THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF MATHEMATICS

IMPLEMENTING A COMMUTING MARKET ACCESS MODEL TO THE WASHINGTON, DC METROPOLITAN AREA

ATHARV GUPTE
SPRING 2020

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Mathematics
with honors in Mathematics

Reviewed and approved* by the following:

Ruilin Zhou
Professor of Economics
Thesis Supervisor

Mark Levi
Distinguished Professor of Mathematics
Faculty Reader

* Electronic approvals are on file.

ABSTRACT

Scholars, policymakers, and real estate developers all agree that constructing new public transportation systems results in increased economic growth and labor productivity. Several empirical models have been constructed that reliably conclude specific economic and social effects that changes in commuting patterns provide. However, the vast majority of these models have only been applied to cities in developing economies, where the construction of transit systems has recently been much greater than in developed economies such as the United States. This has led to considerable uncertainty to the degree of relocation resulting from the construction of new rail or bus lines. Such uncertainty has stalled several transit and urban development proposals.

This paper discusses a commuter market access (CMA) model developed for Bogotá, Colombia. The Bogotá model reliably explains various economic effects of its Transmilenio transit system, which drastically changed commuting patterns across the city. This study then applies the Bogotá model to neighborhoods in the Washington, DC area, a much higher income region. The reliability of this model on various geoeconomic developments in the DC region is then analyzed, and compared to that of Bogotá. In the end, the Bogotá model's merits can be extended to analyze firm relocation preferences in developed economies, but household relocation in such economies is much more nuanced and remains not fully explained.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis supervisor and economics professor, Dr. Ruilin Zhou, for being an exceptional role model for me, both in helping me write this paper and in serving as a mentor for me to shape my career choice. Prof. Zhou has an enormous amount of knowledge across various economic disciplines, and it was a pleasure to get to know her through her class and my thesis at Penn State.

Next, I would like to give a shout out to the Penn State Real Estate Society, which I semi-reluctantly joined the fall of freshman year as a non-business major, only to become hugely immersed with the field and a huge array of fellow Penn Staters. I’ll never forget the 5-hour van rides to New York, or the late night exec meetings in the deep of winter, that would shape my elective choices, thesis topic, and perhaps a future side hustle.

Finally, I wouldn’t have been able to get to this position without my parents, Ranjit and Shilpa Gupte. They have both worked incredible hours to nurture me and afford me the best opportunities to thrive.

Chapter 1

Introduction

Access to transportation has been a key driver to economic growth for hundreds of years. In the past century, use of public transportation, such as trains or buses, has taken off, aiding to decongest rapidly growing cities around the world. Most cities in developed economies such as in the United States, Britain, or Germany have established transit systems that have not seen major overhauls or expansion in the past 50 years, primarily stemming from their stagnation in population growth. However, in developing countries, the opposite is the case, where until the 1990s public transportation primarily consisted of just buses. In just the past 20 years, countries such as India, South Africa, and Colombia have all implemented rapid transit systems that changed the dynamics of their cities.

One specific case example is Bogotá’s Transmilenio Bus Rapid Transit (BRT) system, the most-used such type of system in the world. A BRT utilizes dedicated busways complete with stations and ticketing kiosks, and is much cheaper to implement than a heavy rail line such as a metro or monorail. One paper by Nick Tsivanidis of the University of Chicago frames a simplified equilibrium matrix equation, whose solution vector is a simplified version of a term he names the commuter market access (CMA). Each census tract in Greater Bogotá is assigned a CMA for both households and firms.

However, such empirical studies have rarely been tested on applications in mature countries, such as the United States. Most cities in the US have not seen significant population growth necessitating the construction of new transit lines in the past 50 years. In general, transit lines have been fueled by rapid growth, not minuscule population changes, the latter of which several US metros are experiencing. These increases have slowly begun to strain existing infrastructure around cities. As such, there is a recently renewed push to invest more in public transportation in the US.

One challenge facing urban planners and policyholders regarding transit proposals is a lack of clarity on the types and degrees of resulting growth, such as in property values, productivity, or availability of amenities. This is primarily due to the lack of transferrable studies conducted in high-income economies, as opposed to emerging markets. This paper aims to test the merits of Tsivanidis's CMA model to the Washington, DC metropolitan area. Washington, DC is unique in that its surrounding region is seeing huge growth in recent years, both in population and transit system scope. Its Metro rail system is expanding rapidly to this day, with an extension to Dulles International Airport expected to open in 2020.[1]

The rest of the paper is structured as follows. Chapter 2 discusses various studies on determining outcomes resulting from transit construction. Arguments on the merits of directly linking transportation with property values are addressed, as well as nuances with working with residential and commercial real estate data. Finally, unique challenges of transit development in more developed countries will be analyzed in specific case examples.

Chapter 3 takes a look at Tsivanidis's CMA model, which is constructed to analyze the outcomes of Bogotá's Transmilenio. It starts with a brief explanation for the full, generalized CMA model, and then narrows it down to the simplified system of equations model that is the primary basis of this paper.

Chapter 4 gives an overview of the Washington, DC Metropolitan area along with its WMATA Metro rail system. It then discusses data collection techniques, which are done at the census tract level through a combination of the US Census Transportation Planning Package and CartoVista. The CMAs of each locality within a 15-mile radius of central DC are computed, and an analysis of their values follows. Finally, a case study is conducted analyzing changes in the CMA values along the DC Metro's newly constructed Silver Line.

Chapters 5 analyzes the added effect of the residential CMA, above using student-faculty ratios, in explaining housing prices. The residential CMA ends up being a much weaker element explaining housing valuations when school quality is also included. Chapter 6 analyzes the added effect of the firm CMA, above using employment growth, in explaining per-worker GDP. The firm CMA ends up being a significantly more effective proxy in explaining per-worker output. Both sections also include respective analyses with findings in Tsivanidis's paper.

Chapter 7 summarizes these findings, concluding that Tsivanidis's model can only be extended to developed economies when analyzing firm movement and relocation preferences, but not household tastes. It also includes potential future avenues of study, including using microdata to more effectively tease out an agent's commuting effects from other factors both on the business development and amenity sides.

Chapter 2

Previous Studies on Transportation and Regional Economics

Relating Growth with Transit Development

It is widely agreed upon that access to public transportation increases the ability for nearby neighborhoods to attract long-term residents, workers, and amenities. However, the exact degree and shape of the result effects have continually been hard to predict, especially in more advanced societies.

Regardless, many scholars do agree that access to transportation (and demand to expand transit networks) is not a factor of production, but rather a result of high observed economic growth rates. According to one study, 96% of new rail transit projects in the United States were implemented in regions seeing over a 5% observed population growth in the U.S. Census (per decade).[2] Similar findings can also be shown in developing countries. Overall, however, transportation networks do not predate urbanization; that is, it is very unlikely for a new train line or busway to extend into a low-density area which may see potential growth as a result of the line's completion. Instead, transportation access and reliability synergistically work with growth (either in population, workforce, or leisure activity), after initial growth (a catalyst) has been observed to validate transit construction.

Public Transportation in Developed versus Developing Economies

A major argument that supports the recent discrepancy in transit system expansion between developed and developing economies is that in the former, agents are much more diverse in terms of their preferences, and relocation cannot be predicted as reliably as in lower income economies. In emerging economies, there is a much greater general sentiment to "move up," trying to find the highest quality real estate.[3] In developed economies, relocation is much more nuanced, and relies much more heavily on connections and ties to neighborhoods.

Many scholars claim this to be the primary reason that real estate developers and transit planners are less likely to simultaneously develop huge apartment and commercial complexes in the US like they can in emerging economies.[4] Due to significantly stickier preferences among agents, large swings in demand are much less frequent in higher income regions like the US.

This could be a key reason that transit proposals are much more likely to be stalled in the United States. Los Angeles has the only heavy-rail subway system in North America that commenced operations in 1990 or later, and Washington, DC is the only city whose Metro has continually been expanding since its debut in 1976.[4]

Nevertheless, it is important to note that US development, while less predictable and strong than in emerging economies, has still occurred at a slow and steady rate, which has in the last couple decades led to strains on existing transportation systems. While the vast majority of Metropolitan Statistical Areas (MSAs) have only seen an annualized population growth rate of under 1% since 1970, this translates to about 65% when compounded over 50 years (from 1970 to 2020).[5]

In principle, complex agent relocation preferences and recent overutilization of existing transportation networks has caused gridlock in proposals, stalling several projects needed to increase capacity. This is not the case in emerging economies, where agents are more likely to relocate in denser clusters next to highly accessible transit nodes.

Transportation and Property Values

While it is commonly agreed upon that access to reliable public transportation boosts property values for both commercial and residential real estate, several sources point to the fact that directly relating property values with transit accessibility may cause statistically insignificant results.[5] One such agency, the National Center for Highway Research Planning, even deduced through multiple studies that the transferability of times series models to geographically similar areas is limited.[6] Much of this is caused by significant interaction effects between accessibility and other welcome amenities, like building quality and nearby leisure. Effectively, changes in property value resulting from new transit construction may just be the derivative of the aggregate of the immediate impacts such construction provided, without any significant valuation boost of its own.

Nevertheless, some clear relationships (both spatially and in terms of intensity) exist. One of the most important observations is that the degree of residential property value growth is significantly greater in a city's far-flung suburbs than areas closer in.[7] Such a relationship was also present for commercial property, though not as intense. Additionally, the degree of value change with respect to distance from a transit node (such as a station or highway junction) has also been studied to not follow a consistent gradient over geographies. The willingness of individuals to travel to access such nodes does not primarily depend solely on the distance or time to access them.[8]

It is also important to note that increased transportation access may actually be a bane to desirability, both residential and commercial. While such a phenomenon is not as prominent in Europe, in parts of North America, there is a lingering stigma that rapid transit access may lead to poverty, or diminish an area's "prestige factor," especially for more exclusive communities. This is cited as a primary reason why the DC Metro does not serve Georgetown, a wealthy, primarily residential neighborhood of the city.[9]

Residential versus Commercial Analysis

There are certain benefits and complications when analyzing residential and commercial desirability. In general, however, the decisions made by commercial agents (i.e. firms) can be very different from those made by households. For example, while commute accessibility can be a key factor in relocation preferences for both households and firms, certain amenities, like strong public schools, by themselves may not be valued as much by firms, especially as they may need to pay higher mill taxes.

In general, however, business relocation is much more predictable and less psychological than household relocation. Businesses are likely to relocate in places which give them economic benefits in their goal in maximizing profit. This would imply being accessible to the highest quality talent, as well as finding commercial space that minimizes long-term operating costs, both in terms of rent and utilities.[10] Based on this, and on the fact that many firms hire hundreds of workers, firms generally make decisions that synergistically work with the typical worker, despite some of these workers individually having slightly different preferences.[11] Essentially, a firm's relocation, especially for one that hires hundreds of employees, is in general much more predictable, even in developed economies. However, one challenge in analyzing firm outcomes is that tax data and past quoted rent data may not be representative of what these firms actually pay.[12]

On the other hand, there are several differing psychological reasons that can explain the relocation of residential agents. For example, individuals may be tied to an area with lifelong friends, a particular address, or school district.[13] Most importantly, however, each individual residential agent essentially acts according to the decision of a household, rather than the size of the firm, drastically increasing the variation of movement preferences and exposing several more nuances.[14] These two factors, especially in established, developed economies, have contributed to a lack of certainty among developers regarding certain demographics that could be housed in new developments.

Chapter 3

The Bogotá Commuter Access Model (CMA)

Overview on the Transmilenio System

The Transmilenio is a Bus Rapid Transit (BRT) system in Bogotá, Colombia. Opened in 2000, the BRT system was constructed to alleviate existing road traffic across the city, and seed new opportunities for development near highly-travelled corridors. The Transmilenio carries 2.2 million passengers each weekday, and has become the most-used such type of system in the world.[15]

BRT is a relatively recent concept in which standard buses are given dedicated, grade-separated lanes, known as “busways,” with platformed stations and centralized ticketing. In highly packed cities like Bogotá, BRT has shown to provide commute time decreases nearly as great as a heavy rail metro line, while costing between a third and a fifth of a rail-based system. Moreover, BRT systems can feed into existing communities and bus systems extremely well, as buses can serve existing streets, and then enter the grade-separated BRT system through “portals.” BRT systems are generally above ground or on elevated viaducts (as in Bogotá), but are occasionally built in tunnels.[16]

CMA Model Overview

To analyze land value and use implications of the Transmilenio, Nick Tsivanidis of the University of Chicago constructed an economic framework for the city of Bogotá. In this model, workers were segmented based on skill-level (high and low), and would be attracted to reside in neighborhoods with strong job access and amenities, and lower home prices. All residents were assumed to have public transport access (either through a standard bus or the Transmilenio BRT), but residents also had the option to own a car for commuting. At the same time, workplaces were segmented based on sector, car usage, and would be attracted to places with low floorspace costs and increased access to talent pools. In equilibrium, the price of floorspace and wages would clear all real estate and labor markets. Tsivanidis split up the city into 2,800 census tracts, each of which differed by floorspace availability, amenities, residential and commercial populations, and relative accessibility.

In the full, generalized Bogotá model, Tsivanidis uses properties of the Frechet distribution to define the likelihood of a worker living in tract i and working in tract j, and of type g (either skilled or unskilled), and commute type a (either automobile or public transportation), as:

$$\pi_{ijag} = \frac{(w_{jg}/d_{jsg})^{\theta}}{\sum_s (w_{sg}/d_{isg})^{\theta}} \equiv \frac{(w_{jg}/d_{jsg})^{\theta}}{\Phi_{Riag}}.$$

In this equation, θ is a parameter representing commute disutility, w_{sg} is the wage of an employee working in tract s in a type g role. This equation is where Tsivanidis coins his resident commuter market access (RCMA) term Φ_{Riag}, which is defined as $\sum_s (w_{sg}/d_{isg})^{\theta}$. The RCMA aims to measure a tract's relative competitiveness in supplying potential workers to firms, especially those with higher wages and a short distance away. Moreover, based on the Frechet distribution, Tsivanidis defines the firm commuter market access (FCMA), Φ_{Fjg}, in an intermediate step in calculating firm labor supply:

$$\Phi_{Fjg} = \sum_{i,a} d_{ija}{}^{-\theta} \frac{L_{Riag}}{\Phi_{Riag}};$$

$$L_{Fjg} = w_{jg}^{\theta} \Phi_{Fjg}.$$

So, in principle, the labor supply in a particular tract j depends on both the average wages as well as the relative competitiveness in attracting workers.

It turns out that when all labor markets, residential markets, and floorspace markets clear, Tsivanidis's methodology simplifies to a system of linear equations. Within the generalized model, Tsivanidis defines residential commuter market access (RCMA) and firm commuter market access (FCMA) which aimed to measure relative accessibility of a census tract to workplaces (RCMA) or talent pools (FCMA). It turns out that when the model was simplified to disregard segmentation of workers and residents, and fix floorspace allocations per capita, the RCMA and FCMA could be solved through a system of nonlinear equations:

$$\Phi_{Ri} = \sum_{j} e^{-\theta d_{ij}} \frac{L_{Fj}}{\Phi_{Fj}}$$

$$\Phi_{Fj} = \sum_{i} e^{-\theta d_{ij}} \frac{L_{Ri}}{\Phi_{Ri}}$$

$i, j \in \{1,2, \ldots . n\}$; n = total number of analyzed tracts

There are therefore $2n$ such equations, with Φ_{Ri} and Φ_{Fj} respectively representing the RCMA for tract i and FCMA for tract j. L_{Ri} and L_{Fj} respectively represent the total working population residing in tract i, and number of individuals working in tract j. In principle, these factors represent aggregate labor suppliability and hirabiliy. Finally, d_{ij} represents the average commute time from tract i (residential) to j (workplace), keeping in mind that $i = j$ does not imply $d_{ij} = 0$. θ is a parameter that represents travel disutility; a lower θ value indicates that individuals are more willing to commute further to obtain better employment, at the cost of personal and leisure time. Generally, θ is estimated to be higher for developed economies. In Tsivanidis's paper, θ was assumed to be .05, as an estimation parameter relative to property desirability.

Each RCMA value depends on all n FCMAs, and vice versa. Note that commute time increases disutility, causing both Φ_{Ri} and Φ_{Fj} to decrease. Also, it is clear to see that an increase in the total number employed in a tract (L_{Fj}) has a positive effect on Φ_{Ri}, as a higher quantity of jobs would be accessible. The same argument can be reversed in the case of the FCMA, Φ_{Fj}. Finally, each tract's RCMA is inversely related with all tracts' FCMA values. This makes sense since each tract i's relative residential appeal declines when a particular tract j can easily draw in a lot of workers.

Tsivanidis calculates RCMA and FCMA values for each of the 2,800 census tracts in Bogotá for both 2000 and 2015, and regresses these, as well as other explanatory variables, with floor space prices.

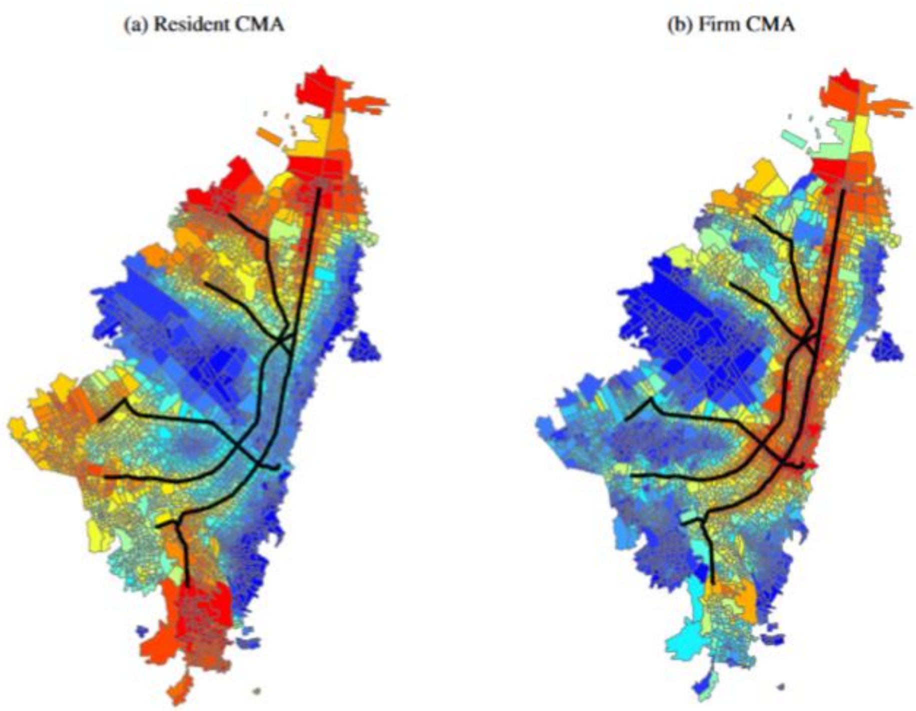

Figure 1. Map of Bogotá with growth in CMA identified by color. Areas in red represent the strongest positive CMA growth.

Based on the simplified model constructed by Tsivanidis, the places with the highest RCMA growth are generally in the outskirts of the city, while the places with the highest FCMA growth are in the city center and along the Transmilenio system (in black).

Chapter 4

Defining the DC Commuter Market

Demography and Transport System of Greater DC

The Washington, DC Metropolitan Statistical Area (MSA) houses over 6 million individuals, making it the sixth most populous metropolitan area in the United States, and the fastest growing major metropolitan region in the Northeast. Consisting of the District of Columbia, along with counties in Maryland and Virginia, the DMV as the region is known, houses a diverse array of individuals and firms, from the National Mall to far-flung exurbs approximately 40 miles away. There is also considerable commuting and social interaction with the nearby Baltimore area to the northeast.

The DMV is served by the Washington, DC Metro, the second-most used rail transit system in North America, after the New York City Subway. The Metro consists of six lines that join in central hubs in Downtown DC, before branching out into the Maryland and Virginia suburbs. The first stations of the Metro opened in 1976, and significant expansion is still ongoing as of 2020. For example, in 2014, the Silver Line opened, connecting Tysons and Reston (two major commuter and suburban office towns in Virginia) to central DC. This line is being expanded further west to serve Ashburn and Dulles.[17] Meanwhile, the New York Subway's Hudson Yards station, which opened in 2015, was part of the first extension to the system in 26 years.[18] Chicago's L hasn't been extended at all since the late 1970s, around the time the DC Metro started operations.[19]

Additionally, the suburbs of the DMV are served by a network of expressways, including the Interstate 495 Beltway which circles around the city. In recent years, several office parks have sprung up around the Beltway, most notably in Bethesda, Silver Spring, and Tysons. The Beltway was primarily a six-lane highway until the late 2000s, in which a widening took place, primarily in Virginia, to increase capacity and decrease traffic.[20]

Pedestrian traffic in the DMV is primarily confined to DC itself and select suburbs, such as College Park, home of the University of Maryland. Most of the DMV has been planned around the automobile, and several of the newer business parks have been designed as such, allowing ample room for parking and wide avenues. Many of the newer Metro stations in the suburbs have been constructed with a park-and-ride concept, with large car parks situated right next to the station.[21] Walking into or out of stations is not regularly done outside DC. All these factors lead the DC Metro to have a typical rider having a higher income than in other transit systems in the US, most notably Chicago’s L and New York’s Subway.

Data Collection Methods

Required data to calculate the RCMA and FCMA for the DC area includes residential working population, total number of workers, and average commute times between tracts, for both 2000 and 2015. Tracts are chosen at the municipality level, as boundaries of US Census Tracts changed substantially over this time period.

All municipalities that are at least partially within a 15-mile radius of the Washington Monument in Central DC would be included for this study, and would be selected through CartoVista, a mapping software. The number of workers living and working in each tract are found from the US Census database, and the Census Transportation Planning Package provides estimates for median travel time between different tracts. There are 95 such municipalities that kept common boundaries between 2000 and 2015, and they are all included for this study. The District of

Columbia is its own municipality, along with 94 others in Maryland and Virginia. A full list of municipalities (along with RCMA/FCMA values) is available in Appendix B.

Figure 2. Map of selected municipalities for 2015 study in CartoVista.

Calculating DC'S Commuter Market Access (CMA)

Since 95 tracts are included in the study, a system of 190 nonlinear equations had to be solved. Nonlinear equations cannot be solved through matrix inversion, but there are several other algorithms that can be used which iterate over a process that converges to the actual solution vector. The algorithm used to calculate the RCMA and FCMA values is the Newton-Raphson iteration, which is able to give a converging sequence to the roots of the equations. The process, along with manipulation to fit the CMA equations, is explained in Appendix C.

FCMA Calculations and Analysis

The FCMA values are logarithmically weighted around 1, i.e. $E\left[L_{Fj}\ln\left(\Phi_{Fj}\right)\right]=0$. Washington City has a 2015 FCMA value of 2.369, which reflects not only a high number of jobs, but also a relative (and proportional) ease of access from the suburban municipalities tabulated in the study. However, Bethesda, MD, has a 2015 FCMA value of 2.714. Bethesda is a major office center just northwest of DC's city limits, and located on the DC Metro's Red Line, along with easy access to the Interstate 495 Beltway and Interstate 270. Such an FCMA value could indicate that while Bethesda by absolute numbers is not a bigger job center than DC and its proximity to a large working population, much of which is relatively inaccessible to many other job centers (thus having lower RCMAs). One intriguing result is that Andrews Air Force Base has the highest FCMA value of 7.075. The AFB serves as a major individual employer, with very limited transit and employment opportunities nearby. It is located in a semi-rural area just barely within the 15-mile radius of Central DC. The FCMA equation more heavily weighs in residential populations that are closer in, and themselves generally less accessible to employment activities. It would be reasonable to conclude that the vast majority of employees of the AFB are using private transportation from nearby towns southeast of DC, which are otherwise very disconnected from the main transportation arteries. Such a find may present a limitation to using the FCMA values as a proxy for business desirability.

On the other hand, communities that not only have relatively fewer jobs, but also have easy access on transportation thoroughfares to job centers, exhibit the lowest FCMA values. For instance, Largo, MD, has among the lowest 2015 FCMA values, at .526. Largo is interesting in that it was relatively recently (in 2004) connected to the Washington Metro through an extension of the Blue Line (which heads into central DC and other suburban job centers). Such a low FCMA value may indicate that while Largo is well connected to transit opportunities, its lack of an existing commercial stronghold has made it relatively uncompetitive with nearby options in attracting jobs, as those other job centers could now more easily be access from Largo. In a sense, FCMA may not necessarily be related with transportation access, if such access decreases an area's relative competitiveness when attracting new jobs.

Places with Highest FCMA Rise		Places with Greatest FCMA Fall	
Marlton CDP, Maryland	397.17%	Brookmont CDP, Maryland	-87.28%
Cloverly CDP, Maryland	283.53%	North Springfield CDP, Virginia	-74.36%
White Oak CDP, Maryland	232.53%	Chevy Chase town, Maryland	-55.95%
Walker Mill CDP, Maryland	170.38%	Rosaryville CDP, Maryland	-53.35%
District Heights city, Maryland	141.96%	Springdale CDP, Maryland	-46.34%

Table 1. Top 5 greatest FCMA risers and fallers, 2000-2015

Interestingly, some of the highest FCMA risers occurred in places well outside the Interstate 495 Beltway. This could be due to residential growth in nearby outlying regions, which in general are not accessible to the biggest job centers (i.e. have a lower RCMA).

RCMA Calculations and Analysis

The residential commuter market access (RCMA) values are more nuanced in terms of potential causes and relationships with urban composure and access to transportation. The highest RCMA in 2015 (8.257) is that of Coral Hills, MD, immediately outside the southeastern border of DC, and within the Interstate 495 Beltway. Coral Hills is on the DC Metro Green Line, but far away from any highways. It also has very limited employment, with the vast majority of working residents commuting out to either DC or other suburbs. The location of Coral Hills is unique in that it is adjacent to places that are hard to access from most workers (i.e., had a low FCMA), but still has enough access to major job centers, via the Green Line. Based on the mechanics of the CMA equations, it would therefore seem that places with a good balance between nearby employment (that is otherwise hard to reach) and access to major job centers would have the highest RCMA.

Another interesting finding is that suburban business centers, such as Tysons and Bethesda, produce among the smallest RCMAs (less than .5 in 2015). Based on intuition of the model, one can conclude that while Tysons residents can easily find work in Tysons itself, the borough doesn't stand out as a place which can attract households working in other areas. Such a finding is interesting because while Tysons is very well connected to transit opportunities (including Interstate 66, the Dulles Parkway, and the Washington Metro's Silver Line), other major employment centers (especially those which are overall harder to access, i.e. lower FCMA) are still long commutes away. Some possible explanations for this are that Tysons is well connected to transit arteries that serve to provide easy access to only the most desirable commercial areas, and that transit times to DC, Bethesda, and other such areas is still relatively high. Essentially, Tysons is only attractive to those working in major job centers, not those working in smaller offices or similar workplaces scattered around the entire DMV area.

Places with Highest RCMA Rise		Places with Greatest RCMA Fall	
Brookmont CDP, Maryland	639.74%	Springdale CDP, Maryland	-86.30%
North Springfield CDP, Virginia	288.72%	White Oak CDP, Maryland	-69.36%
Rosaryville CDP, Maryland	94.54%	District Heights city, Maryland	-55.29%
Berwyn Heights town, Maryland	78.66%	Lincolnia CDP, Virginia	-51.35%
Chillum CDP, Maryland	77.24%	Andrews AFB CDP, Maryland	-46.71%

Table 2. Top 5 greatest RCMA risers and fallers, 2000-2015

The highest risers in RCMA tend to be inner-belt suburbs, close to but not necessarily served by Metro lines.

Comparing 2000 and 2015 CMAs along the Silver Line

The Silver Line was constructed in 2014 between Falls Church and Reston, passing through McClean and Tysons. For both McLean and Tysons, the FCMA dramatically increased. In the case of Tysons, the FCMA changed from 1.021 in 2000 to 5.879 in 2015, and for McLean, the FCMA changed from .851 to 1.282. McLean and Tysons are adjacent stations on the Silver Line, separated by the Interstate 495 beltway approximately 12 miles west of Central DC. Prior to the 2014 opening of the Silver Line, both localities were already well connected to the suburban DMV highway network, which primarily serves higher-population areas with easier access to work (i.e., a high RCMA). However, these places previously lacked quick access to residential communities that are overall less served (i.e., have a low RCMA). Based on the CMA system of equations, all else being equal, access to underserved residential areas is much more valuable than access to highly served residential areas, since the RCMA shows up in each factor's denominator. Therefore, it is clear to see that in 2015, commute times decreased to places with low RCMA, which helped increase Tysons' and McLean's FCMAs. The DC Metro is unique in that it serves several inner suburban communities that are cut off from the DMV's arterial road network. The Metro is the primary mode of transportation to access job centers for such bedroom communities.

The changes in the RCMA provide a more nuanced discussion. In both Tysons and McLean, the RCMA slightly increased but remained especially low. In Tysons, the RCMA increased from .141 to .170, while in McLean, the RCMA changed from .492 to .774. This means that between 2000 and 2015, Tysons and McLean became slightly more attractive to residents wanting to take jobs in underserved areas (i.e. with more job openings and a low FCMA), but they are still relatively unattractive compared to the DMV as a whole. One possible explanation for this is that the Silver Line, unlike the DC Metro's other lines, primarily serves business centers rather than smaller communities. For instance, the Silver Line serves Reston, Tysons, McLean, Falls Church, and Arlington, before heading into DC. All of these places are longstanding commercial centers with high FCMAs. It is very likely, therefore, that the Silver Line only increased accessibility for Tysons and McLean residents to other highly served job centers, and barely affected accessibility to underserved job centers, many of which are more likely to have increased job openings to attract workers. RCMA does not necessarily have a positive correlation with transportation access, unless this transportation access opens up the possibility of commuting to less developed, underserved job centers.

One extremely important remark, however, is that the Silver Line only opened in 2014, one year before the 2015 timestamp of the data. It is very likely that there could be some lagged growth in later years, as companies and residents adjust to the Silver Line's presence in northern Virginia. For example, comparing 2000 with 2020 data could show a much more significant difference in both RCMA and FCMA in these areas, when the Silver Line would be more mature.

Chapter 5

Regressing DC's RCMA with Home Valuations and Student-Faculty Ratios

Overview of CMA Regression Technique

In the Bogotá model, Nick Tsivanidis concluded that changes in floorspace price have a log-log positive relationship with the RCMA and FCMA. A similar approach is used here, regressing the RCMA and FCMA changes respectively with changes in estimated housing valuations between 2000 and 2015, and changes in per-worker GDP between 2000 and 2015. Additionally, a second factor, known as an "amenity term," is added to each of the RCMA and FCMA regressions. Finally, a third term, which measures the interaction between both the RCMA and FCMA regressions' amenity terms, is added to assess the interdependence of RCMA and amenities when estimating property desirability.

Measuring the RCMA's Added Effect over Student-Faculty Ratios on DC Home Values

To compare the relationship between housing desirability and both commuter market access and amenity quality in the DC area, a similar regression technique to the Bogotá model is used:

$$\Delta ln(EHV_i) = \beta_0 + \beta_1 \Delta ln(\Phi_{Ri}) + \beta_2 \Delta ln(SFR_i) + \beta_3 \Delta ln(\Phi_{Ri}) ln(SFR_i) + \varepsilon_i,$$

where EHV_i is the estimated average home valuation in tract i, and SFR_i is the elementary school student-faculty ratio in tract i.

The EHV values are taken from the Zillow Home Value Index (ZHVI), which aims to provide estimated home appraisal values based on sale data, listing frequency, and land valuation.

The ZHVI has monthly data back to 1996, and for this study, January 2000 and January 2015 are used.

The base case in this study is to determine the strength of the relationship between amenity quality and housing values, without regards to commuter accessibility. Hence, regression is performed on the above equation, with the β_1 and β_3 terms removed, in principle leaving a simple regression between the log change in ZHVI and the log change in the student-faculty ratios (SFRs). SFRs are analyzed in all elementary schools in the 95 included census tracts. This data is taken from SchoolDigger. The SFR has shown a very strong ability to serve as a proxy for public school spending per capita. A higher SFR is an indication that a certain school may be underfunded, while a lower SFR indicates a high school budget per pupil, as more teachers would be instructing the same number of children.

Without inclusion of the CMA terms in the regression model, the R^2 is .4567, indicating that approximately 46% of the variations in the log ZHVI changes can be explained solely by the log SFR changes. This means that school quality can be used as a very strong predictor of housing values; the two work relatively well in synergy. However, as shown in the scatter plot in Figure [x], there is huge variability in the ZHVI changes when little change in the SFR is observed. It is only when the explanatory variable's absolute value becomes large (i.e. more negative) that more pronounced conclusions can be made about increases in ZHVI. Additionally, as expected, the estimate for β_2 is negative, indicating that increased student-faculty ratios have a negative impact on housing prices, since high student-faculty ratios could be a proxy for lowered public-school spending per student.

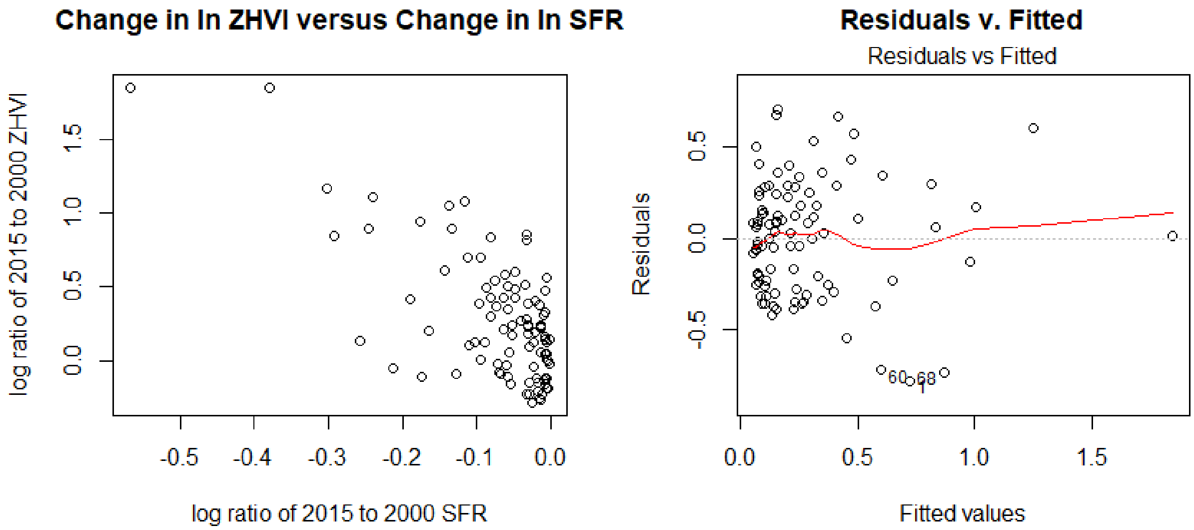

Figure 3. Scatterplot and residuals-fitted plot for simple SFR regression

Interestingly, however, despite the appearance of the scatter plot, the residuals versus fitted plot shows a much more even variance of the residuals, only decreasing at the highest fitted values. This may indicate that lack of univariance may not be a huge issue in this simple regression. When comparing the theoretical with actual quantiles, a slight leftward skew in the residuals is apparent in the distribution of the residuals than a true normal distribution. However, the degree of the skew is rather small.

Next, the full multi-regression model is created, now taking into account accessibility (RCMA). In the full multi-regression model, including the interaction between the RCMA and the SFR, both the estimates for β_1 and β_2 are statistically significant at the .01 significance level. However, β_3, the interaction coefficient, isn't given an estimate sufficiently high enough to conclude that there was interaction in the model (i.e. the null hypothesis of $\beta_3 = 0$ is not refuted). This would indicate that the SFR (and ultimately school spending) is likely independent of the RCMA. One other noteworthy finding is that the estimate for β_2 is -2.8781, much larger in absolute value than that of β_1 (.3982), indicating that the SFR change, pari passu, has a much stronger relationship with home valuations than accessibility to workplaces. β_2 having a negative estimate makes sense because higher SFRs indicate a lower budget per student, translating to a lower quality public education.

The R^2 only slightly increases to .5503 (from .4567 in the simple model), which is another confirmation that the SFR may be a much more reliable explanatory variable to explain differences in home valuations. In the residuals versus fitted plot, shown below, the variability of the residuals around the fitted regression appears to be much more constant, barring three outliers (all numbered). It would seem that, comparing this plot with that of the simple regression, that the differences in SFR likely in fact explain the huge variation in ZHVI values with increasing RCMA values.

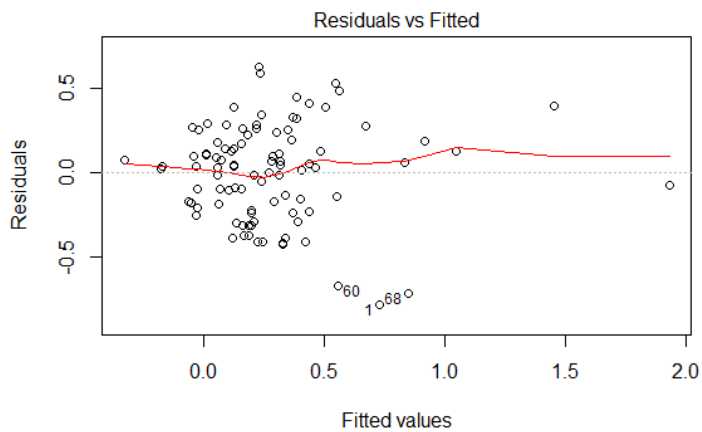

Figure 4. Residuals versus fitted for RCMA multiregression

One interesting plot to check is the scatterplot constructed when just the change in log RCMA is the explanatory variable (i.e. the above regression without the β_2 and β_3 factors). In this case, the R^2 is only .263, which is much lower than that of the simple model with only the SFR. Much of this can be attributed by the fact that increases in the RCMA only seem to increase the upper limit on home value increases, i.e. the variability significantly increases as the RCMA changes get higher. It is only when the RCMA increases significantly that significant home value growth is very likely.

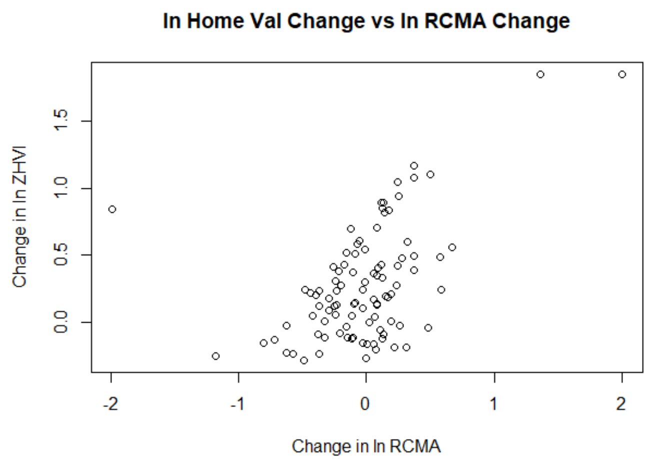

Figure 5. Plot of ln ZHVI change and ln RCMA change

The following table shows the estimates of each of the regression parameters, for both simple models as well as the full model with interaction.

Parameter	Coefficient of:	Simple: SFR	Simple: RCMA	Full Model	p-Value (Full Model)
β_0	Intercept	0.05085	0.29106	0.08841	0.029358
β_1	Change in ln RCMA	-	0.48356	0.03982	0.000171*
β_2	Change in ln SFR	-3.16074	-	-2.87810	0.0000000000618*
β_3	Interaction	-	-	0.51823	0.14335
R-Squared		0.4567	0.2629	0.5503	

Table 3. Regression table of all regressions whose dependent variable is the ZHVI

RCMA: Bogotá Regression Analysis

In his paper, Nick Tsivanidis performs a regression to model the expected change in log floorspace price (FS), i.e. $\ln(FS_{2015}/FS_{2000})$, based on the calculated change in log RCMA, i.e. $\ln(\Phi_{R2015}/\Phi_{R2000})$. He also adds another explanatory variable: $\ln(ES_{2015}/ES_{2000})$, where ES_i

represents the vector of all tracts' college-educated share of working populations for year i. In a similar multiple regression to what is conducted for the DC model, Tsivanidis bases his model off the following equation:

$$\Delta ln(FS_i) = \beta_0 + \beta_1 \Delta ln(\Phi_{Ri}) + \beta_2 \Delta ln(ES_i) + \beta_3 \Delta ln(\Phi_{Ri}) ln(ES_i) + \varepsilon_i.$$

One important deviation in Tsivanidis's regression practice from the DC model is that the simple regression is initially performed with respect to the RCMA (as an explanatory variable) rather than the other factor (in this case, college-educated share). In principle, his base case is the relationship between the RCMA and residential floorspace prices, pari passu.

The simple regression between the floor space and RCMA yields a nearly perfect linear relationship plot, with an R^2 value of .470, indicating that 47% of the variation of the residential price per square meter can be explained by changes in the relative competitiveness to accessing workplaces. For comparison, when constructing a simple regression between the change in log ZHVI and change in log RCMA in the DC area, the R^2 is only .2629.

When adding in the second and third terms (above the simple regression, to account for the ES factor), the R^2 jumps by .27, to .740. Additionally, the β_3 factor, measuring the impact of the interaction term, has a p-value below 10%, but the β_2 factor isn't statistically significant even at the 10% level. These results suggest that in Bogotá, significant synergies exist between residents' education level and relative workplace accessibility. A highly-educated population would need to access high-skilled workplaces, and easy accessibility to a multitude of jobs is key to maintaining higher residential property desirability.

This result is very different when compared to that of the DC model, in which the outside explanatory factor contributes to a much greater share of the housing value than the RCMA.

Additionally, significant synergies seem to be present between the RCMA and an outside factor in the Bogotá model, but not in the DC model.

Implications of Models and Comparisons to Bogotá

One very important thing to note is that the role of the ES factor is not exactly the same as that of the SFR factor for the DC model. While the SFR can be used reliably as a proxy for amenity quality, the ES serves as a better proxy for overall neighborhood quality, including lowered crime rates and gentrification. However, it is nonetheless interesting to compare the two models, as it has been shown that across the world, highly educated parents place a much greater emphasis on their children's education.[22] While the college-educated share may not be as good of a proxy as the student-faculty ratio when assessing school quality, psychological factors could play a key role in providing synergies between highly educated parents and neighborhood school quality.

Running both the simple and multiple regression models shows that in the DC area, school spending is most likely a much bigger factor in determining residential desirability (i.e. home prices) than just workplace accessibility. One possible explanation for this is that in the United States, public schools are primarily funded through real estate taxes (which themselves are a function of property appraisal valuations). If public schools get increased sums of residential tax revenue, it would me more likely for them to increase spending per student, ultimately leading to the hiring of new teachers (lowered student-faculty ratios). Additionally, such an effect serves to increase school ratings, which many US families take a look at when relocating. US public education is very unequal from town to town, and such ratings serve as an important proxy for neighborhood quality (and therefore house prices). In principle, therefore, student-faculty ratio decreases likely work in tandem with increases in home values; neither one by itself causes the other.[23]

In Colombia, however, the public education system is centrally administered by the *Ministerio de Educación Nacional*, pooling funds from generalized public service taxes. While these taxes do come from resident incomes, there are no school districts that can apportion spending to certain areas; all districting is done by the central system. Additionally, in 2014, only 52.1% of all students attended a school in the public school system (compared to 89% in the United States).[24] These factors make school quality much less likely to be tied to neighborhood location than in the United States, where most students attend public schools and individual districts receive specific mill taxes.

One final note is that residential price per square foot is not the same thing has total home valuation, and the former may be a better indication of desirability. However, in the DC area, new home construction is comparatively slower than that in Bogotá, and the model only considers changes in the values, not the values as a whole.

Chapter 6

Regressing DC's FCMA with GDP per Worker and Growth in Employment

Measuring the FCMA's Added Effect over Business Growth on DC Per-Worker Productivity

To compare the relationship between worker productivity and both firm commuter market access and business growth in the DC area, a regression model was constructed very similar to that used to analyze the added effects of the RCMA:

$$\Delta ln(PWGDP_i) = \beta_0 + \beta_1 \Delta ln(\Phi_{Fi}) + \beta_2 \Delta ln(L_{Fi}) + \beta_3 \Delta ln(\Phi_{Fi}) ln(L_{Fi}) + \varepsilon_i,$$

where $PWGDP_i$ is the gross domestic product per worker (GDP per worker) in tract i, and L_{Fi} is the number of individuals working in tract i. Note that while the entire L_F vector is required to construct both the RCMA and FCMAs, the presence of a large number (95) of analyzed tracts should make the effect from the CMA system of equations negligible in a regression model.

The per-worker GDP is calculated by summing together the taxable earnings (i.e. after interest) of all businesses in zip codes contained in a tract i. Commercial taxable earnings are available on the Internal Revenue Service (IRS) zip code data for every year going back to 1998.

Similar to the process used when analyzing the added effect of the RCMA, the initial regression leaves out the β_1 and β_3 terms, effectively making this a simple regression between the per-worker GDP and employment growth. Without inclusion of the FCMA terms, the regression only produces an R^2 of .0937, indicating that less than 10% of the variation in the PWGDP changes can be attributed to solely employment growth. The estimate for β_2 is .057, indicating that the employment changes have a minor positive relationship with the PWGDP.

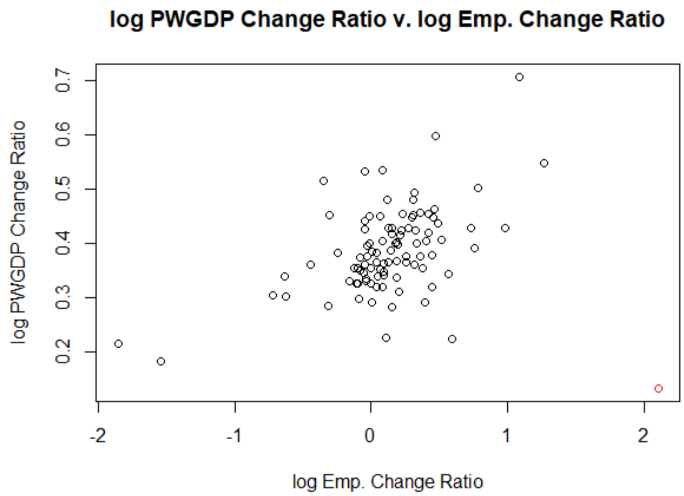

Figure 6. PWGDP versus employment scatter plot

Interestingly, based on this plot, there is one clear outlier, marked in red, which represents a tract with an extremely high employment change ratio. This point corresponds to Springdale, Maryland, which is an extremely small municipality that is located immediately to the east of the Interstate 495 Beltway and is an extremely residential community with low employment. It is very possible that this extreme employment change ratio may be due to a couple retail and restaurant outlets opening in town, many of which mirror those already present.

With this in mind, Springdale is removed from this regression, and the R^2 jumps handedly to .28 from .09. Additionally, the estimate for β_2 rises to .0997. Based on the analysis of the outlying tract and its geographic makeup, a more accurate relationship can reasonably be derived by ignoring it. Below shows the updated scatter plot, as well as the residuals v. fitted plot, without the outlier. There does not seem to be too much of an issue with non-univariance of the residuals in the second plot below.

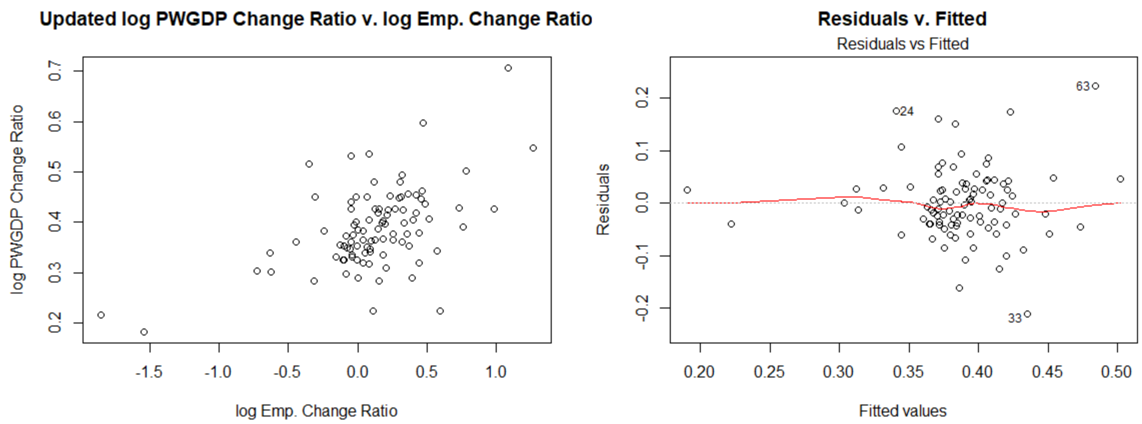

Figure 7. Updated scatter plot and residuals v. fitted plot, after removing outlier Springdale

While keeping Springdale (the outlier) out of the model, the FCMA values, along with the interaction term (the β_1 and β_3, respectively) are added to the linear regression, giving the full equation introduced in the beginning of this section. The R^2 jumps from .280 to .459, a substantial increase. Additionally, the β_2 estimated value is no longer statistically different from zero (the p-value is .683), indicating that removing the employment change from the model would yield a similarly strong relationship. In fact, when the employment change is in fact removed, a simple regression model with only the β_1 term produces an R^2 of .451, and the estimate of the β_1 value only changes from .1098 to .1129. As expected, when plotting just the PWGDP change and FCMA change (with Springdale still removed), a considerably stronger relationship is visible when compared to the simple regression with solely the employment changes. There may be a slight increase in residual variance in the middle range of the fitted values, but it does not seem significant enough to warrant model transformations. β_3, the interaction coefficient, does not have an estimate that is statistically significant from 0. Overall, a log-log change model seems to capture the relationship between PWGDP and FCMA rather reliably.

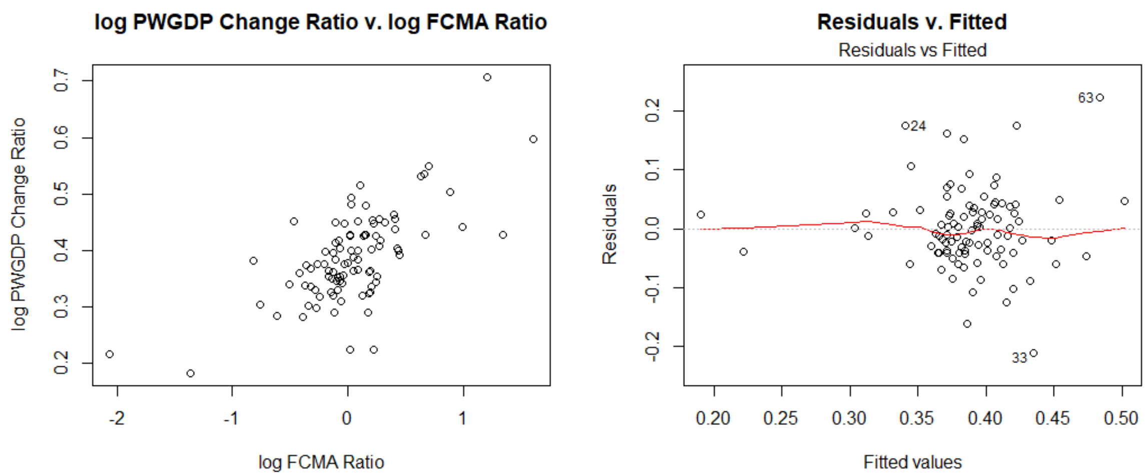

Figure 8. Updated scatter plot and residuals v. fitted plot, considering only FCMA

The following table shows the estimates of each of the regression parameters, for both simple models as well as the full model with interaction.

Parameter	Coefficient of:	Simple: Emp.	Simple: FCMA	Full Model	p-Value (Full Model)
β_0	Intercept	0.37549	0.38315	0.37958	less than 2E-16*
β_1	Change in ln FCMA	-	0.11287	0.10975	0.000000263*
β_2	Change in ln Emp.	0.09978	-	0.00995	0.683
β_3	Interaction	-	-	0.01553	0.245
R-Squared		0.28	0.451	0.459	

Table 4. Regression table for firm-side regressions, with PWGDP as the explanatory variable

FCMA: Bogotá Regression Analysis

Similar to his analysis on RCMA, Tsivanidis performs a regression to model the expected change in log commercial floorspace price (CFS), i.e. $\ln(CFS_{2015}/CFS_{2000})$, based on the calculated change in FCMA , i.e. $\ln(\Phi_{F2015}/\Phi_{F2000})$. and various other explanatory variables, including the number of commercial establishments (F), i.e. $\ln(F_{2015}/F_{2000})$. Importantly, unlike the analysis done in this paper for DC's FCMA and labor growth, Tsivanidis initially considers just FCMA in the model as a base case, then adds in the added effect from growth in the number of firms. In his following equation,

$$\Delta ln(CFS_i) = \beta_0 + \beta_1 \Delta ln(\Phi_{Fi}) + \beta_2 \Delta ln(F_i) + \beta_3 \Delta ln(\Phi_{Fi}) ln(F_i) + \varepsilon_i,$$

the β_2 and β_3 terms are initially left out.

The simple regression in the Bogotá model, considering the FCMA as the only explanatory variable for the commercial floorspace price, the R^2 is .223. Noting that commercial floorspace price and per-worker GDP, while not the same quantity, can serve as reliable proxies for each other, the DC model shows a much higher R^2 of .451 when considering solely the FCMA. This is a key indication that location and competitiveness to accessing workers purely on a geographic sense may not be as essential in Bogotá as in DC. In fact, the R^2 in the DC model was more than double that of the Bogotá model.

More interestingly, when the full regression model is considered, adding back the β_2 and β_3 terms, the R^2 jumps to .365, which is a notable increase, and indicates that a consequential relationship may exist between solely the growth in the number of firms and commercial floorspace prices. One essential remark to make is that the number of firms is not equal to the number of laborers, nor does it necessarily serve as the best proxy. This is because a certain company can increase its employment level or acquire competitors or other nearby firms. However, in the case of Bogotá, acquisitions are rare, and a much higher proportion of employment is through small businesses and contract work, therefore making firm count a more realistic proxy for total employment level.

Implications of Models and Comparisons to Bogotá

The biggest takeaway from the initial calculations and analysis of the multi-regression models for both DC and Bogotá is that FCMA likely matters a lot more to business productivity and desirability in the DC area than in Bogotá. The R^2 of the simple Bogotá model considering only FCMA is much smaller than that of the simple DC model. In fact, the R^2 of the multiple regression Bogotá model, considering both the number of establishments and FCMA, is still considerably lower than even the simple regression model for DC.

One very likely explanation for this is crime, and specifically a much higher variability of crime rates in Bogotá than in DC. According to Mangai Natarajan, an urban planning researcher at the John Jay College of Criminal Justice, the variance in crime rates across major US metropolitan areas is approximately 46% less than in developing countries, including those in Latin America and Asia. Additionally, Natarajan remarks that crime rates (specifically, felony-equivalent charges) are both a significant bane to both residential and commercial desirability, and the bane for commercial desirability jumps staggeringly if the crime rate becomes significant.[25] The DC area is a notably safe region, with only DC itself seeing a significant higher crime rate per capita (1,244 violent crimes per 100,000 residents) than the national average.[26] Meanwhile, as discussed in Tsivanidis's paper, Bogotá's crime map would look almost like a mosaic, outside the city center, with many suburban areas lacking in law enforcement presence and funding. This could be a principal reason stalling commercial floor space increases in some of the outer regions, even those newly connected by the Transmilenio.

Chapter 7

Conclusions

Strengths and Limitations of CMA Value Interpretation in DC Area

Based on the calculations done for this study, it is safe to conclude that Tsivanidis's reduced-form CMA model, initially constructed for an emerging economy (Bogotá, Colombia), can give policyholders in developed economies a much clearer picture in understanding firms' relocation preferences. However, this model's scope in capturing nuances in residential preferences in higher income economies is much more limited.

It is clear that the FCMA can very readily serve as a proxy for a locality's relative ability and competitiveness to attract workers (its initial purpose as outlined by Tsivanidis). The highest FCMAs are observed in edge cities that have access to both expressway systems and the Washington, DC Metro. Such is precisely why Tysons, VA, which was recently connected by the Silver Line, has among the highest growth in FCMA (from 1.021 to 5.879). Previously, Tysons was relatively isolated from Dulles International Airport and Central DC, but that is no longer the case.

The RCMA, however, while producing favorable results showing a positive relationship with transportation access, is much more nuanced. Some of the highest RCMA values are located in inner-belt suburbs on metro lines, but self-contained communities dotting the region. On the other hand, while the RCMA of places along the newly constructed Silver Line, such as Tysons and McLean, did increase, they were still significantly below the average. Based on the intuition of Tsivanidis's model, Tysons and McLean residents are still not well connected to underserved jobsites, where hiring rates and relative wages may be higher. However, this analysis outlines a key limitation of the RCMA in the DC area: all jobs are alike (or still stratified at skilled versus unskilled), and those that are harder to access will have higher wages to clear markets. In the DC area, white-collar businesses (like finance, consulting, and law firms) are heavily clustered, even outside the District itself. Several white-collar employees, therefore, are likely to relocate nearby these jobs. What is more is that each of these white-collar nodes are well connected by the Metro. There are a multitude of condominium complexes being constructed around Tysons and McLean, where major companies like Capital One are headquartered, because prestigious companies in the DC area are clustered together. In Bogotá, on the other hand, while there is a small contingent of white-shoe firms in the city center, the majority of businesses are clerk shops and unskilled worker contractors. Such businesses are not clustered and are instead dispersed around the city.

Regression Implications

It is important to note that in Tsivanidis's paper, the amenity regression is based on reduced-form implications of worker types and choices. Tsivanidis also performs a structural analysis (using maximum likelihood to evaluate general model parameters) of the CMA model in which residents use their own characteristics to make decisions on where to work and live. Due to lack of data availability, specifically to capture multiple nuances of a developed economy like the DMV area, only the simplified model (rather than the full structural model) is run for the DC area in this study. Therefore, the extent of the regression analysis, beyond implications of the CMA values themselves, is relatively limited in scope.

The distributional nature of the CMA system of equations implies that in general, changes in per-unit floorspace prices have a log-log relationship with changes in amenity factors, neighborhood quality factors, and CMA values. In this regard, the FCMA regression done on DC seems to encapsulate a log-log relationship with the per-worker GDP (a good proxy for commercial floorspace prices). Much of this could be attributed to the reasons that certain FCMA being high or low could be backed up by transportation access and nearby residential population.

The RCMA regression exhibits a slightly weaker log-log relationship, that is still slightly quadratic or exponential in nature at the highest values. Only if the RCMA increases drastically can one conclude that the housing values will likely increase; besides that, it is clear that several other factors serve as a much better proxy. One such factor is public school quality (proxied by the student-faculty ratio). When added to the regression alongside the RCMA, the RCMA becomes an insignificant explanatory factor that can explain housing prices in the DC area. Much of this has to do both with the shortcomings of the RCMA's implications in the DMV (and other more developed economies), as well as the fact that Colombia's public education is funded and administered through the centralized *Ministerio de Educación Nacional*, as opposed to individual school districts in the United States.

Potential Extensions to Study

Perhaps the biggest limitation to this study is the lack of easily accessible data (i.e. microdata) to give individuals significantly differing characteristics that would make their travel decision functions endogenous as in Tsivanidis's model. For example, on the business side, there are several nuances in business revenue, workforce dynamics, and real estate prices, that could all be placed into Tsivanidis's general model functions to single out the effect of transit accessibility. On the residential side, there are countless other variables at play, which are mostly independent of per-student school spending, that can capture much of the remaining variation in residential valuations. Microdata at this level is available by the United State Census Bureau, but huge monetary liabilities may be associated with the use of such data.

A more realistic extension to this study, given the data at hand, would be to loosen some of the restrictions in Tsivanidis's simplified model, especially to include multiple job types. Compared to one job type in the simplified CMA model used in this paper, or two job types (skilled and unskilled) in Tsivanidis's generalized model, a model with k different job types could help the RCMA be a much more accurate proxy of residential attractiveness (i.e. the RCMA) in more advanced economics. Such job types can include finance, services, manufacturing, retail, and hospitality, among others. The Census Transportation Planning Package includes such data for businesses in all 95 tracts for both 2000 and 2015. The clustering nature of white-shoe jobs (in downtowns or "edge cities") or retail jobs (in malls) in developed economies provides a huge limitation in interpreting RCMA values, which depend only on access to underserved jobs with high salaries, even though all jobs are the same.

An additional further avenue could be to construct a lagged time series regression of the RCMA and FCMA changes starting in 2000, as the system is still being developed, and projecting prediction intervals for the CMA values for a 5- or 10-year timeframe. For example, the Silver Line, which opened in July 2014, just one year before the end data, may have not caused a full effect on either the CMA or related factors, as the population would still be adjusting to the new service. Five or ten years later, however, the Silver Line would be much more mature. The key here would be to see if a prediction interval system could reliably predict future CMA and amenity values, confirming them by the actual values five or ten years after the initial time series construction.

Appendix A

In-Text Citations

1. DC Metro Average Weekly Boardings by Station.
2. Forkenbrock, David J., and Glen E. Weisbrod.
3. Lewis, Paul.
4. Ibid.
5. Jara-Diaz, Sergio.
6. Forkenbrock, David and Glen Weisbrod.
7. Ibid.
8. Ibid.
9. Li, Zheng.
10. Mohammad, Sara I., et al.
11. Tryfos, Peter.
12. Cervero, Robert, and Michael Duncan.
13. Ko, Kate, and Xinyu Cao.
14. Ibid.
15. Tsivanidis, Nick.
16. Ibid.
17. DC Metro Average Weekly Boardings by Station, 1977-2018.
18. Li, Zheng.
19. Ibid.
20. Lewis, Paul.
21. Ibid.

22. Mohammad, Sara I., et al.

23. Ibid.

24. Immerstein, Silje.

25. Natarajan, Mangai

26. Federal Bureau of Investigation

Appendix B

CMA Values for all 95 Municipalities

Place Name	2015 RCMA	2000 RCMA	2015 FCMA	2000 FCMA
Washington city, District of Columbia	0.422155032	0.379767179	2.36872047	2.633609575
Adelphi CDP, Maryland	3.224641212	2.87345283	0.315629608	0.353492935
Andrews AFB CDP, Maryland	0.140890656	0.264369606	7.075184494	3.774761258
Aspen Hill CDP, Maryland	4.976171399	4.411482102	0.202350547	0.217652211
Beltsville CDP, Maryland	0.531831853	0.491441659	1.882169713	2.05462734
Berwyn Heights town, Maryland	3.000268737	1.679358374	0.342697285	0.521034163
Bethesda CDP, Maryland	0.368439901	0.431689843	2.714001825	2.319430809
Bladensburg town, Maryland	1.826673716	1.816039848	0.551882203	0.583930926
Bowie city, Maryland	1.751897132	2.541864979	0.57146826	0.382454395
Brentwood town, Maryland	2.24587766	1.746551681	0.437340768	0.482066485
Brookmont CDP, Maryland	4.587610289	0.620166833	0.193839792	1.523809619
Burtonsville CDP, Maryland	2.012524687	1.617470606	0.498511403	0.637695506
Calverton CDP, Maryland	1.095954101	0.930905091	0.909174993	1.075989203
Camp Springs CDP, Maryland	2.315042451	1.694054385	0.429750068	0.571135479
Capitol Heights town, Maryland	0.958404789	1.70524184	1.038897704	0.530033822
Cheverly town, Maryland	0.931018587	1.387143898	1.079441431	0.728118754
Chevy Chase town, Maryland	2.885673445	2.42102399	0.322762478	0.732715271
Chevy Chase CDP, Maryland	0.98525692	1.223287022	1.018225357	0.827368166
Chillum CDP, Maryland	6.039772578	3.407773575	0.16464632	0.305309573
Clinton CDP, Maryland	1.710624028	1.720971347	0.583966115	0.585993091
Cloverly CDP, Maryland	3.7398909	5.390581119	0.266082826	0.069376919
Colesville CDP, Maryland	4.56494732	3.506086685	0.226486283	0.31824556
College Park city, Maryland	0.442054553	0.320843245	2.26254059	3.122894905
Coral Hills CDP, Maryland	8.257497665	5.690774676	0.129165339	0.115948338
District Heights city, Maryland	2.692514202	6.021650503	0.3656818	0.151131285
East Riverdale CDP, Maryland	6.474861018	4.009343692	0.153776877	0.24575236
Fairland CDP, Maryland	2.357654164	2.050467549	0.421042199	0.488539046
Forest Glen CDP, Maryland	1.154201077	1.29021584	0.876069681	0.833092584
Forest Heights town, Maryland	4.444451923	4.926652647	0.230707314	0.224264787
Forestville CDP, Maryland	1.261267554	1.186593398	0.792565831	0.867918018
Fort Washington CDP, Maryland	2.259936115	3.49053412	0.442366995	0.289807318
Friendly CDP, Maryland	7.218993124	6.835781138	0.137394896	0.141140025
Glenarden city, Maryland	2.17969856	3.567302583	0.4628325	0.371805645
Glenn Dale CDP, Maryland	2.555107665	2.194537752	0.393505971	0.447636981
Greenbelt city, Maryland	0.926304944	0.854737195	1.076637917	1.156992181
Hillandale CDP, Maryland	0.787998309	0.616986572	1.266059096	1.655289975

Hillcrest Heights CDP, Maryland	5.636303149	5.517949907	0.180083154	0.20065428
Hyattsville city, Maryland	0.791959962	0.997927875	1.26014095	0.992378125
Kemp Mill CDP, Maryland	2.683895551	2.928707915	0.370466593	0.282992894
Kettering CDP, Maryland	3.177365793	4.246383043	0.314612159	0.204336895
Lake Arbor CDP, Maryland	0.819380374	0.972386055	1.217627592	1.012176933
Langley Park CDP, Maryland	5.490449403	7.067173441	0.17951296	0.114644679
Largo CDP, Maryland	1.901475042	2.398452554	0.525772981	0.463537952
Marlow Heights CDP, Maryland	1.131232955	1.250489017	0.891631224	0.73625549
Marlton CDP, Maryland	8.29149188	11.51587763	0.126322507	0.025408511
Mitchellville CDP, Maryland	1.227461922	0.967271778	0.81421132	1.072762349
Mount Rainier city, Maryland	5.083358792	6.581723742	0.193486423	0.162810913
New Carrollton city, Maryland	2.793559706	4.501945038	0.357019742	0.273360631
North Bethesda CDP, Maryland	0.579737049	0.401249774	1.725129233	2.488211281
North Kensington CDP, Maryland	3.027413765	2.681626701	0.333988862	0.378553131
Potomac CDP, Maryland	1.142677661	1.394387314	0.875729804	0.715076647
Riverdale Park town, Maryland	1.056148307	1.415619588	0.9537581	0.72406723
Rockville city, Maryland	0.478541108	0.393874756	2.090558667	2.538260662
Rosaryville CDP, Maryland	6.754477523	3.47208539	0.146648744	0.314340048
Seat Pleasant city, Maryland	2.671384602	2.988649665	0.373868777	0.367522441
Silver Spring CDP, Maryland	1.027847238	1.305604204	0.972670054	0.765176798
South Kensington CDP, Maryland	3.469893585	3.269346403	0.28047077	0.246035629
South Laurel CDP, Maryland	3.269637038	3.509500955	0.306567211	0.282051586
Springdale CDP, Maryland	7.239667717	52.85341466	0.097071626	0.180909232
Takoma Park city, Maryland	1.331738646	1.481586045	0.753644275	0.690244155
Temple Hills CDP, Maryland	4.313954274	6.593762359	0.235804201	0.157124455
Travilah CDP, Maryland	4.347102983	3.833211791	0.224198706	0.218070165
Walker Mill CDP, Maryland	6.990080788	6.495773694	0.147397826	0.054514747
White Oak CDP, Maryland	0.79851307	2.60599541	1.25112348	0.376247277
Woodmore CDP, Maryland	2.330258321	3.358667704	0.424077279	0.350278301
Alexandria city, Virginia	0.936179379	0.948672917	1.068795268	1.054528343
Annandale CDP, Virginia	1.547911468	1.071644382	0.646334995	0.933326112
Arlington CDP, Virginia	0.763687565	0.706026156	1.309517624	1.4164858
Bailey's Crossroads CDP, Virginia	1.223837124	1.006509725	0.816215725	0.968700791
Belle Haven CDP, Virginia	1.61294746	1.44249355	0.609603245	0.698471909
Burke CDP, Virginia	3.677709566	4.162454635	0.272354133	0.25126882
Dunn Loring CDP, Virginia	1.887507593	1.986422855	0.523933903	0.453322245
Fairfax city, Virginia	0.346883278	0.378609279	2.881324616	2.639510002
Falls Church city, Virginia	0.653039535	0.669933235	1.529483204	1.467065407
Fort Hunt CDP, Virginia	3.971628533	4.094685233	0.253716818	0.261081067
Franconia CDP, Virginia	1.67465924	3.126837711	0.595006465	0.307487286
Great Falls CDP, Virginia	2.119870567	1.609943098	0.473991293	0.58412489
Groveton CDP, Virginia	3.049791242	3.542504866	0.326131322	0.272175903

Huntington CDP, Virginia	4.815129224	3.329619911	0.209200822	0.309900067
Hybla Valley CDP, Virginia	1.651317654	2.032483208	0.60234467	0.494394915
Idylwood CDP, Virginia	2.986559457	2.341899163	0.335010635	0.462436148
Lake Barcroft CDP, Virginia	2.911735107	2.670727379	0.345116561	0.389943812
Lincolnia CDP, Virginia	1.521487932	3.127543248	0.657136902	0.326709109
McLean CDP, Virginia	0.77499175	0.492014162	1.281771912	1.37155098
Mantua CDP, Virginia	2.09955699	3.059871161	0.472545402	0.378773415
Merrifield CDP, Virginia	0.56421558	0.341298509	1.771962104	2.932336442
North Springfield CDP, Virginia	4.567977152	1.175130294	0.218330198	0.851385435
Oakton CDP, Virginia	2.012454813	1.859370464	0.4962376	0.524196672
Pimmit Hills CDP, Virginia	2.307610087	2.014243076	0.434617164	0.451982825
Seven Corners CDP, Virginia	1.814146883	2.519906022	0.551733724	0.429351935
Springfield CDP, Virginia	0.641574343	0.658818407	1.558033498	1.515900532
Tysons Corner CDP, Virginia	0.170316696	0.141315468	5.879268608	5.8015906
Vienna town, Virginia	0.586181672	0.51411845	1.706977939	1.945192592
West Springfield CDP, Virginia	2.759752964	3.514275797	0.364580867	0.265612524
Wolf Trap CDP, Virginia	3.824227235	4.48600853	0.261434965	0.223313679

Appendix C

Newton-Raphson Algorithm Explanation

The Newton-Raphson Algorithm provides a mathematical framework to solve a system of nonlinear equations, in which simple matrix inversion would not suffice. The following system of 190 equations is an example of such a case:

$$\Phi_{Ri} = \sum_j e^{-\theta d_{ij}} \frac{L_{Fj}}{\Phi_{Fj}}$$

$$\Phi_{Fj} = \sum_i e^{-\theta d_{ij}} \frac{L_{Ri}}{\Phi_{Ri}}$$

$$i, j \in \{1,2, \dots .95\}.$$

One method that may converge to a solution (which works with the data for this study) is known as the Newton-Raphson algorithm. First, one needs to start with defining the vector of all variables as a function of all the variables, i.e. $\mathbf{0} = F(\boldsymbol{\Phi})$, where $\boldsymbol{\Phi}$ has 190 individual components. Next, one needs a starting vector from which the algorithm can iterate closer to the solution. For this case, a vector of length 190, with each component equaling one, would suffice.

To easily get $F(\boldsymbol{\Phi})$ from the data, a design matrix is first constructed, which consists following 190×190 matrix where each "cell" represents 95×95 elements.

$$\boldsymbol{D} = \begin{pmatrix} \mathbf{0} & \boldsymbol{e}^{-\boldsymbol{\theta d}}\boldsymbol{L_F}) \\ \boldsymbol{e}^{-\boldsymbol{\theta d}}\boldsymbol{L_R} & \mathbf{0} \end{pmatrix}$$

Here, the lower left and upper right "elements" represent all possible distances from tracts i to j, multiplied by the number of workers or residents in each of the tracts (arranged vertically).

From this, it follows that:

$$\mathbf{0} = F(\boldsymbol{\Phi}) = \boldsymbol{D} \times \left(\frac{\mathbf{1}}{\boldsymbol{\Phi}}\right) - \boldsymbol{\Phi},$$

where $(\mathbf{1}/\boldsymbol{\Phi})$ is a 190×1 vector with each element equaling $1/\Phi_i$.

Using a vector of $\mathbf{1}$ as the initial guess, the Newton-Raphson method iterates over the following procedure:

$$\boldsymbol{\Phi}_{\boldsymbol{n}+\mathbf{1}} = \boldsymbol{\Phi}_{\boldsymbol{n}} - \boldsymbol{J}(\boldsymbol{F}(\boldsymbol{\Phi}_{\boldsymbol{n}}))^{-1}\boldsymbol{F}(\boldsymbol{\Phi}_{\boldsymbol{n}})$$

where $\boldsymbol{J}$ is the Jacobian matrix of $\boldsymbol{F}$.

There are instances, based on the stability of the inverses required in this procedure, in which the Newton-Raphson method does not converge to a solution. However, in this study, 100 iterations do suffice in finding the solution vector $\boldsymbol{\Phi}$.

Appendix D

Links to Code and Master Data File

All code and a master file of used data can be found in the following Git Hub link:

https://github.com/apnowithae/CMA-Thesis-Code-and-Data

BIBLIOGRAPHY

Cervero, Robert, and Michael Duncan. “Transit’s Value-Added Effects: Light and Commuter Rail Services and Commercial Land Values.” Transportation Research Record: Journal of the Transportation Research Board, vol. 1805, no. 1, Jan. 2002, pp. 8–15. DOI.org (Crossref), doi:10.3141/1805-02.

DC Metro Average Weekly Boardings by Station, 1977-2018. 6 Jan. 2019. Raw data. Washington Metropolitan Area Transportation Authority, Washington, DC.

Forkenbrock, David J., and Glen E. Weisbrod. Guidebook for Assessing the Social and Economic Effects of Transportation Projects. National Academy Press, 2001.

Georgia Institute of Technology, and Arthur Nelson. “Transit Stations and Commercial Property Values: A Case Study with Policy and Land-Use Implications.” Journal of Public Transportation, vol. 2, no. 3, Sept. 1999, pp. 77–95. DOI.org (Crossref), doi:10.5038/2375-0901.2.3.4.

Hanson, Susan. The Geography of Urban Transportation. 4th ed. London: Guilford, 1995. Print.

Immerstein, Silje. “Education in Colombia.” *WENR*, 7 Dec. 2015, https://wenr.wes.org/2015/12/education-in-colombia.

Jara-Diaz, Sergio R. “ON THE RELATION BETWEEN USERS’ BENEFITS AND THE ECONOMIC EFFECTS OF TRANSPORTATION ACTIVITIES.” Journal of Regional Science, vol. 26, no. 2, May 1986, pp. 379–91. DOI.org (Crossref), doi:10.1111/j.1467-9787.1986.tb00826.x.

Junge, Jason R., and David Levinson. “Financing Transportation with Land Value Taxes: Effects on Development Intensity.” Journal of Transport and Land Use, vol. 5, no. 1, Apr. 2012, pp. 49–63. DOI.org (Crossref), doi:10.5198/jtlu.v5i1.148.

Ko, Kate, and Xinyu Cao. “The Impact of Hiawatha Light Rail on Commercial and Industrial Property Values in Minneapolis.” Journal of Public Transportation, vol. 16, no. 1, Mar. 2013, pp. 47–66. DOI.org (Crossref), doi:10.5038/2375-0901.16.1.3.

Lewis, Paul. "It's Not Too Late for a Better Silver Line Contract." The Washington Post. WP Company, 14 Dec. 2018. Web. 10 Apr. 2019.

Li, Zheng. “The Impact of Metro Accessibility on Residential Property Values: An Empirical Analysis.” Research in Transportation Economics, vol. 70, Oct. 2018, pp. 52–56. DOI.org (Crossref), doi:10.1016/j.retrec.2018.07.006.

Martínez, L. Miguel, and José Manuel Viegas. “Effects of Transportation Accessibility on Residential Property Values: Hedonic Price Model in the Lisbon, Portugal, Metropolitan Area.” Transportation Research Record: Journal of the Transportation Research Board, vol. 2115, no. 1, Jan. 2009, pp. 127–37. DOI.org (Crossref), doi:10.3141/2115-16.

Meyer, Gereon, and Susan Shaheen. Disrupting Mobility Impacts of Sharing Economy and Innovative Transportation on Cities. Cham: Springer International, 2018. Print.

Mohammad, Sara I., et al. “A Meta-Analysis of the Impact of Rail Projects on Land and Property Values.” Transportation Research Part A: Policy and Practice, vol. 50, Apr. 2013, pp. 158–70. DOI.org (Crossref), doi:10.1016/j.tra.2013.01.013.

Natarajan, Mangai. “Crime in Developing Countries: The Contribution of Crime Science.” *Crime Science*, May 2016.

“Table 4.” *FBI*, https://ucr.fbi.gov/crime-in-the-u.s/2014/crime-in-the-u.s.-2014/tables/table-4/table_4_crime_in_the_united_states_by_region_geographic_division_and_state_2013-2014.xls. Accessed 15 Apr. 2020.

Tryfos, Peter. Methods for Business Analysis and Forecasting Text and Cases. New York, NY: John Wiley & Sons, 1998. Print.

Tsivanidis, Nick. “The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá’s TransMilenio .” Dartmouth College, vol. 1, no. 1, Feb. 2019.

ACADEMIC VITA

Atharv Gupte

atharvgupte@gmail.com

Education

The Pennsylvania State University, Schreyer Honors College **State College, PA**
Bachelor of Science in Mathematics 2016-2020
Minor in Economics and Statistics

- **Relevant Coursework:** Multivariate probability theory, Calculus-based intermediate economics, Real analysis

Work Experience

General Reinsurance Corporation **Stamford, CT**
Actuarial Intern (Life Division) Jun-Aug 2019

- Helped mitigate reserve margin requirements of 12 carriers by over 15% through matching prescribed risk classes with similar relative mortalities, calculated from the Society of Actuaries (SOA) Relative Risk Tool
- Streamlined mortality improvement efforts through the development of a looping macro which performed reasonability checks on 32 life tables for select period duration, attained age, and demographic metrics
- Fixed premium mapping of a multipolicy deal through inspecting treaties and coding in the GGY AXIS pricing software

Unum Group **Portland, ME**
Actuarial Intern, Long-Term Care Jun-Aug 2018

- Optimized a disabled-life reserve model's reliability and performance through testing and comparing algorithm design in the Prophet financial modeling software with an Excel/VBA-based model replicator
- Stratified data table formation and organization for 980,000 policyholders through an SQL-based model tabulation
- Supported experience analysis in multiplier factor research by performing a 7-year monthly backcast of claim data

Comfort Inn and Suites Allentown **Allentown, PA**
Summer Management Intern May-Aug 2017

- Increased revenue per available room in 120-unit hotel by analyzing reports on occupancy and pricing strategy of competition
- Checked in and provided exceptional service to approximately 60 parties daily, ranging from families to business travelers

Leadership Experience

Penn State Real Estate Society **State College, PA**
Club President and Head of Analysis Dec 2017-Present

- Foster an interest in markets and real estate analysis through hosting guest speakers and developing investment competitions
- Spearhead a 30-student organization through delegating treasury and social responsibilities with 5 club officers
- Added depth and discussion to student meetings by innovating on presentation media through use of simulations and workshops

Schreyer Honors College Freshman Orientation **State College, PA**
Mentor Feb-Aug 2017

- Selected from competitive application to assist 12 incoming freshmen with insight on Schreyer Honors College culture
- Served on the Move In and Arrival team – Managed logistics of room key handling, showed incoming students room/class locations, and helped oversee campus tour of 300 incoming Penn State Schreyer Honors students

SKILLS, ACTIVITIES & INTERESTS

Languages: Fluent in English, Marathi; Limited Proficiency in French
Technical Skills: R, Python, Visual Basic, SQL
Activities: Real Estate Society President, Actuarial Science Club Senior Mentor, Schreyer Honors College Mentor, Green Studies