hierarchical spatial model should take

under the multifamily context.

2. Examining whether subsidized housing programs influence multifamily rents. Multifamily housing provides accommodation for low-and moderate income households. Federal government and local governments initiated many subsidized programs to make multifamily housing affordable for low-and moderate income households. However, there are not many empirical studies to test whether the subsidized housing programs decrease the multifamily rents significantly. Scholars are also eager to know which subsidized programs influence multifamily rents, particularly from a micro-level perspective.

3. Multifamily housing rent quantile regression analysis. Different tenants value different multifamily housing attributes at different prices. Low-income tenants are less likely to own their own car(s). As such, low-income tenants may value the proximity to a rail transit station and other public transportation more than high-income tenants. As such, it would be interesting and important to conduct multifamily housing rent quantile regression analysis. Unfortunately, discussions on quantile regression analysis of multifamily housing is lacking in existing literature, although quantile analysis is carried out extensively for single-family housing.

4. Will the multifamily housing rents that are proximity to the rail transit stations continue to increase after the Purple Line is in service? The effects of rail transit on multifamily housing rents may vary across time. It would be interesting to tract the effects across time, especially after the rail transit is in service. Furthermore, the geographical scope of the effects may expand from a half-mile catchment to a one-mile catchment. Such analyses require future empirical studies.

5. Does construction of rail transit displace residents and make neighborhoods more segregated? Rail transit project is expected to improve neighborhood accessibility to jobs, recreation, and reduce residents travel cost. The neighborhoods newly connected by rail transit become more attractive and cause housing prices to rise. This may displace current residents and create social impacts. Current literature fails to empirically examine these potential social effects resulting from the rail transit, but such analysis is extremely important.

6. Appendix I: Empirical Data Validation

This appendix tries to validate the survey data I use in Essay #1 and Essay #2. The survey data is from the Montgomery County rental housing survey. Montgomery County Code 29-51 requires all landlords of rental housing units to participate in the Annual Rental Facility Occupancy Survey conducted by the Department of Housing and Community Affairs (DHCA). Landlords who do not comply with the annual rental survey may be subject to a civil citation and /or fine. Since 2015,the survey is conducted from April 1-30 each year and tracks vacancies, turnover rate, rents, and amenities. Since 2014, the survey has been conducted online. Facilities located within Montgomery County's unincorporated areas as well as the municipalities of Rockville, Gaithersburg, and Takoma Park participate in the survey.

Table 6.1: Multifamily facilities in the survey

Description(LicenseStatusType)(group)	Accessory Apartment Class 1	Accessory Apartment Class 3	Condominium(Garden,High Rise)/Co-op Apartments	Multi-Family Survey Out of Jurisdiction	Multi-family Apartment Complex	Registered Living Unit	Single Family,Townhouse, Back to Back,Duplex	Unknow
Approved						529		
Exception	171	15	4,555	2	80	15	14,533	1,232
Pending	2	12	92		10	59	131	
Pending Renewal	1	2	18		5		46	
Potential	42	15	3,046		8	16	10,723	1,404
Registered					**2**			
Licensed	189	136	9,237		**628**		16,782	
Survey Only				**281**	**19**		1	

The Montgomery Rental Survey is sent to the Licensed/Registered multifamily Rental facilities (boxed in red below) across the county. To date there are 630 non-municipal Rental Facilities that are within the Montgomery County Jurisdiction, and another 300 that are licensed through their respective municipality (Gaithersburg, Rockville, and Takoma Park); these are all sent the survey. Out of these facilities, 846 answered the facility portion of the survey in 2018, and even fewer completed the unit portion. I further compared the number of multifamily rental units in the survey data to number of rental units in American Community Survey data for validation.

Table 6.2: Comparing census housing counts to my data

	Census Data		Rental Survey Count in 2017	Difference in 2017	Rental Survey Count in 2018	Difference in 2018
	Estimate	Margin of Error				
Total:	373,219	+/-2,431				
Owner occupied	243,517	+/-3,486				
Renter occupied:	129,702	+/-3,715	80,575	49,127	72,202	57,500
No bedroom	7,605	+/-1,464	3,684	3,921	3,681	3,924
1 bedroom	37,551	+/-3,293	32,743	4,808	30,750	6,801
2 bedroom	47,398	+/-3,369	37,698	9,700	32,088	15,310
3 bedroom	27,511	+/-2,779	6,186	21,325	5,420	22,091
4 bedroom	6,634	+/-1,239	264	6,370	254	6,380
5 or more bedrooms	3,000	+/-985	11	2,989	9	2,991
Note: Census data from 2017 1-year estimates; table B24042 in ACS;Housing tenure by Bedrooms. For survey data, I have dropped duplicated entries and outliers.						

The U.S. Census Bureau asks the tenure of occupants regardless of the facility type or ownership of the units (single landlord, accessory apartments, single family detached rental unit). As the multifamily facilities in the survey table shows, there are many facility types around the county for which it would be impossible to obtain survey information from all of the landlords in the county. The rental survey is only for licensed multifamily properties. The 2018 survey is missing approximately 100 facilities, and 80% of these properties reside in Takoma Park, which is technically outside the jurisdiction. Below is a table representing all "multifamily properties" by their license status as of 2018.

Table 6.3: Multifamily properties taking 2018 survey

Description(LicenseStatusType)	Did not Complete 2018 Survey	Completed 2018 Survey
Condemned	1	
Eliminated	1	
Exception	57	8
Inactive	83	
Licensed	12	600
New Application	42	
Pending	3	6
Pending Renewal	2	3
Potential	8	
Registered		2
Survey Only	92	186
Unlicensed	4	39

Table 6.4: Comparing census housing median rents to my data

	Census Data		Rental Survey Count in 2017	Difference in 2017	Rental Survey Count in 2018	Difference in 2018
	Estimate	Margin of Error				
Median gross rent Total:	**1,708**	**+/-25**	1480	228	1503.0	205
No bedroom	1,381	+/-94	1329	52	1410	-29
1 bedroom	1,473	+/-37	1345	128	1385	88
2 bedroom	1,727	+/-28	1525	202	1550	177
3 bedroom	1,980	+/-76	1798	182	1800	180

4 bedroom	2,351	+/-113	1624	727	1873.5	478
5 or more bedrooms	2,372	+/-288	1557	815	2550	-178
Note:Census data from 2017 1-year estiamtes;table B24042 in ACS;Housing tenure by Bedrooms. For survey data,I have dropped duplicated entries and outliers						

In summary, the data I used in the dissertation cover more than 90% of multifamily facilities in Montgomery County; I consider these data valid and representative. These high-quality data provide a solid foundation for this research.

7. Appendix II: Letter to Dean of the Graduate School

UNIVERSITY OF
MARYLAND

SCHOOL OF ARCHITECTURE, PLANNING, AND PRESERVATION
THE NATIONAL CENTER FOR SMART GROWTH RESEARCH AND EDUCATION
COLVIN INSTITUTE OF REAL ESTATE DEVELOPMENT

College Park, Maryland 20742-1411
301.405.2158 Tel 301.314.5639 FAX
www.arch.umd.edu
www.smartgrowth.umd.edu

September 5, 2019

Dr. Steve Fetter
Associate Provost for Academic Affairs and Dean of the Graduate School
2123 Lee Building
College Park, MD 20742

Dear Dr.Fetter,

I am writing to certify that the examining committee has determined that Qiong Peng made substantial contribution to a work that included in his dissertation as one essay in a three-essay format dissertation. As Qiong's dissertation advisor and a co-author on the piece, I can vouch for his contributions to the work. The piece, titled "Do multifamily unit rents increase in response to planned light rail construction" is the Essay #1. Qiong Peng was lead author on the piece coauthored by Nicholas Finio and myself. Qiong took part in data collection and analysis and wrote much of the article.

The inclusion of this work in this dissertation has been approved by myself as Qiong's dissertation advisor and by Casey Dawkins, the PhD program Director in Urban and Regional Planning and Design.

Sincerely,

Gerrit Knaap

Professor, Urban Studies and Planning Program
Director, National Center for Smart Growth
Associate Dean, School of Architecture, Planning and Preservation
University of Maryland, College Park
Email: gknaap@umd.edu

Bibliography

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). When should you adjust standard errors for clustering? (No. w24003). National Bureau of Economic Research.

Action Committee for Transit,2020. Montgomery County's Advocates for Better Transportation. Retrieved from http://www.actfortransit.org/purple_line.html

Agostini, C. A., & Palmucci, G. A. (2008). The anticipated capitalisation effect of a new metro line on housing prices. *Fiscal studies, 29*(2), 233-256.

Allen, M. T., Springer, T. M., & Waller, N. G. (1995). Implicit pricing across residential rental submarkets. *The Journal of Real Estate Finance and Economics, 11*(2), 137-151.

Anselin, L. (2013). *Spatial econometrics: methods and models* (Vol. 4): Springer Science & Business Media.

Anselin, L., & Lozano-Gracia, N. (2009). Spatial hedonic models. In *Palgrave handbook of econometrics* (pp. 1213-1250): Springer.

Arbia, G. (2006). *Spatial econometrics: statistical foundations and applications to regional convergence*: Springer Science & Business Media.

Asabere, P. K., & Huffman, F. E. (1996). Negative and positive impacts of golf course proximity on home prices. *The Appraisal Journal, 64*(4), 351.

Babawale, G., Koleoso, H., & Otegbulu, C. (2012). A hedonic model for apartment rentals in Ikeja area of Lagos metropolis. *Mediterranean Journal of Social Sciences, 3*(3), 109-120.

Baltagi, B. (2008). *Econometric analysis of panel data*: John Wiley & Sons.

Benjamin, J., Chinloy, P., & Hardin, W. (2007). Institutional-grade properties: Performance and ownership. *Journal of Real Estate Research, 29*(3), 219-240.

Benjamin, J., Sirmans, G., & Zietz, E. (1997). Security measures and the apartment market. *Journal of Real Estate Research, 14*(3), 347-358.

Bogdon, A., & Ling, D. (1998). The effects of property, owner, location, and tenant characteristics on multifamily profitability. *Journal of Housing Research, 9*(2), 285-316.

Boucq, E., & Papon, F. (2008). Assessment of the real estate benefits due to accessibility gains brought by a transport project: the impacts of a light rail infrastructure improvement in the Hauts-de-Seine department.

Brunauer, W., Lang, S., & Umlauf, N. (2013). Modelling house prices using multilevel structured additive regression. *Statistical Modelling, 13*(2), 29.

Cao, X., & Lou, S. (2017). When and How Much Did the Green Line LRT Increase Single-Family Housing Values in St. Paul, Minnesota? *Journal of Planning Education and Research*, 0739456X17707811.

Cervero, R., & Duncan, M. (2002). Benefits of proximity to rail on housing markets: Experiences in Santa Clara County. *Journal of Public Transportation, 5*(1), 3.

Chasco, C., & Gallo, J. L. (2013). The impact of objective and subjective measures of air quality and noise on house prices: a multilevel approach for downtown Madrid. *Economic Geography, 89*(2), 22.

Clapp, J. M., & Salavei, K. (2010). Hedonic pricing with redevelopment options: A new approach to estimating depreciation effects. *Journal of Urban Economics, 67*(3), 362-377.

Damm, D., Lerman, S. R., Lerner-Lam, E., & Young, J. (1980). Response of urban real estate values in anticipation of the Washington Metro. *Journal of Transport Economics and Policy*, 315-336.

Dawkins, C., & Buehler, R. (2010). Promoting affordable housing near public transit: The role of planning. *Policy Paper, 3*.

Devaux, N., Dubé, J., & Apparicio.P. (2017). Anticipation and post-construction impact of a metro extension on residential values: The case of Laval (Canada), 1995–2013. *Journal of Transport Geography, 62*, 12.

Djurdjevic, D., Eugster, C., & Haase, R. (2008). Estimation of hedonic models using a multilevel approach: An application for the Swiss rental market. *Swiss Journal of Economics and Statistics, 144*(4), 23.

Dubé, J., Legros, D., & Devaux, N. (2018). From bus to tramway: Is there an economic impact of substituting a rapid mass transit system? An empirical investigation accounting for anticipation effect. *Transportation Research Part A: Policy and Practice*, *110,* 15.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International regional science review, 26*(3), 244-268.

Elhorst, J. P. (2014a). Dynamic spatial panels: models, methods and inferences. In *Spatial Econometrics* (pp. 95-119): Springer.

Elhorst, J. P. (2014b). Spatial panel data models. In *Spatial econometrics* (pp. 37-93): Springer.

Elhorst, J. P. (2014c). Spatial panel models. In *Handbook of regional science* (pp. 1637-1652): Springer.

Ferguson, B., Goldberg, M., & Mark, J. (1984). *The pre-service impacts of the Vancouver advanced light rail transit system on single family property values*.

Frew, J., & Jud, G. (2003). Estimating the value of apartment buildings. *Journal of Real Estate Research, 25*(1), 77-86.

Gatzlaff, D. H., & Smith, M. T. (1993). The impact of the Miami Metrorail on the value of residences near station locations. *Land Economics*, 54-66.

Gelfand, A. E., Banerjee, S., Sirmans, C. F., Tu, Y., & Ong, S. E. (2007). Multilevel modeling using spatial processes: Application to the Singapore housing market. *Computational Statistics & Data Analysis, 51*(7), 13.

Gipe, G. W. (1976). Developing a Multiple Regression Model for Multi-Family Residential Properties. *Real Estate Appraiser, 42*(3), 28-33.

Giuliano, G., Gordon, P., Pan, Q., & Park, J. (2010). Accessibility and residential land values: some tests with new measures. *Urban studies*, *47*(14), 28.

Glaesener, M. L., & Caruso, G. (2015). Neighborhood green and services diversity effects on land prices: Evidence from a multilevel hedonic analysis in Luxembourg. *Landscape and Urban planning*(143), 12.

Golub, A., Guhathakurta, S., & Sollapuram, B. (2012). Spatial and temporal capitalization effects of light rail in Phoenix: From conception, planning, and construction to operation. *Journal of Planning Education and Research, 32*(4), 15.

Goodman, A. C., & Thibodeau, T. G. (1998). Housing market segmentation. *Journal of Housing economics, 7*(2), 23.
Guntermann, K. L., & Norrbin, S. (1987). Explaining the variability of apartment rents. *Real Estate Economics, 15*(4), 321-340.
Hardin III, W., & Cheng, P. (2003). Security, gated communities, and apartment rents. *Journal of Real Estate Research, 25*(2), 145-158.
Hardin III, W., Hill, M., & Hopper, J. (2009). Ownership structure, property performance, multifamily properties, and REITs. *Journal of Real Estate Research, 31*(3), 285-306.
Henneberry, J. (1998). Transport investment and house prices. *Journal of Property Valuation and Investment, 16*(2), 144-158.
Hill, R. J. (2013). Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of economic surveys, 27*(5), 879-914.
Hoch, I., & Waddell, P. (1993). Apartment rents: another challenge to the monocentric model. *Geographical analysis, 25*(1), 20-34.
Hoesli, M., Thion, B., & Watkins, C. (1997). A hedonic investigation of the rental value of apartments in central Bordeaux. *Journal of Property Research, 14*(1), 15-26.
Jaffe, A. J., & Bussa, R. (1977). Using a simple model to estimate market rents: A case study. *Appraisal Journal, 45*, 7-13.
Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in ecology and evolution, 5*(9), 3.
Jones, K., & Bullen, N. (1993). A multi-level analysis of the variations in domestic property prices: Southern England, 1980-87. *Urban studies, 30*(8), 18.
Jud, G. D., Benjamin, J. D., & Sirmans, G. S. (1996). What do we know about apartments and their markets? *The Journal of Real Estate Research, 11*(3), 243-257.
Jud, G. D., & Winkler, D. T. (1991). Location and amenities in determining apartment rents: an integer programming approach. *Appraisal Journal, 59*(2), 266-275.
Kain, J. F., & Quigley, J. M. (1970). Evaluating the quality of the residential environment. *Environment and Planning A, 2*(1), 23-32.
Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics, 17*(1), 99-121.
Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review, 40*(2), 509-533.
Kelejian, H. H., & Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics, 118*(1-2), 27-50.
Knaap, G. J., Ding, C., & Hopkins, L. D. (2001). Do plans matter? The effects of light rail plans on land values in station areas. *Journal of Planning Education and Research, 21*(1), 32-39.
Kreft, I. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*: Sage.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate behavioral research, 36*(2), 29.
Lancaster, K. (1966). Change and innovation in the technology of consumption. *The American Economic Review, 56*(1/2), 14-23.
Lee, L. F. (2004). Asymptotic distributions of quasi‐maximum likelihood estimators for spatial autoregressive models. *Econometrica, 72*(6), 1899-1925.
LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*: Chapman and Hall/CRC.
Lin, J.-J., & Cheng, Y.-C. (2016). Access to jobs and apartment rents. *Journal of Transport Geography, 55*, 121-128.
Loomis, J., Santiago, L., & Lopez de Jesus, Y. (2012). Effects of construction and operation phases on residential property prices of the Caribbean's first modern rail transit system. *Urban Public Economics Review, 17*.
Marks, D. (1984). The effect of rent control on the price of rental housing: an hedonic approach. *Land Economics, 60*(1), 81-94.
McDonald, J. F., & Osuji, C. I. (1995). The effect of anticipated transportation improvement on residential land values. *Regional science and urban economics, 25*(3), 261-278.
McMillan, A., & Lee, S. (2017). Smart growth characteristics and the spatial pattern of multifamily housing in US metropolitan areas. *Urban studies, 54*(15), 3500-3523.
McMillen, D. P., & McDonald, J. (2004). Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983–1999. *Real Estate Economics, 32*(3), 463-486.
Mejia, L. C., & Potter, K. D. (2015). The Value Behind the Value-Add: Multifamily Rent Growth After Renovations. *Journal of Housing Research, 24*(2), 211-220.
Miller, N. G. (1982). Residential property hedonic pricing models: a review. *Research in Real Estate, 2*(1), 31-56.
Mulley, C. and Tsai, C.H.P., 2016. When and how much does new transport infrastructure add to property values? Evidence from the bus rapid transit system in Sydney, Australia. Transport Policy, 51, pp.15-23.
Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed‐effects models. *Methods in ecology and evolution, 4*(2), 10.
Ogur, J. D. (1973). Higher education and housing. *American Journal of Economics and Sociology, 32*(4), 387-394.
Orford, S. (2002). Valuing locational externalities: a GIS and multilevel modelling approach. *Environment and Planning B: Planning and Design, 29*(1), 13.
Portnov, B., Genkin, B., & Barzilay, B. (2009). Investigating the effect of train proximity on apartment prices: Haifa, Israel as a case study. *Journal of Real Estate Research, 31*(4), 371-395.
Quigley, J. M. (1985). Consumer choice of dwelling, neighborhood and public services. *Regional science and urban economics, 15*(1), 23.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1): Sage.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy, 82*(1), 34-55.
Shear, W. B. (1983). A note on occupancy turnover in rental housing units. *Real Estate Economics, 11*(4), 525-538.
Shenkel, W. M. (1975). The Valuation of Multiple Family Dwellings by Statistical Inference. *Real Estate Appraiser, 41*(1), 25-36.
Shin, W. J., Saginor, J., & Van Zandt, S. (2011). Evaluating subdivision characteristics on single-family housing value using hierarchical linear modeling. *Journal of Real Estate Research, 33*(3), 32.
Sirmans, S., & Sirmans, C. (1992). Property manager designations and apartment rent. *Journal of Real Estate Research, 7*(1), 91-98.
Sirmans, S., Sirmans, C., & Benjamin, J. (1989). Determining apartment rent: the value of amenities, services and external factors. *Journal of Real Estate Research, 4*(2), 33-43.
Smith, C. A., & Knoll, M. (1988). Improving Estimates of Potential Gross Income in Multifamily Properties through Market Research. *Appraisal Journal, 56*(1).
Snijders, T. A., & Bosker, R. J. (1999). *An introduction to basic and advanced multilevel modeling*: Sage.
Song, Y., & Knaap, G.-J. (2003). New urbanism and housing values: a disaggregate assessment. *Journal of Urban Economics, 54*(2), 218-238.
Song, Y., & Knaap, G.-J. (2004). Measuring the effects of mixed land uses on housing values. *Regional science and urban economics, 34*(6), 663-680.
Team, R. D. C. (2013). nlme: linear and nonlinear mixed effects models (Version R package version).
Treg, C. (2010). A multilevel property hedonic approach to valuing parks and open space.
Uyar, B., & Brown, K. (2007). Neighborhood affluence, school-achievement scores, and housing prices: cross-classified hierarchies and HLM. *Journal of Housing Research, 16*(2), 20.
Wilson, B., & Frew, J. (2007). Apartment rents and locations in Portland, Oregon: 1992-2002. *Journal of Real Estate Research, 29*(2), 201-218.
Wolverton, M. L., Hardin, W. G., & Cheng, P. (1999). Disaggregation of local apartment markets by unit type. *The Journal of Real Estate Finance and Economics, 19*(3), 15.
Yan, S., Delmelle, E., & Duncan, M. (2012). The impact of a new light rail system on single-family property values in Charlotte, North Carolina. *Journal of Transport and Land Use, 5*(2), 60-67.
Zietz, E. (2003). Multifamily housing: A review of theory and evidence. *Journal of Real Estate Research, 25*(2), 185-244.