

The Mode Most Traveled: Parking Implications and Policy Responses

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Engineering and Public Policy

Olutayo G. Fabusuyi

B.S., Computer Science, University of Ife, Nigeria
M.S., Public Policy & Management, Carnegie Mellon University
M.Phil. Economics, University of Oxford, United Kingdom

Carnegie Mellon University
Pittsburgh, PA

December, 2016

Acknowledgement

The last three years have been time well spent – there is nothing I would rather have done than spend these years in getting my Ph.D. Having said this, it would be disingenuous to take ownership for why the process – both journey and destination – has been so fulfilling. I have always argued that a Ph.D. program is a marriage with limited shelf life with the student as the junior partner – I am yet to hear of an analogy more fitting. One could come out of it with feelings of nostalgia or absolutely hating it.

That the former is my case, is in large part attributable to my primary supervisor – Robert Hampshire. Robert has an uncanny ability to know when to step in to offer advice, guidance or direction on a project and when to give me the latitude to do things my way. More importantly, this thesis is informed to some degree by his innovative ideas, and the quality of the product has benefited immensely from the insights and inputs he has provided.

I would also like to express my gratitude to my thesis committee co-chair – Granger Morgan. From the get-go, Granger provided suggestions on navigating the Ph.D. program and provided valuable input during the thesis proposal and write-up stage. The contribution of my other doctoral committee members – Mitch Small and Sean Qian is also gratefully acknowledged. Their suggestions and viewpoints have enriched the thesis quality. The contribution of two of my coauthors in an earlier paper – Katsunobu Sasanuma and Tori Hill is also recognized. Finally, the department of Engineering and Public Policy provided me with the right environment, and I have benefited immensely from conversations with faculty, staff and students of the department during the program.

I have been blessed to have come from a family that places an invaluable premium on education. To my mom and dad, I would not have made it this far without the nurturing, guidance and value that you imbued in me during my formative years, and particularly to my Dad, whose level of emotional investment in the Ph.D. program rivals mine. And to my siblings – Gbolade, Temilolu, Fikemi, and Dolapo, thanks for the constant support and words of encouragement. You are appreciated. I would also

like to express my heartfelt appreciation to Tori Hill, who encouraged me to apply for the EPP Ph.D. program, provided invaluable advice along the way and has been a constant fixture by my side as I go through the program.

As the saying goes, money makes the world go round. I gratefully acknowledge financial assistance from the National Science Foundation (NSF) CAREER award, the NSF collaborative research grant, the NSF I-Corps program and the Department of Engineering and Public Policy Tuition Scholarship. Access to these financial resources considerably lessened the burden of 70 hour work- weeks, made up of endless hours spent in basements at Carnegie Mellon University and the University of Michigan as well as numerous coffee shops dotted all over Pittsburgh and Ann Arbor.

Abstract

A disproportionate number of Americans drive to work alone – at the individual state level, the figure ranges from a low of 58% in New York to 85% in Alabama. What factors explain this travel behavior and what sort of policy responses are required to effect changes? In responding to this question, we used multiple test beds and made the case for a demand side management approach to remedy parking problems particularly observed in cities central business districts. To this end, we provide an overview of travel behavior and information on policy levers by generating detailed profiles that capture the travel behavior of commuters in the Pacific states of the continental United States. Building off the result that revealed San Francisco as an outlier, we examine the efficacy of modifying parking rates, via elasticity measures, to influence the demand for parking by developing a two-stage panel data regression optimization model for managing parking in the City of San Francisco. A key contribution of the research is deriving these price elasticities of parking demand estimates using panel data methods. Coefficient estimates from the panel data regression are used to fit a linear prediction model that is the primary input to the optimization model. The balance of the thesis focuses on parking information by discussing the design and implementation of ParkPGH, a novel smart parking application that provides real time and predictive information on garage parking availability in downtown Pittsburgh. At its core is a predictive model that uses as input historical parking, weather and event data to provide estimates of available parking spaces. We provide an example of the model implementation using data from the Theater Square garage where we utilize neural network-based predictors and multiple net searches to generate both continuous and binary estimates of parking availability. Provision was made for the binary classifier given the need to reduce the possibility of Type II errors.

CONTENTS

Acknowledgement	i
Abstract.....	iii
Contents	iv
Figures and Tables.....	vi
Chapter 1. Introduction	1.1
1.1 Motivation.....	1.1
1.2 Scope and Research Questions	1.2
1.3 Study Approach	1.4
1.4 Novelty and Practical Relevance	1.6
1.5 Bibliographic Notes	1.7
1.6 Thesis Outline	1.9
Chapter 2. Profiling Commuters' Travel Behavior in the Pacific States of the Continental U.S.	2.10
2.1 Introduction and background context	2.10
2.2 Review of Existing Literature	2.12
2.3 Data Sources and Descriptive Statistics.....	2.15
2.4 Empirical Strategy and Model Specification	2.16
2.4.1 Computing Population Estimates.....	2.16
2.4.2 Econometric Analysis and Model Specification.....	2.17
2.5 Estimation Results and Discussion of Findings.....	2.20
2.5.1 Probability and Parameter Estimation Results.....	2.20
2.5.2 Examining differences in commuting profile based on segmented population	2.22
2.5.3 Spatial variation analysis	2.26
2.6 Conclusion	2.29
Chapter 3. Rethinking SFpark's Performance Based Parking Pricing.....	3.32
3.1 Introduction.....	3.32
3.2 Literature Review	3.34
3.3 Empirical Strategy and Data Sources	3.39
3.4 Optimization model	3.44
3.5 Regression Analysis.....	3.51
3.6 Results and Discussion of Findings	3.56
3.6.1 Coefficients and price elasticity estimates.....	3.56
3.6.2 Figures of merit and assessing SFpark's effectiveness.....	3.59
3.7 Associated Caveats and Policy Implications	3.64
3.8 Conclusion	3.66
Chapter 4. Addressing Parking Challenges in Downtown Pittsburgh	4.68
4.1 Introduction.....	4.68
4.2 Needs Assessment.....	4.69
4.3 Systems Development.....	4.74
4.4 Decision Analytics	4.78

4.4.1	Data Description	4.79
4.4.2	Neural Network.....	4.80
4.4.3	Measures of Predictive Accuracy	4.82
4.4.4	Rationale for using Predictive Classification Methods.....	4.83
4.4.5	Predictive Classification Methods	4.84
4.5	Evaluation: Post Deployment Measures	4.87
4.5.1	Evaluation Approach	4.87
4.5.2	Post Deployment Measures.....	4.89
4.5.3	Program's Estimated Impact.....	4.92
4.6	Conclusion	4.94
Chapter 5. Summary		5.96
5.1	Summary of Findings and Contributions.....	5.96
5.1.1	Travel Mode Preferences	5.96
5.1.2	Parking rate	5.98
5.1.3	Parking information	5.99
5.2	Future Research Direction	5.100
5.2.1	Measurement errors	5.100
5.2.2	Omitted variables	5.101
5.2.3	Methodological challenges	5.103
5.3	Looking Forward	5.103

FIGURES AND TABLES

Figure 2-1: Commuting profile across gender based on travel duration	2.23
Figure 2-2: Condensed commuting profile with or without kids based on travel duration	2.24
Figure 2-3: Commuting profile based on race with respect to travel duration	2.25
Figure 2-4: Solo drivers as a fraction of all vehicle commuters	2.27
Figure 2-5: Solo drivers as a fraction of all vehicle commuters in LA County (left) and as a fraction of all commuters in Northern California (right)	2.28
Figure 3-1: Panoramic representation of the two-stage panel data regression/optimization model.....	3.43
Figure 3-2: Distribution of price elasticities of parking demand	3.58
Figure 3-3: Progress log of the \$.25 price change without the revenue constraint...	3.62
Figure 3-4: Progress log of the \$.25 price change with revenue modelled as a soft constraint.....	3.63
Figure 4-1: Theatre Square Garage Parking Vacancies Mean (left) and Variance...	4.72
Figure 4-2: Smart Parking Application Schema	4.74
Figure 4-3: Pittsburgh's Downtown Map with available parking spaces	4.77
Figure 4-4: ParkPGH iPhone parking prediction output.....	4.78
Figure 4-5: Bar Chart showing relative variable impact on GRNN predictions.....	4.81
Figure 4-6: RMSE values for different sizes of the testing data set	4.83
Figure 4-7: Scatter plot comparing predicted values to actual parking utilization ...	4.84
Figure 4-8: Daily requests by selected delivery channels.....	4.89
Table 2-1: Mode of Transportation to Work.....	2.16
Table 2-2: Travel mode option estimated probability.....	2.21
Table 2-3: Predicted margins for carpooling by Asian commuters	2.22
Table 3-1: Panel Data regression result	3.57
Table 3-2: Comparison across key metrics of performance	3.60
Table 4-1: Stakeholders' Objectives and Measures	4.70
Table 4-2: Baseline Data on Key Program Objectives	4.71
Table 4-3: Variable list for the prediction model and their associated definitions...	4.79
Table 4-4: Summary of the Neural Network Analysis	4.80
Table 4-5: Net Search Result using RMSE measures.....	4.82
Table 4-6: Summary result of the classification based prediction methods	4.86
Table 4-7: Process related measures (%)	4.91
Table 4-8: Outcome related measures.....	4.92
Table 4-9: ParkPGH Estimated Impact.....	4.93

CHAPTER 1. INTRODUCTION

1.1 MOTIVATION

Within the continental United States, the percentage of commuters who drive solo relative to all commuters varies from a low 58% in New York to 85% in Alabama.¹ The disparity in commuting patterns is however, more pronounced within states. In the state of California for example, the percentage of commuters who drive alone to work as a percentage of all commuters across the US Census public use microdata area² (PUMA) ranges from 27% for PUM area code 2203 in San Francisco to 84% for PUM with area code 3302 in Fresno County. Finer segmentation of mode choice reveals negligible variation in the percentage of commuters who carpool although significant differences exist in the percentage of commuters that used the public transit system – a reflection of the differences in the availability of transit services across the geographical areas.

The enormity of this realization begs a host of questions. What factors explain the disproportionate proportion of Americans that drive alone and what sort of policy responses are required to effect a change in travel behavior? What are the parking implications of the observed travel mode behavior and in what way can we collectively, as a society, transition to a more sustainable situation? How critical will these issues be in the coming years given the increased urbanization trend and the linkages between travel mode and other societal objectives such as reducing pollution, congestion and greenhouse gas emissions?

Our research efforts speak to these questions. In addressing them, we use a three part empirical study that examines the effectiveness of demand side interventions in managing

¹ These figures including the subsequent ones for states and cities were all obtained by generating population estimates from the microdata sample set from the US Census using 2014 data.

² A PUMA is a geographically designated enumeration unit by the US Census with a population in excess of 100,000 but below 200,000 residents.

travel behavior and the attendant parking demand such behavior creates. A supply side intervention may not be feasible given that as a result of weak finances, an appreciable number of cities have limped along for decades, delaying much needed investments in critical infrastructure. Further compounding the problem is that higher tiers of government; county, state, federal, with infrastructure responsibilities often pay scant attention to these assets thus forcing financially stretched cities to pick up the slack. Urban infrastructure deficits are further exacerbated in cities' central business districts (CBDs) because of the increased concentration of workers and visitors.

We draw from disparate disciplines and use a myriad of datasets that include parking availability from electronic gate counts, parking meters, , agencies' administrative data, events and weather data for specific localities and survey data from the US Census Public Use Microdata Sample (PUMS) dataset. The primary objective is examining the use of demand side interventions in influencing travel mode choices and parking options that individuals exercise, and examining how the interventions could be implemented in such a way that incentivize individuals to exercise parking options that are socially beneficial. With the demand side intervention as the common thread, the thesis provides a detailed picture of travel behavior and possible policy levers that could be used in influencing travel behavior; examines the effectiveness of parking rate changes in influencing individuals' spatio-temporal parking decisions and documents the conceptualization and implementation of a parking app that provides parking information to patrons in downtown Pittsburgh.

1.2 SCOPE AND RESEARCH QUESTIONS

Following up on the motivating factors for the research, we itemize the research questions below with the thesis chapter associated with each question:

- a. Chapter 2 - what factors explain commuters' travel behaviour and can they be influenced to incentivize commuters to choose more socially optimal commuting pattern?

Rationale: A key determining factor of demand for parking is the travel mode choice made by individuals. The innovation that this task provides is generating representative travel mode profiles with respect to spatial, temporal and demographic characteristics. In addition to the profiles, marginal effects on travel mode choice will also be computed. The rationale for proceeding along this line is twofold – for one, in looking at the gradient of the marginal effect functions, we could make an assessment with regards to how sensitive commuters are to travel mode choice given a marginal difference in the covariate of interest assuming other explanatory variables stay invariant. Secondly, we explore policy levers that could be effective in shaping individuals' travel mode preferences.

- b. Chapter 3 - how effective are parking rates in influencing the spatio-temporal parking decisions individuals make?

Rationale: The justification for this question is two-fold. For one, there is the need to estimate plausible price elasticity of parking demand that could subsequently be used to predict occupancy in a manner superior to the reactionary presently being used by SFpark. Secondly, running multiple scenarios based on these predictions can provide policy guidelines on the most preferred course of action.

- c. Chapter 4 - can a demand-side intervention, specifically providing real-time and predictive information, address the mismatch between parking demand and supply in cities' CBDs?

Rationale: This question was motivated by our observations and the conversations we had with stakeholders who argued that dearth of parking spaces in downtown Pittsburgh was more of a perception problem. In their assessment, if drivers utilized the fringe parking facilities and employees who need parking for extended periods of time chose to park farther at lower cost, it would appreciably address the perceived deficit. Finally, we know that people are inherently creatures of habit, so can their habitual patterns be influenced by giving them information on other possible options? In making an assessment on this question, we will be examining if the data we have could

show that people distributed themselves more effectively across the available parking facilities.

1.3 STUDY APPROACH

The overarching theme of the thesis is determining the effectiveness of demand side interventions in achieving sustainable parking. To this end, we have focused on empirical driven analyses with practical relevance. Each chapter employs an empirical approach ideally tailored to answer the question associated with the chapter. Detailed descriptions of the analyses are provided below:

Chapter 2 or the overview and policy levers component of the research provides a better understanding of commuters' travel mode choices by generating detailed profiles that capture the travel behavior of commuters in the Pacific states of the continental US. These profiles were created using discrete choice models (DCM) with the data coming primarily from the US Census Public Use Microdata Sample (PUMS) data. The microdata sample set allows for the estimation of fine-grained commuting profiles that reveal how individual commuters make travel mode choices. Our results show appreciable spatial variation in travel modal choices and statistically significant differences in modal choices across population segments. One of the revealing findings suggest that significantly enhancing public transit is an order of magnitude more effective at reducing the number of solo-drivers compared to encouraging carpooling. Policy implications are highlighted with regards to the insights provided by the study on making long-term transportation-related infrastructure decisions.

Using a two-stage panel data regression optimization method, Chapter 3 builds on the findings from the second chapter by focusing on the pricing experiment conducted by the City of San Francisco with the aid of a grant from the Department of Transportation (DOT). The policy instrument is changing parking rates via elasticity measures to influence

demand for parking spaces. A key contribution of the research is deriving these price elasticities of parking demand estimates using panel data methods. Coefficient estimates from the panel data regression are used to fit a linear prediction model that is the primary input to the optimization model. The optimization model seeks to minimize parking occupancy deviation from the desired range by modifying parking rates via the price elasticity of demand obtained from the regression model. The model effectiveness is evaluated by comparing actual occupancy/rate tuples with the optimized result to ascertain the potential improvement in the program's performance. Policy scenarios are subsequently explored by carrying out sensitivity analysis primarily through modifying SFpark pricing rules.

Chapter 4 or the information component of the research provides the balance of the research effort. It addresses the design and implementation of a parking application to provide real time and predictive information on parking availability in Pittsburgh's CDB. The task includes conducting a needs assessment to ascertain whether there is a problem and if so, the nature of the problem. The insight obtained from the needs assessment phase is subsequently used in the design and implementation of the parking application. The predictive model uses as input historical parking, weather and event data to provide estimates of available parking spaces. We provide an example of the model implementation using data from Pittsburgh's Theater Square garage where we utilize neural network-based predictors and multiple net searches to generate both continuous and binary estimates of parking availability. Provision was made for the binary classifier given the need to reduce the possibility of Type II errors. Measures of predictive accuracy were also computed for these estimates by comparing the actual outcomes from the testing dataset with the predicted outcomes from the net searches. We summarize by documenting the impact of the application on garage utilization, parking search time and search time variability, the policy implications of the findings and outline suggestions on areas for further study.

1.4 NOVELTY AND PRACTICAL RELEVANCE

Associated with our research is a couple of firsts. This is the first research to derive price elasticity of parking demand using panel data methods. The derived elasticity measures improve significantly on previously derived elasticities that used the same dataset because of the lack of analytical rigor in the manner the measures were computed. In addition, this is the first study to document the decision analytics and the evaluation carried out for a parking predictive app, which incidentally, happened to be the first of its kind. More importantly, compared to documenting the decision analytics and evaluation of the app, an appreciable amount of the research effort was expended on the design and conceptualization of the app.

With regards to operational and practical relevance, the fine granular commuting profiles provide crucial insights on how policy measures could be targeted to population subsets and lays the foundation in providing the information required in estimating demand for transportation services as a result of changing population dynamics. Not only do the research findings provide this, but recommendations are also given on potential policy levers that could prove effective in achieving socially optimal travel behavior. Finally, the visualization using representative individual for each segment of population provides an easily readable chart that often times, shows the stark difference in commuting profile in a more convincing manner.

Other key innovations include addressing the loss of predictive power at the extremes of the predictive continuous parking information using categorical classifiers in such a manner that explicitly address precision, recall and sensitivity measures and that have direct revenue implications to garage owners or parking challenges for patrons. Of relevance is also using the response shift bias theory in artificially increasing the number of pre-deployment respondents with the objective that we have an enough decent sample size before the parking application addressed in Chapter 4 goes live. The predictive

information also holds promise in changing travel mode, considerably extending the reach of the analysis beyond parking issues.

While not new by itself, the context in which the optimization model used in Chapter 3 and the findings obtained make them worthy of being mentioned in this section. The optimizations highlight the importance of moving towards a predictive regime that allows for proactively managing the parking program compared to a purely reactive approach based on the prevailing parking occupancy. Relative to SFpark's figures, our model achieved approximately 10% improvement in systems performance when measured by the cumulative deviation from the desired occupancy target. Of importance are the directional guidelines provided by the policy related scenarios. The scenarios enrich the analysis by providing an advocacy tool and by generating insights on key policy levers that can point the agency managing the program in the right direction with regards to the most effective options available to it in furthering the program's goals. It is also relevant that the veracity of the results is anchored on the primary policy instrument – the spatio temporal price elasticities of parking demand.

1.5 BIBLIOGRAPHIC NOTES

A number of peer reviewed publications are associated with our research. The publications, including working papers, manuscripts submitted with review pending are itemized below:

Peer Reviewed Journal Articles

1. **Fabusuyi T.**, Hampshire R., Hill V. & Sasanuma K. (2014) "Decision Analytics for Parking Availability in Downtown Pittsburgh." *Interfaces* Volume 44, #3
2. **Fabusuyi T.**, Hampshire R., & Hill V. "Evaluation of a Smart Parking System", *Transportation Research Records: Journal of the Transportation Research Board of the National Academies*, # 2359 – Car Sharing, Demand Management and Parking. Washington D.C. 2013

Peer reviewed conference proceedings

3. **Fabusuyi T.**, and Hampshire R. “Addressing Parking Challenges in Downtown Pittsburgh.” Proceedings of the 14th Computers in Urban Planning and Urban Management (CUPUM) Conference, Cambridge, MA, 2015
4. Hampshire, R.C., **Fabusuyi T.**, Hill V. and Sasanuma K., “Predictive Modeling for Smart Parking: Case of ParkPGH,” Proceedings of the 18th World Congress of the *Intelligent Transportation Systems*, Orlando, FL, 2012

Submitted - review pending

5. **Fabusuyi T.**, and Hampshire R. “The Mode Most Traveled: Parking Implications and Policy Responses.” Book chapter for the 15th Computers in Urban Planning and Urban Management (CUPUM) Conference, Adelaide, Australia, 2017 (Extended abstract accepted)

To be submitted

6. **Fabusuyi T.**, and Hampshire R. “Rethinking SFpark’s Performance Based Parking Pricing.” To be submitted to *Transportation Research Part A: Policy and Practice* (Fall 2016)
7. **Fabusuyi T.**, Hampshire R. and Qian S. “Profiling Commuters’ Travel Behavior in the Pacific States of the Continental U.S.” To be submitted to *International Journal of Sustainable Transportation* (Fall, 2016)

In addition to the list above, the research endeavor has also seeded a number of synergistic products. These includes presentations at peer reviewed conferences and annual meetings, awards including the National Science Foundation (NSF) I-Corps grant, a couple of media mention, an article in a professional magazine and a technical report.

1.6 THESIS OUTLINE

The thesis is structured into five chapters. Chapter 1 provides the motivation for the current research, itemizes the research questions addressed in the study and documents the novelty of the research and its relevance to real life issues. Overview and policy levers are addressed in Chapter 2. The chapter creates detailed commuting profile for small segments of the population and sheds insight on the spatial variation in travel behavior across the Pacific states of the Continental US. The effectiveness of modifying parking rates to influence the spatio-temporal demand for parking spaces is examined in Chapter 3. Recommendations on preferred course of actions are also provided using simulated policy scenarios. Information as a demand side intervention is examined in Chapter 4. The chapter documents the steps required to conceptualize, design, implement and evaluate ParkPGH, a real time and predictive parking app for the City of Pittsburgh. Chapter 5, the concluding chapter, summarizes, identifies areas for further studies and offer some insight on what the future landscape of parking may look like.

CHAPTER 2. PROFILING COMMUTERS' TRAVEL BEHAVIOR IN THE PACIFIC STATES OF THE CONTINENTAL U.S.

2.1 INTRODUCTION AND BACKGROUND CONTEXT

The factors that explain commuters' travel mode have been the subject of numerous studies starting from the seminal work of Daniel McFadden on urban travel demand in 1974 (1). Of late, there has been a resurgence in this line of research with much of the attention being devoted to understanding why a predominant percentage of commuters drive alone (2). Key drivers of this interest include the linkage between travel mode and greenhouse gas emissions (3); the smart cities concept and the shifting demographic towards cities (4) and what this trend portends specifically with regards to travel mode choices and the attendant demand for parking spaces it creates. The increased demand is acutely felt in cities' central business districts (CBDs) that are becoming attractive not only for work, cultural events and sporting activities but also as the residential area of choice (5).

Modeling travel mode choice has historically been carried out as a function of the measurable attributes of the commuter and the attributes of the travel mode alternative (6). An appreciable degree of modification has since been made to this approach with the impetus coming largely from the travel behavior research community. The rationale behind the rethinking is the realization that travel mode choice are influenced by lifestyles, preferences and an individual's values (7). These often manifest as unobserved heterogeneity of preferences across individuals and a key concern with the econometric approaches is how to address the attendant assumption violations they create.

Our approach to examining what factors motivate an individual to drive to work alone builds on the aforementioned approaches, albeit with a slightly different twist. Using the US Census Public Use Microdata Sample (PUMS) data from the pacific region of the US, specifically the states of California, Oregon and Washington, we generate travel mode profiles for representative individuals for each geographical area of interest. The public use

micro-data sample set provides an excellent resource for understanding commuting patterns for segments of the population at sub-regional or city level, a feat that was not possible in the past given the relatively small sample of existing surveys. Apart from determining the variations in travel behavior across demography and geography, a key objective of this research is examining how sensitive different segments of the commuting population are to travel mode choices given changes in the covariates of interest.

Findings from the study shows that achieving a travel mode shift – from the representative commuter across the three states to that of a San Francisco resident could reduce the number of vehicles plying the road on any working day by more than 10 million. We do not find appreciable variation in the preferences exhibited by commuters to carpool. Consequently, the mode shift could only come from municipalities, counties and city governments investing in effective public transit systems and/or implementing land use strategies that seek to achieve job-housing balance.

In addition to the spatial variation observed, the innovative approach used in segmenting the population reveal key insights on the need for targeted policy measures for subsets of the population. Even within a population cohort, we observed appreciable variation in the travel mode choices they exhibit. Newly arrived immigrants in the workforce, particularly those with lower education, are twice as likely to use the public transit system compared to the mainstream population however, the probability of carpooling is not significantly different compared to the native commuting population. Targeted policies such as transit vouchers with sunset provisions could prove effective for such a cohort.

The rest of the chapter is organized as follows. The next section provides an exhaustive review of the existing literature. Section 2.3 identifies the sources and provides descriptive statistics of the data employed for the study. The empirical strategy and model specification are addressed in the fourth section while section 2.5 presents the results from the study and discusses the findings. The concluding section recaps and provides insight on the caveats associated with the study.

2.2 REVIEW OF EXISTING LITERATURE

The theoretical basis for most travel choice models is the random utility model (8) and the classic approach examines how changes in socio-economic and demographic composition impact commuting patterns. Empirical work on subsets of the population include those on gender differentials in commuting pattern (9), (10), (11); travel mode choices of immigrants (12), (13), (14); and the works of Guiliano (15), Blumenberg and Smart (16) that examine differences across racial and ethnic identities. Some of these analyses have been directed to addressing the changing demographics, specifically with respect to immigration and the aging population and how these trends will impact aggregate travel mode choices.

An appreciable volume of literature within the body of work on commuters' travel mode choice revolves primarily around driving unaccompanied. Ferguson (17) identified three traditionally economic factors responsible for this: more cars per capita due to an increase in average wealth; a decrease in gas prices; and an increase in the education level of the average worker. Bhat and Sen (18), in a model of travel mode choice, found that income, number of vehicles per worker, work duration and employment density are all significant variables in the decision to drive solo. Workers with high incomes, higher ratios of household vehicles to workers and longer duration at work are all positively related to driving solo. Barbour (19) in a study on the state of California found that women, African Americans and high income groups have longer commutes, and that factors influencing travel mode choices include vehicle ownership, parking cost and parking availability.

Research efforts in this area have been broadened to include the impact of neighborhood characteristics such as residential sorting on travel mode choice. Some of these studies, such as Thakuriah et al (20) and Waddell et al (21) incorporate land use policies and job location decisions. Other studies with residential sorting focus include a study by Cervano and Duncan (22) that found that after residential sorting is taken into account, the impact of neighborhood attributes diminishes significantly. In another study accommodating for

residential sorting, Zhang (23) used an instrumental variable approach in a model that included residential location, auto ownership and travel mode choice. He found that dependency on automobiles was highly sensitive to street network connectivity and automobile availability.

A number of the empirical research on travel mode options have been carried out using PUMS data or by utilizing PUMS data in combination with other datasets. The dataset has particularly found much relevance in conducting comparative assessment of travel mode choices for population segments or in examining spatial variation across geographical areas given that the dataset has national coverage and provides an even platform with regards to data collection. Examples of studies utilizing PUMS data to analyze travel behavior include the 2000 PUMS data that was employed by Blumenberg and Shiki (24) to examine the travel mode choices of the foreign-born population of California. The study also looked at the relationship between transit usage rates and the tenure of U.S. residency, while controlling for other factors that might influence transit choices. PUMS data has also been used to show differences in commuting pattern across gender (25), (11); in the analysis of access to job opportunities (26); and for land use policies (27). Further credence to the use of PUMS data in examining travel behavior is also provided by the works of Thakuriah et al (20) on job access and reverse commute program; activity based demand model by Donnelly et al (28) and Deal et al (29) that examine the impact of growth management policies on travel mode choice.

Finally, there are significant differences in the types of models used to estimate these effects though, of late, there has been a movement towards the use of more robust models. Huang et al (30), for example, posit that preferences for autonomy, including self-reliance, independence and privacy, in addition to economic preferences are key in determining travel mode choice. Added to this, Charles and Kline (31) argue for the need to consider social capital as a key factor in determining carpooling. They make the case that while traditional economic factors come into play, social identification and social capital also

matter - a view point shared by Bluemenberg & Smart (32) that looks at the carpooling behavior of the immigrant population in Southern California.

In light of the aforementioned, there has been considerable refinements to the methodology with regards to unmeasurable attributes - preferences, values and attitudes of commuters and the attendant heterogeneity, self-selection and endogeneity issues they create. Econometric approaches employed that go over and above the measurable variables such as commuters' socio-economic characteristics and attributes of the travel alternatives include hybrid choice models (33) that pushes for more behavioral realism by incorporating both latent attitude and variables that capture perception in the modeling process. Other econometric models employed in explaining how lifestyle preferences and social interaction influence travel behavior include the structural economic models (34) and the latent choice class model (35). Latent variables could also be introduced through the use of an Integrated Choice and Latent Variable (ICLV) model that employed a multinomial discrete choice model and a latent variable model (36). This approach takes into consideration the abstract constructs that influence commuters' travel mode choices.

While these approaches have done better in integrating unobservable explanatory variables, it has been observed that they are not ideal for long term policy formulation (37), (38). Recognizing these limitations and driven by the need to reflect the unique attributes of our dataset, we model commuters' travel mode choice using the multinomial probit (MNP) approach. The model is robust enough as to be relevant with regards to the assumptions made particularly with regards to the transportation field (39). Our approach improves on existing studies by creating rich commuting profiles for each population cohort, thanks to the microscopic census profiles. This is done by examining multiple dimensions of the variation using distinct segments of the population sampled with the objective of ascertaining the responsiveness of each of the subset to changes in the explanatory variables. We subsequently generate marginal effects of the said explanatory variables not only at the mean but by graphing the probabilities of an individual selecting specific travel mode based on the control variable of interest. This presents a more visual and readable

option and shows the differences among the subsets of the sampled population by stacking their travel mode profiles on the same graph. Local means for these population segments are used in the regression analysis.

2.3 DATA SOURCES AND DESCRIPTIVE STATISTICS

The study utilized data obtained from the US Census PUMS dataset. The use of the PUMS dataset presents several advantages. For one, the utilization of individual and household records that could be linked through an identifier provides researchers with a great deal of flexibility in manipulating the dataset. Secondly, the large sample size affords us not just higher statistical power but also validity with regards to generalizing our findings to the population of interest and enables more granular analysis of different segments of the commuting population. The study covers the pacific states of Washington, Oregon and California – an area representing a combined population of 49.3 million as of July 2013 (40). The geographical area provides the ideal environment to examine the impact of policy measures given that subsets of this area have been test bed for innovative programs such as the SFPark project and LAExpress Park and has historically been at the fore-front of the debate on mode conscious commuting. The 2014 dataset, the most recent year used in the study, has approximately half a million observations from which we generated a total population estimate of 49.3 million for the three states. Of this figure, 28.2 million are in the workforce with 21.6m coming from California, 4.2 million from Washington and the balance of 2.4 million is from Oregon. A detailed breakdown of commuting pattern including vehicle occupancy is shown in the Table 2-1.

Table 2-1: Mode of Transportation to Work

Means of Transportation	California (16.86 million)	Oregon (1.744 million)	Washington (3.22 million)
<i>Car, Truck or Van</i>	84	82	83
<i>Drive solo</i>	73	71.2	71.8
<i>2-person carpool</i>	8.3	8.5	8.4
<i>2+ person carpool</i>	2.7	2.3	2.8
<i>Transit</i>	5.4	4.2	6.3
<i>Non-motorized</i>	10.7	13.8	10.7

2.4 EMPIRICAL STRATEGY AND MODEL SPECIFICATION

The empirical strategy employs a dual pronged approach – one examines differences in commuters’ travel mode choice across population cohorts while the second explores locational variation in commuting options. The strategy utilized a multi-level approach – analysis at the level of individual respondents using the PUMS dataset and a higher level analysis that address the spatial dimension of the data. The microdata sample set was used to generate the population estimates and the associated margins of error. The dataset was also used for the econometric analysis that provides insights to commuters’ travel mode choice. Finally, the study included a visualization component using the PUMS area as enumeration units. Specifically, the empirical strategy entailed the following:

- Generating the population estimates of travel mode choices from the sample dataset using balanced repeated replication (BRR) approach to examine the variation in travel mode choice across the PUM areas and;
- Determining the impact of potential covariates on travel model choice and inferring differences in travel mode choice across space and population segments using multinomial probit (MNP) models

2.4.1 Computing Population Estimates

Using the US Census Public Use Micro-data Sample dataset (41), we generated population estimates of commuters’ transportation modes. In calculating the estimates, we employ two

types of weights: person weight and replicate weights. The person weight is required for the point estimates and both person weight and the replicate weights are needed to calculate the standard errors.

Given that the analysis is conducted on small population segmentation, Fay's variant of the Balanced Repeated Replication (BRR) method was utilized in calculating the standard errors (42). Fay's approach, called the Modified Half Sample (MHS) improves on the BRR by addressing the problem of perturbed weights and decreased sample size using an adjustment factor called the Fay coefficient. This coefficient was set to 0.5 for the PUMS data. Building on the sampling variance:

$$\hat{V} = \frac{\sum_{i=1}^N (X_i - X)^2}{N}, \text{ Fay's MHS variance equals:}$$

$$\hat{V}_{mhs} = \frac{1}{(1-m)^2} \hat{V} = \frac{1}{(1-m)^2} \frac{\sum_{r=1}^N (X_r - X)^2}{N}$$

where m is the Fay's coefficient, X_r the replicate estimate and X , the full sample estimate. Since the PUMS person records file has $N = 80$ replicates and $m = 0.5$, the expression above reduces to:

$$\hat{V}_{mhs} = \frac{\sum_{r=1}^{80} (X_r - X)^2}{20}$$

2.4.2 Econometric Analysis and Model Specification

In determining the contributory factors that explain travel mode choice by commuters, we borrow from the typical discrete choice models that are based on random utility theory using the expression below:

$$V_{ij} = \beta X_i + \epsilon_{ij}$$

V_{ij} represents the indirect utility function, and X_i represents a vector of the individual's characteristics. The individual will choose the travel mode for which V_{ij} is the highest. Therefore, the probability that an individual i chooses mode j will be:

$$P_{ij} = Pr(V_j > V_k) \forall j \neq k$$

$$= Pr(\beta_j X_i - \beta_k X_i) > \epsilon_{ik} - \epsilon_{ij}$$

The expression above constitutes the basic premise of DCM. The pioneering work on discrete choice models was McFadden (1) seminal article. It is assumed that rational individuals are exercising an option out of a set of alternatives, each with a series of attributes from which the individuals derives or enjoys some utility. Implicit in this assumption is that the utility is derived from a deterministic function and measurable. However, there are uncertainties and to reflect this reality, probabilities are used to describe travel mode preferences. In calculating these probabilities, there is an observable portion that could be characterized by the attributes of the travel mode; the commuters' characteristics and the interactions between the modes' attributes and commuters' characteristics. The unobservable portion that capture the uncertainties is subsumed underneath the error term which is assumed to have some well-behaved randomness that is assumed independently and identically distributed with a Gumbel distribution. Given this assumption, $(\epsilon_{ik} - \epsilon_{ij})$ will have a logistic distribution.

If;

$$V_{ij} = \begin{cases} 1 & \text{for } V_{ik} \leq V_{ij} \\ 0 & \text{otherwise} \end{cases}$$

and given the assumptions of the error term, the parameters of the utility model could be estimated using a multinomial logistic model as shown below:

$$\Pr(V_{ij}|j) = \frac{\exp(\beta_j X_i)}{\sum_{j=0}^{J-1} \exp(\beta_j X_i)}$$

From the expression above, an increase of the same magnitude in the utility of each alternative should have no impact on the choice probability equation. Since the probabilities have the same denominator, the ratio of any probability pair j, k is:

$$\frac{\Pr(j)}{\Pr(k)} = \frac{\exp(V_j)}{\exp(V_k)} = \exp(v_j - v_k)$$

It is obvious that the ratio of the probabilities for any pair of alternative remain invariant to changes in the attributes of other alternatives within the choice set that do not feature in the ratio. This is known as the Independence of Irrelevant Alternatives (IIA) assumption, a supposition that seems counter-intuitive particularly for travel mode choices. For example, a policy measure that removes a designated HOV lane is expected to shift more commuters to driving alone, consequently changing the ratios of driving alone relative to (say) public transit or non-motorized or bipedal travel mode. Secondly, the assumption made with regards to the error term precludes the heterogeneity in taste preference with regards to travel choice across commuters.

The empirical strategy takes into consideration these two shortcomings by using multinomial probit (MNP) model in estimating the parameters of the DCM. The MNP approach makes provision for such random variation provided the taste parameter vector is multivariate normal and the utility function is linear in the parameters. With regards to the IIA assumption, a feasible option is to relax the assumption of the error terms being independently distributed from one another. This modification means that the off-diagonal elements for the error terms are not constrained to take only zero values and thus, are correlated and this is precisely what the MNP model achieves at the cost of more demanding computation resources.

2.5 ESTIMATION RESULTS AND DISCUSSION OF FINDINGS

The analysis is based on the population records for the 2009-13 5 year US Census PUMS dataset. Variables employed for the analysis include commuter related variables - Age, Economic wellbeing, Sex, Race, Education attainment, Geographical Location, Pre-teen kids; trip and mode related variables - travel duration, time of travel, travel mode choice and interaction (including squared) terms. The economic wellbeing, measured relative to the federal poverty level of each state analyzed, was dropped given that it is highly correlated with education attainment. The covariates also include extra regressors created by interacting two (or more) variables to capture differential slope effects. The observed travel mode choice, which is the dependent variable, is reclassified into four broad categories: Non-motorized and bi-pedal options; Public Transit; Car-Pool; and Drive Alone. The estimation results are classified into two broad sections. One section presents differences in travel mode choices across socio-economic and demographic profiles while the second addresses spatial variation in commuting pattern across selected metropolitan statistical areas.

2.5.1 Probability and Parameter Estimation Results

Table 2-2 shows the adjusted predicted probability of exercising specific travel mode options for selected population cohorts identified using dummy variables while keeping other covariates at the mean. These estimates are calculated using Stata's margins command that allows covariates of interest to vary while other explanatory variables are held at fixed values, typically the group or subgroup's average.

Table 2-2: Travel mode option estimated probability

Covariates	Non-Motorized	Public Transit	Car-Pooling	Drive Alone
Female	0.045	0.064	0.124	0.767
Male	0.060	0.048	0.119	0.773
Without Kids	0.053	0.055	0.120	0.771
With Kids	0.042	0.041	0.141	0.776
Non-Californian	0.055	0.072	0.127	0.746
Californian	0.052	0.051	0.119	0.778
Non-Asian	0.050	0.053	0.125	0.772
Asian	0.094	0.062	0.102	0.742

The probability across the column for each row, representing a population segment, sums up to 1. Irrespective of the population segment, approximately three out of every four commuters drive unaccompanied with females using the public transit and carpooling options more compared to male. Across the three states surveyed, Asians drive less and carpool less compared to other races and compared to other states, Californians drive more unaccompanied by a margin of 4%. All the coefficients are significant at the 99% confidence interval for each parameter estimate provided.

Given that the population is dichotomized, the adjusted prediction for any two groups – say men and women, is simply the difference of their marginal effect. Our analysis revealed that a White commuter’s probability of car-pooling to work is 12.6% compared to 9.9% for the average Black. The way to interpret this is to say that assuming we have two otherwise identical commuters – one White and the other, Black, the White commuter’s probability of car-pooling to work will be 2.7% points higher compared to the average Black commuter, an outcome that may be due to more dense networks. Table 2-3 improves on this by providing and contrasting the computed adjusted predictions for car-pooling at representative values with respect to education attainment for individuals who identified as Asians. The adjusted probability prediction to carpool for Asians is 10.2% from Table 2-2. Table 2-3 shows that this probability varies significantly depending on the respondent’s education level – ranging from 15% for 3 years of formal education to only 2.2% for

individuals with 24 years of formal education. Finer segmentation is possible by creating smaller cohorts by using interaction terms for the dummy variables.

Table 2-3: Predicted margins for carpooling by Asian commuters

Years of Education	Predicted Margin	Standard Error	z-Value	P> z	[95% Conf. Interval]	
3	0.150	0.018	8.290	0.000	0.115	0.185
6	0.134	0.015	8.910	0.000	0.104	0.163
9	0.111	0.013	8.710	0.000	0.086	0.136
12	0.087	0.011	7.850	0.000	0.065	0.109
15	0.065	0.010	6.750	0.000	0.046	0.083
18	0.046	0.008	5.720	0.000	0.030	0.062
21	0.032	0.007	4.870	0.000	0.019	0.045
24	0.022	0.005	4.200	0.000	0.012	0.033

2.5.2 Examining differences in commuting profile based on segmented population

Here, we show findings on commuting profiles carried out based on distinct population segments. The commuting profile analysis builds on MNP regression analysis and the average predictions and marginal effect calculations provided earlier. Across gender, and specifically, with regards to education, our analysis revealed that women have a higher probability of using the public transit system compared to men with lower rate of probability beyond 12 years of education. Similar education threshold is also observed for driving alone although higher education levels compared to the threshold is associated with higher probability of women driving alone relative to men. Travel mode profiles across gender with respect to travel time presents a different picture as shown in Figure 2-1. Irrespective of gender, there is a pronounced preference for using the public transit system, relative to driving alone, once the commute time is in excess of an hour. However, it may be the case that women appreciate flexibility more, given that the probability of carpooling to work is much less utilized by women. This may be the case particularly for women from households with young children as shown in Figure 2-2.

Figure 2-1: Commuting profile across gender based on travel duration

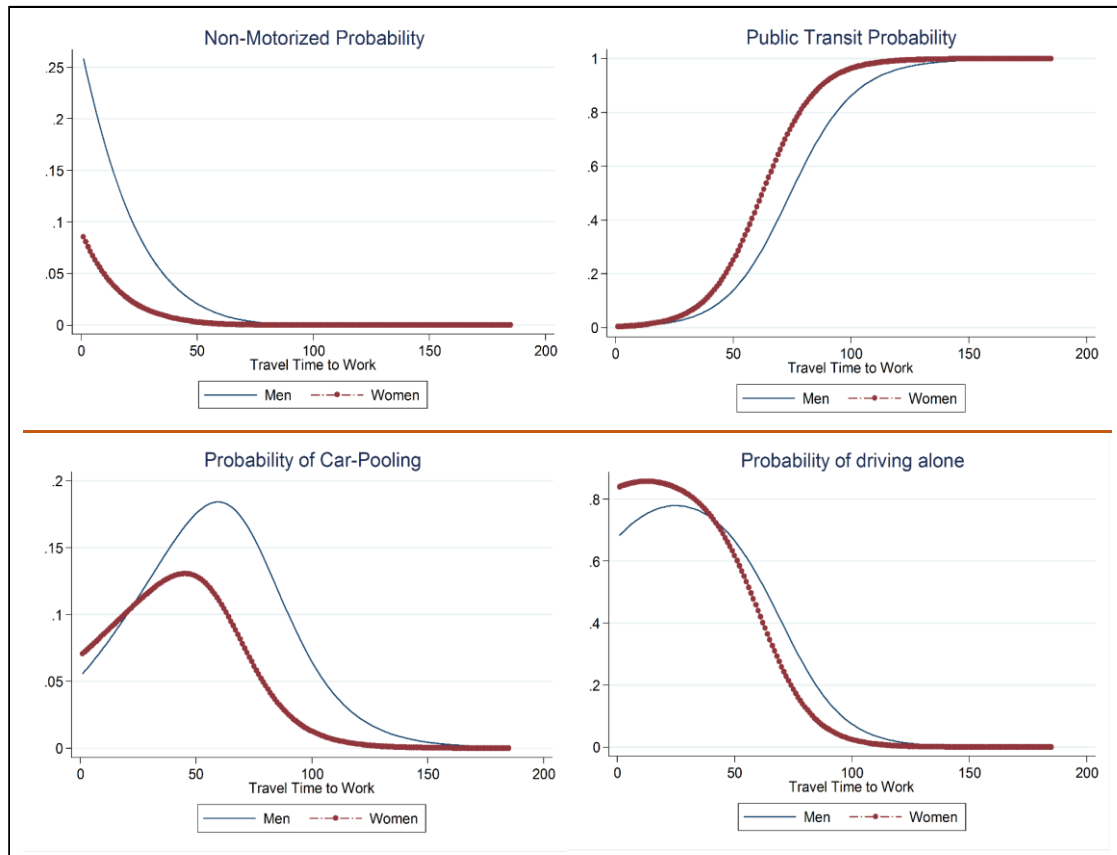
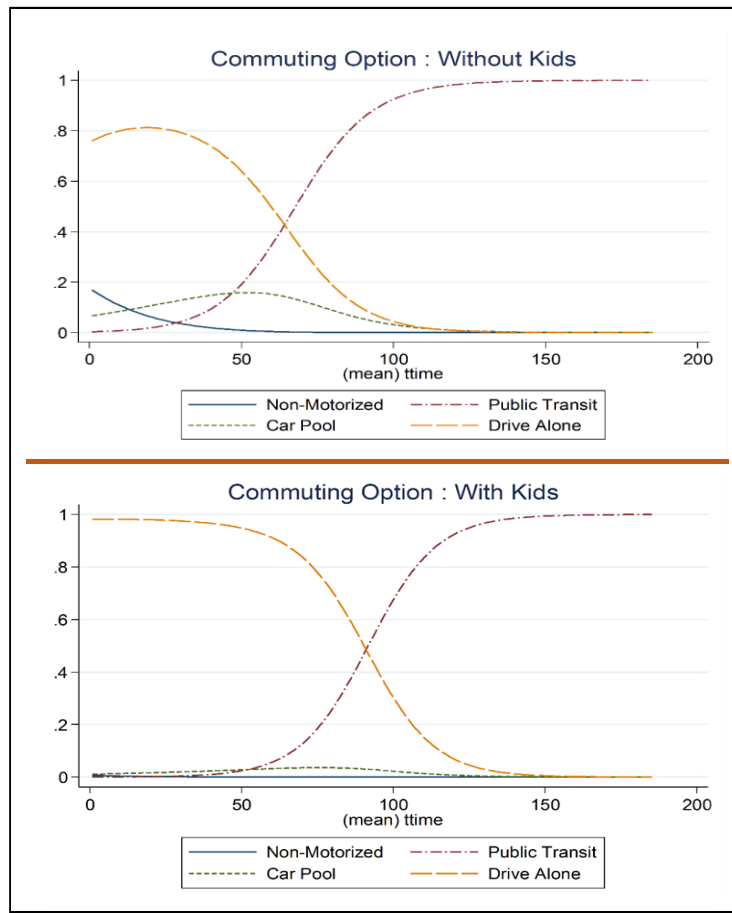


Figure 2-2 presents a condensed profile based on travel time. –The non-motorized and the carpooling travel mode options are rarely used and barely registered a blip particularly for commuters with children. Eyeballing the charts shows that except for the probability of driving alone, the probability of any mode choice for individuals with kids trail those without kids at every level of travel duration. Compared to individuals without kids, the carpooling and non-motorized mode options are appreciably muted. The same pattern (not shown) is also observed when the profiles are generated with respect to education attainment. The distinct mirror image for the computed probabilities observed between public transit travel mode and driving alone was also visibly obvious when the commuting profiles are generated with respect to years of schooling though this may just be picking up the size effect.

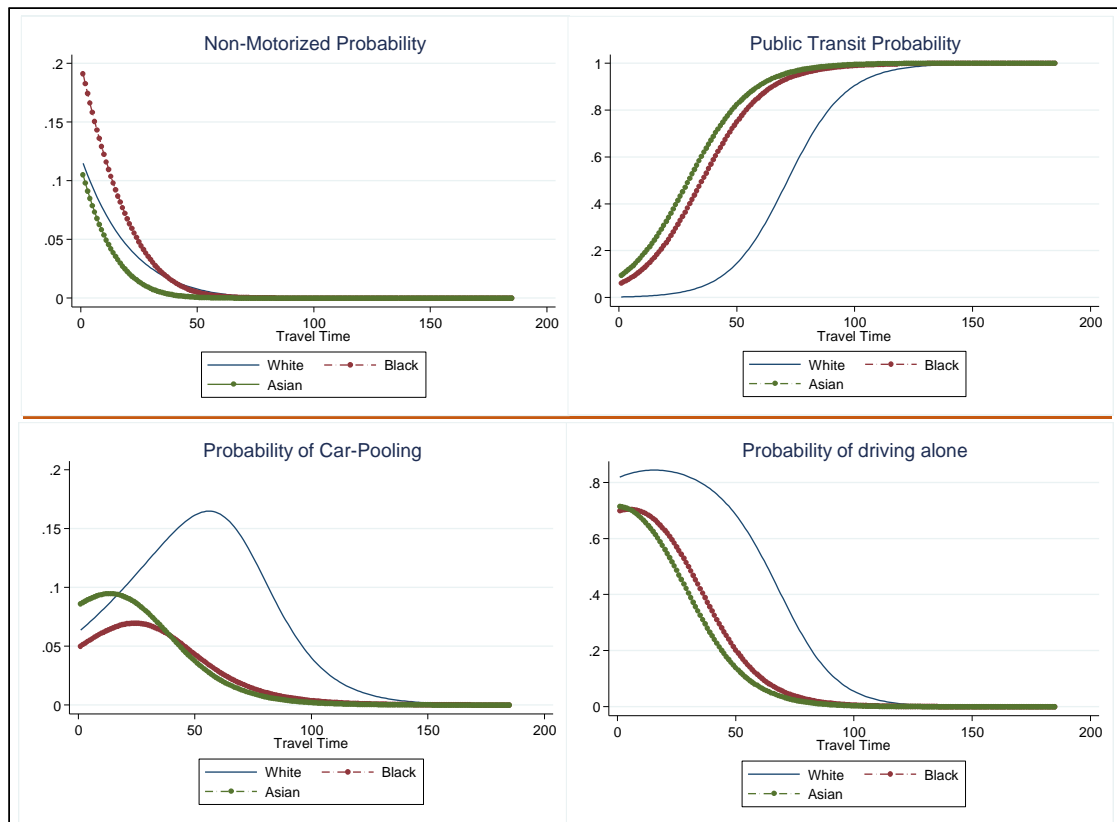
Figure 2-2: Condensed commuting profile with or without kids based on travel duration



Commuting profile based on race however shows a noticeable difference, a finding that corroborates earlier empirical work in this area. An earlier referenced article (2) showed that the differences in commuting behavior are most pronounced based on analysis associated with the characteristics of different ethnic groups. For example, the article revealed that Non-Hispanic White workers drive to work solo at a rate almost 10% points higher compared to other racial groups. Figure 2-3, with the Y-axis not drawn to scale across the travel mode options, sheds more light on this. Our analysis reveals that by far, White respondents utilized carpooling or driving alone to work modes the most, whether the probabilities were computed based on education attainment or travel duration. A couple of observations is particularly worth mentioning. For one, the public transit probability profile shows that once an average White has the equivalent of a high school

diploma, the probability of her taking the public transit travel mode becomes almost negligible. Secondly, with respect to travel duration, the preference for utilizing the public transit systems kicks in when the commute duration is an hour or more for Whites while other ethnic groups exhibit preference for the public transit system much sooner. Part of this observation may come purely from learned behavior. For example, white kids who grew up in the suburbs who are not used to being bussed to school may opt out of using the public transit system once they enter the workforce.

Figure 2-3: Commuting profile based on race with respect to travel duration



The balance of this subsection examines the travel mode choices of natives compared to non-natives. A native is defined simply as an individual who was born in the US and a non-native as one who is foreign born. The PUMS questionnaire merely asked “where was this person born?” and the response is coded “native” if the individual is born in the US and “foreign born” otherwise. By and large, there is a consensus from existing empirical studies

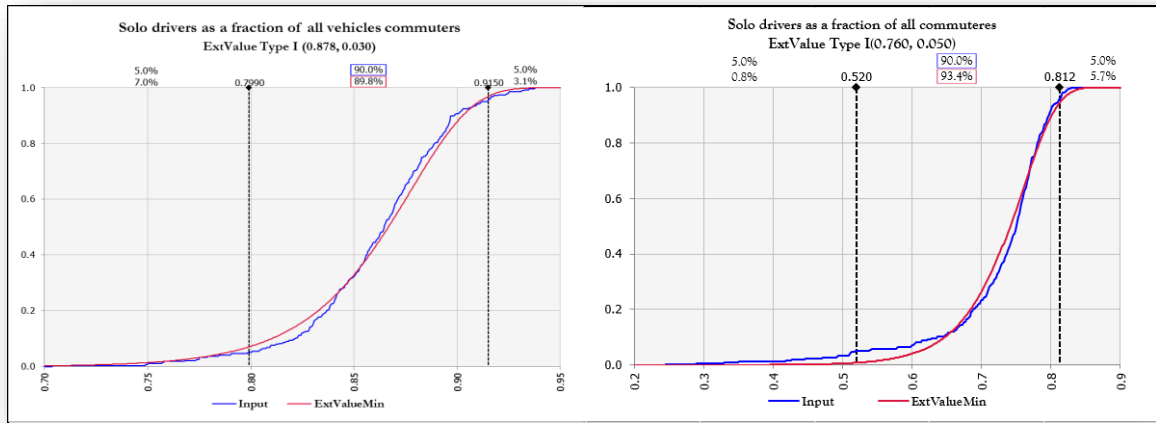
that immigrants use the public transit more compared to natives (24), (43) though the commuting profile, irrespective of the mode choice, of immigrants converges to that of the native born population with extended tenure (13), (44).

Of the individuals in the workforce that provide information on their commute, 30.3% self-report to be foreign born. At low levels of education, immigrants use the public transit almost twice as compared to natives while no discernable difference is observed in commuting profile across the two population segments for individuals with at least four years of college education. However, our findings, with regards to carpooling contrasts with earlier empirical research that found that immigrants have a higher probability of carpooling compared to natives (31), (32). A plausible explanation, related to the network size and its effect on mode choice, is that the analyses for the referenced studies were conducted irrespective of the trip purpose while the present study utilized only trip to work data. This result is particularly noteworthy as immigrants are expected to drive the population growth in the nearest future (45).

2.5.3 Spatial variation analysis

The spatial variation section utilized BRR method in determining differences in commuting pattern across geographical areas. The PUM area constitute the unit of analysis for the BRR approach with a total number of 306 PUMS units analyzed. The percentages of drivers who drive unaccompanied relative to all vehicle commuters and all commuters are shown in Figures 2-4a and 4b respectively. The charts were obtained by fitting point estimates generated for each PUMS area to a distribution function. The best fit, chosen based on information criteria measures, was the extreme value distribution. The charts provide the feasible space or possible realization of individuals who drive solo relative to the total commuting population or all private vehicle users. The objective is to provide insights or evidence of clustering or localized commuting pattern. In our analysis, of the 306 PUMS across the three states analyzed, only four PUMS areas have less than 30% Solo drivers relative to all commuters ratio. All these PUMS are in the San Francisco area.

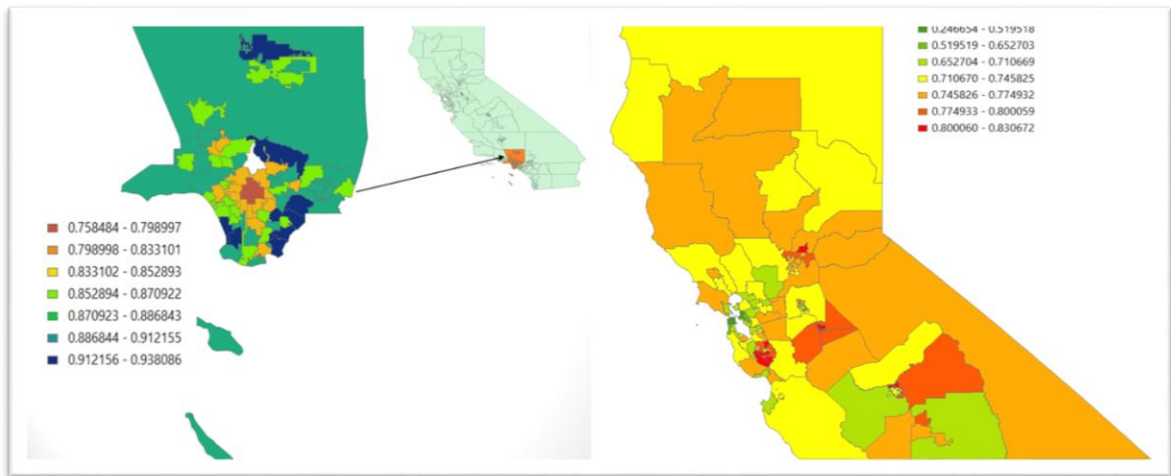
Figure 2-4: Solo drivers as a fraction of all vehicle commuters



A disaggregated analysis at the state level using both box plots and kernel density charts revealed the presence of outliers for California. This was observed for any of the ratios used in Figure 2-4 – fraction of individuals who drive unaccompanied relative to all commuters or relative to the population of all private vehicle users. What is particularly telling is that the best performing PUMS area or the worst are concentrated not just in particular MSAs but in specific counties. Of the five PUMS area with the least ratio of commuters who drive solo relative to the total commuting population, the first four are from San Francisco County and the fifth is from Alameda County.

Complementing the charts in Figure 2-4 are thematic maps of Los Angeles County and Northern California showing the percentage of commuters that travel to work unaccompanied relative to either all vehicle commuters or all commuters. Merely eyeballing the maps in Figure 2-5 reveals that more densely populated and urban areas have lower percentages compared to the suburbs. However, there is appreciably more variability even across densely populated PUMS area, revealing that population density does not explain all that was observed, a finding that buttresses an earlier result that shows that of the three states analyzed, LA County has four of the five worst performing PUMS area when measured by the ratio of commuters driving alone as a fraction of all private vehicle users.

Figure 2-5: Solo drivers as a fraction of all vehicle commuters in LA County (left) and as a fraction of all commuters in Northern California (right)



A more granular analysis looking solely at the seven PUMS area that make up the City of San Francisco shows that only 35% of the workforce drive to work unaccompanied compared to 74% for the three states analyzed. Even more revealing is the fact that the average San Francisco resident is more than seven times (34.4% to 4.8%) likely to use the public transit system compared to a representative commuter across the three states. In addition, a higher percentage of San Franciscans (23.7%) either work close to home or use some form of non-motorized or bi-pedal travel option compared to 10.7% for the balance of the commuting population. Very little variation in carpooling is observed between the two cohorts with regards to carpooling probability.

If we were to assume a shift in travel mode from the average commuter across the three states analyzed to the representative commuter who is resident in San Francisco, we estimate, for the three states analyzed, a reduction in excess of 10 million vehicles put on the road on any working day, each of which will require a parking spot when it reaches its destination. The analysis was carried out using three forms of vehicle occupancy – individuals who drive alone; 2 person carpools and 3 or more person carpools and subsequently applying the proportionate breakdown of the forms of transportation and

vehicle occupancy to the relevant cohort – the representative San Francisco resident commuter and the average commuter for the balance of the population. The equivalent number of vehicles was obtained by dividing the number of vehicles by the occupancy number while taking a conservative view by assuming that maximum occupancy for any vehicle is three.

The analysis shows that commuters are quite resistant to carpooling given the minimal spatial variation observed in the percentage of commuters who carpool. However they may be amenable to using a functional public transit system. As a policy tool, it will be more effective getting commuters who drive to work to shift to public transit than it is to get them to carpool. Presumably this has a lot to do with the decrease in flexibility that accompanies carpooling. For an individual who carpools, her home and home and work departure times are necessarily fixed, a constraint that is appreciably relaxed with the public transit given that she has multiple time-based options for leaving for work and home.

2.6 CONCLUSION

The findings from the study have yielded insights on the key determinants of the travel mode choices individual make; the impact on the mode choice assuming these factors are varied and shown how commuting behavior varies either across distinct population segments or across geographical areas. An advantage of the present study is the use of microdata set that allowed for the estimation of a disaggregated model which showcased how individual commuters, endowed with some characteristics, make travel mode choices. The theoretical basis utilized naturally lends itself to this approach in that each commuter is assumed to exercise a choice that maximizes his utility. The empirical strategy implements this by computing probabilities for each mode choice using an approach that maximizes the likelihood that the predicted travel option is actually the mode utilized in real life by the commuter.

The results obtained from the study show that statistically significant differences on travel mode options were observed across distinct population segments. Same observations apply within a specific cohort assuming individuals have different endowments with regards to the explanatory variables. Not only were these differences significant, the magnitude of the variation was substantial. Among Asians for example, there is almost a ten-fold difference in the probability of carpooling between the most and the least educated commuters. Appreciable spatial variations in commuter patterns were also observed either based on states or MSAs. PUMS areas from the state of California occupy both ends of the distribution spectrum when looking at both the fraction of individuals who drive unaccompanied relative to the total commuting population or those who use personal vehicles. The City of San Francisco provided the most socially optimal travel behavior. Within the state of California, the total number of vehicles plying the road will be reduced by up to 8.6 million on any working day assuming the commuting pattern observed in San Francisco applies to the rest of the state.

While these findings are illuminating, our approach is not without limitations. The identified shortcomings are either related to the assumptions made or the data challenges encountered while carrying out the research. The travel choices, for example, are assumed made with perfect information, which is often not the case in reality. New movers may not have the requisite information on the available commuting options and tenured commuters may exhibit their choices in a myopic manner – a plausible trait given that most of us are creatures of habit. Consequently, the heterogeneity we may be observing may have less to do with taste variation but driven more by asymmetric information.

Apart from the aforementioned, there are data issues related specifically to the paucity of supply-side transportation data or alternative specific data. An example is the lack of data on the quality of the public transit service - frequency, routes covered, cost, timeliness, etc. This data paucity precludes us from establishing the direction or magnitude of travel mode switch assuming a change in transit fare, service improvement, or the elimination of some routes. In fact it could be argued that the possible heterogeneity in travel mode options is

being driven in part by the differences in the quality of the transportation service, particularly public transit, of which we have no data. There is thus the need to extend the analysis by examining how local jurisdictions make provision for transportation infrastructure or implement travel demand management policies.

From existing literature, travel cost is arguably the most important covariate of travel mode choice that was not explicitly taken into consideration in our analysis as no information exist on the costs associated with each travel mode. However, we posit that this does not invalidate the research results given that oftentimes, public transit fares do not reflect the true value and when it does, it is not unusual that the travel cost is not wholly borne by the commuter. Even when market rates are used, transit fares could still be priced differently for dissimilar segments of the population through the use of vouchers, transit subsidies, seniors' rate, and employers' assistance program to mention just a few.

Building off the data issues itemized above, we had to contend with methodological limitations originating primarily from the nature of the dataset. While we have itemized the distinct advantages to the use of the PUMS dataset, these attributes come at a price. For one, the PUMS datasets are cross – sectional, case specific variable datasets obtained using a complex survey. Thus, most of the stepwise routines or nested logit commands will not work for survey data. In addition, its case-specific nature precludes us from using conditional or mixed logit approaches. The paucity of data, particularly those specific to the travel mode and the absence of proxy data to capture un-observables prevented us from establishing concretely if observed differences in commuting options are a result of differences in individuals' taste parameters (preferences, related to demand) or the quality, accessibility and availability of the transportation infrastructure in existence (policy – related to supply) or the combination of both (where policy influence individuals' preferences).

CHAPTER 3. RETHINKING SFPARK'S PERFORMANCE BASED PARKING PRICING

3.1 INTRODUCTION

In an effort to eliminate circling and reduce parking search time and cruising, SFpark, an innovative demand-responsive pricing program, was implemented by the City of San Francisco. Over a period of 13 months, the program was piloted across seven San Francisco neighborhoods made up of over 6,000 metered spaces with the objective of moving parking blocks towards a 60 to 80% parking occupancy range. This objective contrasts sharply with vehicle turnover goals and/or revenue considerations that are typically the interest of cities' parking programs. Other secondary benefits associated with the SFpark program include reduced greenhouse gas emissions, traffic congestion and ensuring the economic viability of the City's central business districts.

Historically, on-street parking rates are static; often set lower compared to off-street parking options and decoupled from transportation goals (46). The rates are typically determined by City Hall with the objective of funding, for example, a pension plan that may be underwater and at risk of being taken over by the state. Consequently, demand for parking spaces are not taken into consideration. In addition, perverse incentives are created for circling and cruising by not reflecting the true cost of curb parking relative to off-street parking. Finally, no provision exists with regards to exploiting synergies between the program and the city's broader transportation goals.

The SFpark program improves on these shortcomings by implementing a spatio-temporal price adjustment mechanism that modifies parking rates across both parking period and parking blocks based on presently observed occupancy levels. For any adjustment period, the latest parking occupancy levels are compared with the desired 60 to 80% parking occupancy range. Prices are revised downwards by \$0.25/hr. for blocks with less than 60% average parking occupancy in the previous period and increased by \$0.25/hr. for those

blocks with occupancy levels in excess of 80%. Rates remain invariant for blocks within the desired 60 to 80% parking occupancy range. This pricing rule wrestles control of parking rates from City Hall but more importantly, ties the rate changes to parking demand. An evaluation conducted by San Francisco Municipal Transportation Agency (SFMTA), the agency directly responsible for managing the SFpark program, concluded that considerable value has been added by the program and that the program has been able to move parking occupancy in the right direction (47), a viewpoint echoed by the Pierce and Shoup (48). It bears stating that apart from the concerns raised by Pierce and Shoup on the abuse of placard signs, they underscored the need for the program to embrace a more proactive approach through prediction compared to merely a reactionary approach that presently obtains.

The program's effectiveness, particularly with regards to the relationship between price changes and parking occupancy levels, has however been met with mixed reviews by other studies (49), (50), (51). Millard-Ball et al (51), for example, were able to replicate Pierce & Shoup's finding using simulated parking demand derived from a simple random process. Based on this finding, they argued that no evidence exists of the impact on parking occupancy from rate changes and suggested that arriving at a conclusion on the effectiveness of SFpark's demand response pricing may be premature. This assertion was corroborated by the work of Chantal and Manville (49) that found no relationship between the parking rate modifications and parking availability.

Of direct relevance to the conflicting findings and the present study is the lever by which price changes influence demand – price elasticity of parking demand measures. Approximately 35% of the price elasticity measures obtained in Pierce and Shoup were positive, a finding that runs contrary to the fundamental law of demand. Secondly, about 40% of the elasticity figures have absolute values in excess of 1 – indicating elastic measures that are contradicted by previous empirical works on parking demand (52) or by what conventional wisdom would have one believe. It is thus problematic making the case for using these figures as the primary input for a policy instrument given the aforementioned

shortcomings. We assert that a key requirement in ensuring the program's effectiveness is utilizing the correct price elasticity measures. This is particularly the case for a forward-looking policy instrument that uses the elasticity figures as inputs.

The present research delivers on this by using price elasticity measures derived from a panel data regression analysis in an optimization model. Estimating parking elasticity measures using panel data methods extends the boundary of knowledge on this subject given that, to the authors' knowledge, this is the first empirical study to derive plausible price elasticity of parking demand estimates using panel data methods. Predicted parking occupancies are subsequently generated by modifying parking rates via the estimated elasticity measures. Particularly noteworthy is that the model has been designed in such a way that does not put an onerous burden on SFMTA; it facilitates improvement to the program through policy prescriptions and provides flexibility that allows for the use of different priority structures and for addressing black swan events.

The balance of the paper is organized as follows. A brief review of existing literature and related works on vehicle parking are provided in Section 3.2. Section 3.3 addresses data sources and provides an overview of the empirical strategy. Detailed information on the subsets of the empirical strategy - the optimization model and the panel-data regression model is provided in Sections 3.4 and 3.5 respectively. Results and the discussion of findings are presented in section 3.6. Caveats associated with the findings are discussed in Section 3.7. Conclusion and future works are discussed in the last section.

3.2 LITERATURE REVIEW

Parking in the United States takes up a vast amount of space with the total land area for a typical given day estimated to be as large as the state of Massachusetts (53). Apart from the land area dedicated to parking, a by-product of the demand for parking spaces, especially for cities' central business districts is cruising and the attendant congestion, pollution and

other negative externalities it creates. Shoup (54) for example, observed that in a 15-block business district in Los Angeles, more than 100,000 hours are wasted annually by drivers cruising for parking. Arnott and Inci (55) has characterized this behavior – cruising for parking, as an example of the tragedy of the commons.

A major focus of parking research is on how to reduce the congestion caused by demand for parking. Historically solutions to parking problems have been supply-oriented, but with limited public financing for new supplies of parking, these initiative have shifted, of late, to a more demand-side approach (56). These include policy prescriptions oriented towards providing information or modifying parking rates with the objective of achieving desired average parking occupancy thresholds by time and space. While the implementation of this concept may be a relatively novel development, the idea itself is far from new. Vickrey (57) as far back as 1954 argued for temporal-spatial differential pricing for parking; charging different prices for parking at different times and locations, but technological challenges made it difficult. It is only of recent that some cities, San Francisco, Seattle and Los Angeles to mention a few, have embraced temporal-spatial parking pricing with the trend currently being embraced or actively explored by other cities.

One of the reasons provided for the disproportionate interest in the demand for on-street parking is that the pricing does not reflect its real value, particularly when measured relative to off-street parking. In fact, most models of curbside parking as observed by (55) assume that curbside parking is underpriced. This observation has been followed up by a number of papers that examine the competition deriving from geographic differences in parking as well as curb vs. garage parking. A key finding is that proximity is a crucial issue for most patrons with most exhibiting preference for parking within a few blocks of their destination (58), (59), (60), (61), and (62).

An appreciable number of the studies have been directed to examining lopsided parking rates and the mismatch this creates between the demand of on-street parking relative to its closest substitute – off-street parking. Calthrop and Proost (63), for example, examine the

relationship between curb parking and off-street garage parking and posit that cruising for parking will persist for as long as there is differential pricing between curb and garage parking that unduly places demand on curb parking relative to off-street or garage parking. However, if garage parking costs less than curb parking, there will be underutilized curb parking and subsequently, the opportunity exists to price right such parking assets. Calthrop and Proost's work highlights how rival parking options, in this case curb and garage parking, can cause inefficiencies if pricing for both are not considered together.

The inertia on getting the prices right for curb parking has typically emanated from the public that perceives curb parking as a public asset whose consumption should be subsidized –in contrast to garages that are oftentimes, privately owned. However, this pricing strategy creates perverse incentives given that the highest demand for parking tends to be in the most densely populated and commercial centers of urban areas. In the absence of a supply side approach that has proven to be prohibitively expensive, a spectrum of demand side interventions, of which intelligent pricing of parking is one, has increasingly being pushed as a viable option. This improves considerably on the static parking meters that was first introduced in Oklahoma city with the objective of increasing the parking turnover rate, approximately eighty years ago (56), (50).

The pricing strategy employed for the smart parking meters is demand response pricing. Vickerman (64) made the case that this is a variant of congestion pricing as pricing parking in order to reduce cruising had the same end result as congestion pricing, but with a lower cost. This is certainly the case as pricing and charging cars which are parked is much easier than doing so with cars that are in motion. This pricing approach has been implemented in a number of cities though with different nomenclature such as performance based pricing or variable pricing. The findings from these programs have been documented by a number of research studies itemized in the next couple of paragraphs.

Clinchant et al (65), for example, examined Los Angeles' LA Express Park, in which 6,324 on-street sensors were installed onto the streets of Los Angeles to monitor local parking

behavior; his findings emphasize the need for methods of preprocessing and more robust models for analyzing large data sets, as well as a need to examine policies on placard use and incentives for walking. In Seattle, Ottosson et al. (52) describe findings from a performance-based parking pricing implementation there, where the target occupancy rate was set at 71–86%. Similar to San Francisco’s SFpark, short-term parking is found to be inelastic; though over time, drivers did make adjustments and park longer on blocks where prices were lowered and parked for shorter times where prices rose.

A meta analyses of smart parking applications was also carried out by Lin (66) where she examines a number of smart parking implementations and divides them into three groups; information collection, service deployment and service dissemination. Each of the groups depends heavily on technology that may include smart phones, sensors and network infrastructure. Her work gives guidance to cities that are considering designing and launching smart parking services while highlighting the policy take-away of the programs analyzed. Apart from the aforementioned, policy implications of performance based pricing also includes the work of Russo (67) on voting preferences and parking; employer provided parking (68), (69) with a focus on how benefits distorts public welfare as a result of narrowly targeted benefits; and employee cash out policies (70), (71), (72), (73), (74), (75) in which employees are given the choice of taking the cash equivalent of employer-provide parking, and the role this option plays in moderating downtown demand for parking

Numerous studies have been carried out specifically on the SFpark program including a comprehensive report that itemized the program’s objectives and preliminary findings that were documented in SFpark 2013 evaluation report (47). The evaluation report revealed that the program has been effective in getting more parking blocks within the desired 60 to 80% desired parking occupancy range with attendant benefits of less circling for parking; reduced parking search time; reduced green gas emissions; an easier payment system and a decrease in peak period congestion. The independent evaluation of the pilot program has however met with mixed reviews particularly with regards to the relationship between price changes and occupancy levels. Pierce and Shoup (48) painted a rosier picture of the

program's effectiveness by demonstrating that the program has made appreciable progress towards its goal of ensuring parking availability though they admitted that work needs to be done particularly with regards to the price elasticity figures obtained and the need to address the widespread abuse of disabled placards. In addition, the authors acknowledge that a purely reactive approach to modifying prices will be less effective compared to a predictive one that allows SFpark to proactively manage the program.

Another group of researchers (50), (51), however dispute the claims made by Pierce & Shoup on SFpark program's impact. Millard-Ball et al. (51) believe that Shoup's analysis does not take into account the endogeneity of prices, neither does it consider that SFpark's rate adjustment rules could be triggering price changes due to fluctuations in demand. The authors employed a regression discontinuity method to address endogeneity issues in Pierce and Shoup's model, with a finding that the short-run elasticity is indistinguishable from zero. Utilizing an alternative approach, the authors make the case that price elasticities obtained by Pierce and Shoup could be replicated merely by simulating parking demand as a random process totally devoid of the parking rate influence.

This dissenting voice on the program's impact was also echoed by Chatman & Manville (49). Apart from administrative data from SFpark, the authors collected information using surveys from patrons and concluded that not only did the response of parking demand to rate changes run contrary to expectations, similar patterns were also observed for turnover, carpooling and parking duration. A multitude of reasons were advanced for why these results came to be. For one, they opine that the expected relationship may not hold because of the patrons who park for free – typically individuals with the handicap placard. In addition, they argued that other changes, such as an easier payment system and relaxing the parking duration were implemented simultaneously as SFMTA was rolling out the pilot phase of the program, thus making it problematic to isolate the impact of price on occupancy levels. Further compounding establishing the program's impact are the caps placed on price changes and price ranges. If it is indeed the case that parking is inelastic as demonstrated by numerous numerical studies (52), (76), (77), it will be difficult to

demonstrate significant effects within that price range, even more so given San Francisco wealthier (relative to any other comparable city in the U.S.) workers and residents.

3.3 EMPIRICAL STRATEGY AND DATA SOURCES

The empirical strategy employs a two-stage model made up of two parts – a panel data regression analysis that provide estimates of price elasticities of parking demand for each block for each parking period and an optimization model that employs the estimated elasticities to generate new parking rates. The strategy is designed with the goal of moving San Francisco’s curb parking occupancy in the right direction using performance based pricing. The pricing strategy varies rates based on parking demand by location (city’s block) and time of the day (morning, afternoon or evening). Information on occupancy was obtained either through the parking sensors or the smart meters. The objective is to keep average parking occupancy between 60 and 80% for any block for each of the parking periods itemized. Prices are lowered for blocks with low occupancy and increased for those with higher than acceptable parking occupancy levels. There has been some value added by the program given the observed reduced cruising and the decrease in the number of blocks whose occupancy fall outside the 60 to 80% desired occupancy range (46), (47).

Apart from the noise inherent in using smart meters to proxy parking occupancy data, stemming largely from patrons who park but do not pay, there has been other identified shortcomings of SFpark’s performance pricing strategy (51). For one, price is the sole policy instrument used in influencing parking occupancy and the wide range and magnitude of price elasticity of parking demand measures that were in excess of ± 7 obtained by Pierce & Shoup, 2013 is an indication that other factors are at play. While the range provides an indication of the degree of variation across blocks at each parking period of the day, even more insightful and more troubling are the magnitudes and the signs of the elasticity measures obtained. Approximately 35% of all the elasticity measures calculated by Pierce & Shoup were positive, results that violate the basic law of demand. Except for Giffen goods, whose consumption increases with increase in price and vice-versa, the price elasticity

measures of demand should be negative. Finally, borrowing from empirical studies (52), (78) and conventional wisdom, demand for parking should be fairly price inelastic, that is, the magnitude of the elasticity measures should be less than 1. It is thus problematic that the primary policy instrument, influencing occupancy by modifying prices runs contrary to economic theory.

Utilizing realistic elasticity measures is a *sine qua non* for moving to an approach that uses prediction compared to merely a reactionary approach in setting prices that presently obtains with SFpark. Implicit in the concept of elasticity is the “*ceteris paribus*” assumption – *all other things being equal*. Controlling for factors, which explain demand for parking spaces over and above price changes, should in theory, provide more realistic elasticity measures that could subsequently be used for predicting parking occupancy that SFpark can better plan for and respond to. Ideally, this should be done in a manner that does not place undue burden of data collection and analytical expertise on SFpark. This is even more imperative given that the program is meant to be a demonstration program that could be replicated in other cities.

The present study addresses these issues by using an approach that employs estimates of the price elasticities of parking demand obtained through a panel data regression model in an optimization process. The objective of the optimization is to minimize deviations from the 60 to 80% parking occupancy target. The price elasticities derived guarantees that we are not navigating with the wrong instrument and ensures that the policy levers – adjusting parking rates to influence parking occupancy via the price elasticities of parking demand – are credible. The integrated model facilitates a more systematic and holistic approach that examines not just a block but the totality of the program. In addition, sensitivity analysis could be carried out and evaluated with respect to figures of merit that allows for objective comparisons, either relative to SFpark’s approach or simulated policy scenarios.

The focus of the optimization is the policy instrument of choice – the parking rate adjustment that are the model's decision variables. SFpark's rule based pricing is reflected as policy constraints in the optimization model – price changes (-\$.5 to \$.25); price range (\$.25 to \$6). By making modifications within these permissible ranges, we can examine the value added by the pricing regime that will be proposed by the optimizer relative to some figures of merit – number of blocks that fall outside the desired range or the cumulative deviations from this range. In addition, policy insights could be gleaned by examining how relaxing these constraints impact on SFpark's objectives. This include extending the range of possible permissible prices or allowing more flexibility with regards to price changes – for example, allowing a \$.05 price increase instead of SFpark's \$.25 cent mandated price changes. The policy scenario entails running multiple optimization using plausible parameter and right hand side values for the constraints. These changes pertain to modifications to rate and parking rate range using both continuous and discrete jumps in price.

To ensure comparability, the model was run using the same random number generator. Parking supply is known and fixed at the maximum capacity and observations will be block by time band or a long form type variable that concatenates time and location. The objective function seeks to minimize deviation from the desired 60-80% occupancy threshold. Decision variables are the price changes for each block by time band and parking occupancy is predicted by changing price and funneling it into a linear prediction model that uses parameter estimates obtained from the panel data regression. Parking occupancy prediction is carried out for just one time period into the future. This approach allows us to compare the hold out dataset with the predicted outcome. The model uses both OptQuest and Genetic Algorithm for the optimization.

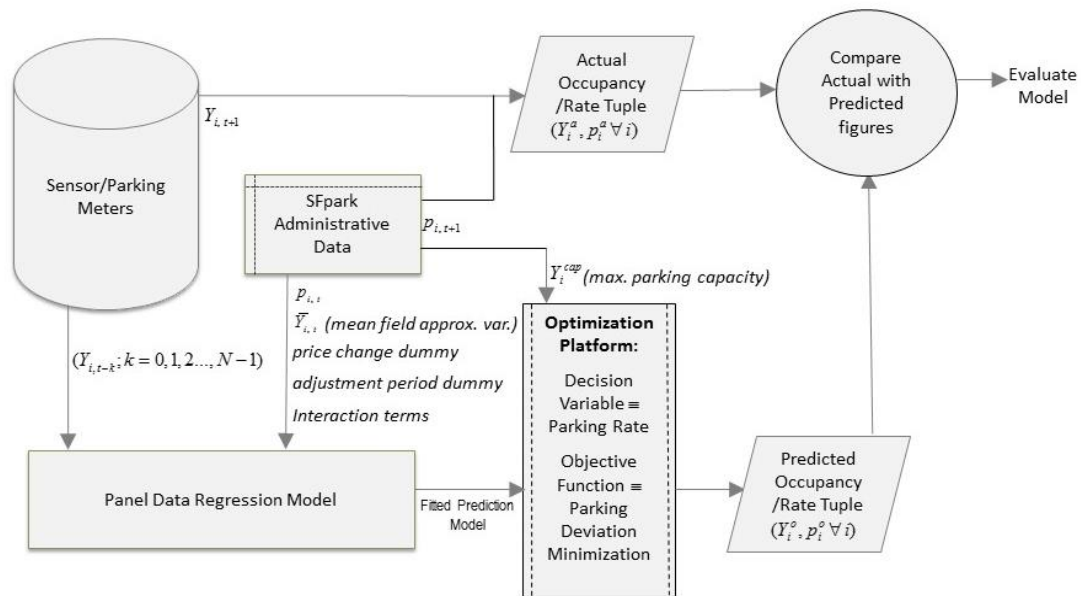
The model was also used to make the case for a revenue neutral approach for the performance based pricing. For one, this calms down the public by guaranteeing that the program is not being turned into a revenue generating venture by City Hall. Secondly, it ensures that revenue streams across the periods are comparable and that public policy

programs that are financed by the revenue generated, such as public transit, will not be jeopardized. In instances where this is not the case, insights from the model could sensitize SFMTA as to the need to make alternative arrangement if a lower revenue stream is expected. The revenue goal is structured as a soft constraint that could be violated but with penalty attached. A recipe model thus becomes the natural default to use for the optimization model. Finally, our approach avoids situations where a price change across board is implemented for the blocks as witnessed in first quarter of 2016. A well designed policy instrument should have the capacity to fully internalize such decisions.

A key innovation used for the panel data regression stage is the use of a simplified form of mean field approximation (79), (80) as an added regressor to control for parking demand in proximal blocks. This approach reduces the n-body system to a 1-body system by averaging, for the present problem, parking occupancy across the n-blocks that share a boundary with block i . The data are sourced from SFpark sensors or parking meters. Addressing this factor using average occupancy for the relevant blocks not only controls for correlation across blocks but also yields more realistic price elasticity estimates.

Figure 3-1 presents a panoramic view of the two-stage panel data regression/optimization model. Data for the panel data regression come from sensor/parking meters and SFpark administrative data. Coefficient estimates from the regression are subsequently used to fit a linear prediction model that is the primary input to the optimization platform. The platform, which employs both OptQuest and genetic algorithm solution methods, seeks to minimize parking occupancy deviation from the desired range by modifying parking rate via the price elasticity of demand obtained from the regression model. The model effectiveness is evaluated by comparing actual occupancy/rate tuples with the optimized result to ascertain the potential improvement in the program's performance. Sensitivity analysis are carried out by amending SFpark pricing rules and a discussion is provided on the associated policy implications.

Figure 3-1: Panoramic representation of the two-stage panel data regression/optimization model



The data used for the study are sourced solely from the City of San Francisco parking sensors, smart meters and SFpark administrative data. Variables associated with the sensors and the parking meters include the dependent variable – the future parking occupancy; the mean factor approximation variable and the present parking occupancy for each block time band. SFpark administrative data related variables include the identity and parking rate of each block time band; parking supply for each block which equals the maximum parking capacity. In addition to these variables, dummies for the price adjustment period and blocks that experienced price changes were also created. Finally, a composite variable was derived by interacting each block time band by the associated parking rates.

The time windows are classified broadly into six: three – morning, afternoon and evening each for weekdays and weekends. Each time window is typically a three hour period and runs usually from 9 am till noon; noon till 3 pm; and from 3 pm till 6 pm though there are variations in these hours depending on the neighborhood. Using this classification, each block in the pilot program will have six distinct observations. Working off the

nomenclature used by SFpark, a sixth digit is added to the blocks to uniquely identify the reference parking period. For example, 41905 represents the 500 block of Franklin Street – a sixth digit that captures the time period, is concatenated with the five digit label with 1, 2, 3 values representing morning, afternoon and evening periods during the weekday and 4, 5, 6 the same periods during the weekend.

3.4 OPTIMIZATION MODEL

The optimization is made up of an objective function that minimizes deviations from stated goals; goal related constraints that include occupancy and revenue targets; policy constraints; system related constraints and technical and non-negativity constraints. The optimization is carried out for blocks in the City's Civic Center neighborhood across six time bands – morning, afternoon and evening for both weekdays and weekends. The model is formulated in a myopic way – a rolling horizon basis with a one period look ahead prediction. Detailed information on each of the model component and the algorithms employed are provided below.

Objective function

Let d_v^+ represent the amount by which occupancy deviate upward from goal v and d_v^- represent the amount by which it deviates downward from goal v .

$$Min Z = f(d_v^-, d_v^+) = \sum_{v=1}^V P_v(d_v^-, d_v^+)$$

The expression on the right is the cumulated weighted ceiling and floor goal constraints. The essence is to minimize the upward deviations (d_v^+) for the ceiling constraints and minimize the downward deviation for the floor constraints (d_v^-). Therefore;

$$\text{Min } Z = \sum w_{cei} d_{cei}^+ + \sum w_{flr} d_{flr}^-$$

Here, the ceiling constraint is the 80% upper threshold specified by SFpark and the floor constraint is the 60% lower threshold. The objective above could be made specific to SFpark's agenda by rewriting the expression as specified below?

$$\phi_l \leq \frac{\sum_{n=1}^N x_{itn}}{nC_i} \leq \phi_u$$

Since parking rates are hourly, C_i is the hourly capacity for each block; x_{itn} is the n th hour number of occupied parking spaces for block time band i at parking period t ; n is the number of hours over the duration of the existing parking rate; the expression in the middle is the duration's average block occupancy and in SFpark's case, $\phi_l = 0.6$ and $\phi_u = 0.8$. Presently, the deviations are equally weighted though the objective function could be easily modified to allow for differential weighing. This minimization objective is subject to the following:

Goal Related Constraints

Occupancy target

This is the primary goal constraint represented by the expression below. In the equation, the deviations are commensurate and could be cumulated to provide SFpark with a figure of merit. Priorities could also be attached to each deviation by using weights. For example, upward deviations could be weighted differently compared to downward ones.

$$\phi_l + d_i^- \leq \frac{\sum_{n=1}^N x_{itn}}{nC_i} \leq \phi_u - d_i^+$$

Observe that the deviations are computed for each block for a specific time band summing together the occupancy within the duration. n represents the number of hours the pricing regime was operational for any given block. In our analysis, there are 352 distinct block ids, each with six time bands, yielding a total of 2112 block by time bands. Our analysis however focuses solely on the Civic Center neighborhood and excluding the *block time band* with missing data and those with positive price elasticity measures, the number of the observations reduces to 124.

Revenue

To ensure comparability with SFpark's occupancy and rate figures in time $t + 1$ the optimization is run without the revenue constraint in time t to generate predicted occupancy and rate figures for period $t + 1$. We compare *what is* with *what ought to be* using these results with regards to key performance metrics – number of blocks that fall outside the desirable occupancy threshold; cumulative deviations from the desired occupancy level across the blocks, number of blocks with price changes and revenue volume using 15th adjustment period as the base period and predicting one period into the future. We subsequently introduce revenue as an added constraint – structured with the objective of maintaining near-revenue neutral price adjustment. This is reflected by minimizing the absolute deviation of the present revenue compared to the revenue obtained for the previous period. In the expression below, \mathbf{X} is a column vector of occupancy and \mathbf{R} is a row vector of parking rates. This constraint is flexible and could be modified easily to reflect the agency's preferences. For example, SFMTA may desire that the program be self-sustaining i.e., the revenue it brings in could sustain its operations. Alternatively, it could be related to what such revenue is devoted to – for example, it could be tied to the amount San Francisco uses yearly to subsidize the public transit system or a $\pm 1\%$ maximum revenue deviation compared to the previous period as shown below,

$$0.99M_{-1} \leq \sum_{i=1}^I \sum_{n=1}^N XR \leq 1.01M_{-1}$$

where N is the total number of hours each metered block i is in operation over the duration of the adjustment period and M is the revenue.

Policy Constraints

Pricing – parking rates

Hourly parking rates cannot be less than \$0.25 or more than \$6. The optimization could be run with these constraints relaxed with the result providing a directional guideline or serving as a policy advocacy tool.

$$25 \leq r_i \leq 600$$

Pricing - occupancy bands and price changes

Parking rates (in cents) are adjusted based on predefined average occupancy thresholds. If average occupancy is $\geq 80\%$, prices are adjusted upwards by 25cents; if occupancy rate is between 60 and 80%, there is no price change and there is a 25cents decrease in parking rate if occupancy is below 60%.

$$r_i^n = r_i \text{ if } 0.6 \leq \frac{\sum_{n=1}^N x_{itn}}{nC_i} < 0.8$$

$$r_i^n - r_i = 25 \text{ if } \frac{\sum_{n=1}^N x_{itn}}{nC_i} \geq 0.8$$

$$r_i^n - r_i = -25 \text{ if } \frac{\sum_{n=1}^N x_{itn}}{nC_i} \leq 0.6$$

Alternatively, the expressions above could be replicated using a mutually exclusive equation with multivalued indicator function as shown overleaf:

$$r_i^n = r_i + gZ$$

Where g is the change in price (=25¢) and Z is an indicator function that returns either 1, 0 or -1 depending on how the average occupancy compares to the threshold values.

Elasticities and parking demand

Prices are used to influence demand thus it is imperative that elasticity measures that relate changes in price to changes in parking demand be derived for each *block time band*. Parking demand is estimated using a fitted regression equation from the linear regression equation below:

$$\ln Y_i = y_i = \sum_{k=1}^K \beta_k X_{ki} + u_i$$

If we assume that $X_{1i} = 1 \forall i$, the second variable, the natural log of prices ($\ln P_i = p_i$) and the third, the interaction term between the natural log of prices and entity specific dummies ($p_i \# D_i$), then the regression equation could be rewritten thus:

$$y_i = \beta_i + \beta_2 p_i + \beta_3 (p_i * D_i) + \sum_{k=4}^K \beta_k X_{ki} + u_i$$

and an analogous fitted linear regression that generates the conditional mean of the parking occupancy is specified thus:

$$\hat{y}_i = E(y_i | X_1, X_2, \dots, X_K) = b_1 + b_2 p_i + b_3 (p_i * D_i) + \sum_{k=4}^K b_k X_{ki} + u_i$$

With this expression, given the log-log relationship between the dependent variable and prices, $b_2 + b_3$ will be the percentage change in demand for parking occupancy given a 1% change in prices. Essentially, it represents the price elasticity of parking demand for each observation. Observe that an equivalent derivation of the elasticity estimates could be obtained by assuming that the parking demand function was generated from a variant of a Cobb-Douglas utility function specified below:

$$Y_i = A W_i^\beta P_i^{\gamma_i} e^{u_i}$$

taking natural log of both sides give:

$$\ln Y_i = y_i = \alpha + \beta W_i + \gamma_i p_i + u_i$$

where the lower case letters represent natural log of the explanatory variables and γ_i is the coefficient of interest. The equation above, when fitted, yields the same conditional mean expression derived earlier. Observe that γ is observation specific while the intercept estimate and the slope parameters for the other explanatory variables are observation neutral.

System Related Constraints

Occupancy

The number of occupied parking spaces for any metered block cannot be more than its capacity. These could be modeled as hourly capacities which apply irrespective of the time band and provide a straight forward interpretation given that the parking rates are hourly.

$$x_i \leq C_i \forall i$$

Technical and non-negativity Constraints

Non-negativity requirement

$$x_{it}, C_i, d_i^+, d_i^- \geq 0$$

$$d_i^+ d_i^- = 0$$

Algorithm and Model Implementation

During the optimization, the decision variable is modified within the allowable bandwidth as specified in the model and for each set of values, the model is recalculated. For the search, algorithms are used to comb through the possible combination of decision variables that yield the most preferred solution. Two types of algorithm were employed for the study – OptQuest and Genetic algorithm (GA). OptQuest uses a combination of techniques that

blend Tabu Search, Neural Networks and Scatter search while GA draws on the biological principle of natural selection to iteratively improve on the solution, with each progress log providing better or no worse off solution set compared to previous ones for the search process. The recipe solving method was used and the model was implemented using Evolver software.

We have endeavored to have comparable scenario runs by implementing procedures that guarantee that these runs could be placed on an even keel. For one, the stopping condition was the same across each optimization. The stopping condition was specified as a 0.01% threshold percentage improvement (reduction) in the objective function using the immediate past 10,000 trials. The optimization stops if the improvement is less than 0.01% of the objective value. Apart from this, the same initial seed was used across the runs. These are random number generators that are relevant to both OptQuest and GA. Finally, the same value for population, crossover and mutation rates and the same set of genetic operators were employed for the optimization.

The choice of the optimization engine was determined automatically by the software depending on whether changes to the decision variable are continuous or discrete. OptQuest is used for discrete changes while the software defaults to the GA for continuous changes in the decision variable. Evolver uses a steady-state approach which means that replacement is done on a piecemeal basis as compared to wholesale changes that may affect a whole generation. The crossover and mutation rates are subsequently applied in determining how solution sets are chosen or generated from the current solution possibilities. Often this includes trading how much resources, typically computing power that will be expended on finding new solution space compared to the time spent on improving the identified solution space.

3.5 REGRESSION ANALYSIS

Regression analysis was conducted using data from all the blocks that featured in SFpark's pilot program. An observation is a *block time band* – a specific block for a specific time window. Associated with each observation are the following variables:

- $Ln y_{it}$, future parking occupancy of the it th observation, is the dependent variable.
- $Ln \bar{y}$, which represents the average parking occupancy for blocks in close proximity to the it th $blkid_tiba$. The variable is used to proxy the *mean field approximation factor*;
- lnp_p , the present price in natural log form expressed in cents which represent the primary policy instrument;
- lny_p , the present period occupancy for the it th observation. To some extent, this variable captures the *block_time_period_band* specific features;
- pr_dum , price dummy for each observation that equals one if there has been a price change in the previous period, and zero otherwise. It could be argued that it complements lny_p in its ability to control for observation specific features;
- adj_per , dummies for the adjustment period that capture the time effects;
- $lnp_p \# blkid_tiba$, the interaction terms between the natural log of price expressed in cents and block by time band dummies.

A plausible model specification is to pool the data across all the time periods, thus, constraining the idiosyncratic effects to be driven solely by individual observations. Essentially, what we are doing in this situation is running a pooled or population-averaged model across the adjustment periods.. This is the same model as in cross-section case, with adjustment for correlation over time for given individual *blkid_tiba*. The formal test of the restriction is carried out using the F test to establish the joint explanatory power or significance of the adjustment period variables. The test uses the null hypothesis of all the partial regression coefficients associated with the adjustment periods being equal to zero compared to an alternative hypothesis that has at least one of these partial regression coefficients, not equal to zero. The expression for the hypothesis test is:

$$H_0 : \beta_j = \beta_{j+1} = \dots = \beta_k = 0$$

$$H_1 : \text{at least one } \beta \neq 0$$

Where the regressors X_j, X_{j+1}, \dots, X_k , $j < k$ with corresponding parameter $\beta_j, \beta_{j+1}, \dots, \beta_k$ are all associated with the price adjustment period dummies – 16 adjustment periods in total with the last one used as a hold-out sample set. For identification purposes, we have the option of setting either the intercept term or one of the dummies to zero. Depending on the option used, the parameter dummies are interpreted either as the adjustment period effect or the difference between the effects of the “ t ” adjustment period and the reference period. Assuming the former option was exercised and given T periods, $T-1$ dummies will be required with the excluded one being the reference period.

Pooling across the adjustment periods is only acceptable if the test of restriction that imposes a null hypothesis of zero parameter values on the adjustment period dummies hold. Secondly, strong and restricted assumptions are needed to make it acceptable. For example, the error terms should be uncorrelated, identically and independently distributed with zero mean and constant variance - In this situation, the usual ordinary least squares (OLS) approach provides the best linear unbiased estimator. However, rarely in practice do we have such well-behaved error term. The norm is that there are other relevant predictors that are not taken into consideration. What is typically the case is that such omitted variables get subsumed underneath the error term. For example, assume we have a model that in the ideal world, has error terms with properties listed above. However, the model is mis-specified because of the absence of a constant variable d , specific to each observation that is unobserved. Given that this variable is not controlled for, d gets absorbed into the error term. Consequently, there is a new composite error term that comprises of the idiosyncratic component u and d , the unobserved heterogeneity specific to each block for each time period.

This issue is of relevance because of the violation of a couple of regression requirements. The Gauss Markov conditions – zero covariance among the disturbance terms and zero covariance between the disturbance term and the explanatory variable may be violated leading to biased and inconsistent estimators. The violation may exist within a cross sectional dataset when the unobserved heterogeneity is correlated with any of the predictors, leading to inconsistent and biased OLS parameter estimates (81). Even when the OLS approach produces unbiased estimates for a single cross section, there may be efficiency losses when the observations are pooled across the adjustment periods given the correlation among the error terms. For example, if we assume a T time period or T observations per block with the regression equation specified thus:

$$Y_{it} = \sum_{k=1}^K \beta_k X_{kit} + u_{it} \quad \forall t = 1, 2, \dots, T$$

where $u_{it} \leq d_i + v_{it}$ then the covariance of the error terms across two time periods, say $t=1$ and $t=2$ is not equal to zero given that the error terms will be correlated as a result of the unobserved heterogeneity term, d . Such unobserved effects, also called error components are best addressed using panel data methods. The term *effect* here is defined to mean association.

In addition to its effectiveness in addressing assumption violations of this nature, a panel data regression lends itself naturally to the dataset – we have data on the same set of observational units ($n=1, 2, \dots, N$ - blocks that feature in the pilot for each time period) across a number of adjustment periods ($t=1, 2, \dots, T$). Panel data offer multiple advantages over cross sectional data. For one, it can isolate short run effects from long run ones and if properly specified, could be effective in separating static effect (related to size or scale economies) from dynamic ones (learning or technology innovation). Of crucial importance is the form the regression parameters are allowed to assume. For example, given the regression equation expressed in scalar form overleaf:

$$Y_{it} = \beta_{1it}X_{1it} + \beta_{2it}X_{2it} + \dots + \beta_{kit}X_{kit} + u_{it}$$

Let us assume that we have a balanced panel data and that $X_{lit} = 1 \forall i, t$ and that we are interested in explaining parking occupancy as a function of the “K-1” explanatory variables. Using the expression above, that allows the intercept term, the slope coefficients and the disturbance terms to vary for each observation unit across each period, we will need to estimate $K \times N \times T$ parameters, an infeasible task given that we only have $N \times T$ observations. However, if we assume that the intercept term does not vary across the distinct observations and that the slope coefficient for the policy relevant variable may vary across each block time band, then, the shift parameter is the same for all entity. With this construct, differences in the estimated parking demand could only come from differences in magnitude of the predictors (endowment); and/or the varying slope (returns to predictors). If these assumptions hold, any differences in parking occupancy, other than differences in the levels of the explanatory variables should be captured by the varying slope term. We investigate examining the impact of these explanatory variables further by using fixed and random effect models that are widely applicable in situations where the number of cross sectional units or observations (N) is appreciably more than the adjustment time periods (T).

Using different model specification and functional form, a formal test of the adjustment periods parameters using an F test to determine their joint explanatory power revealed a p value = 0.000 with F test value of 17.41, showing that they are statistically significantly different from zero, thus we reject the null. We subsequently ran a panel data regression made up of N observations (number of blocks by time period) over T adjustment periods and on the basis of the adjustment periods being significant, made the case for a two-way effects model where the effect is not only attached to the individual observations – the distinct *blkid_tiba* but also to each adjustment period. Effectively, we conditioned out both the block by time band and the adjustment period effects. The choice between fixed and random effects model was determined by the data context, statistical properties of the error

term, the validity of the assumptions made and the objective of the regression analysis. A formal test could also be performed to determine if the error terms are correlated with the regressors.

Clarke et al (82) argued that assuming a rich enough dataset and a policy related objective, the preferred modeling choice should be the random effects model. They argued that even though the RE method is premised on strong assumptions, the flexibility this affords may more than compensate for the attendant caveats when these assumptions do not hold. They argued that this is particularly the case when policy relevant inferences have to be drawn from the analysis. Establishing the validity of the assumptions is extremely problematic except experimental data are available and in the absence of this, a more feasible objective may be one of producing estimates that are policy relevant even though they exhibit bias. This viewpoint was echoed by Snijders et al (83) who showed that in instances where the random effects (RE) assumption of the error term is not valid, the RE approach may still provide superior results partially because of the flexibility of the modeling process and the efficiency advantages (with regards to the estimators) it facilitates. Given the context of the dataset and the aforementioned reasons, we have chosen to use the RE modelling approach.

Of more importance is the heterogeneity across the block time bands and how these variations that are not related to chance influence data on parking occupancy. A common effect or association is assumed for fixed effect model while the effect size is allowed to vary for random effect models. It stands to reason that the effect size may be higher for blocks that receive a higher dosage of the intervention, say, the frequency to which parking payment compliance personnel patrol the blocks. It is also plausible that such monitoring may not be uniform across blocks. For example, it is expected that blocks closer to the core, those that bring in more revenue or those that have historically proven to be more problematic may be monitored more closely compared to those in the peripheries or that do not share the attributes mentioned. Given the aforementioned reasons, we have chosen to use the RE modelling approach.

Modifications were made to the generic RE model to adequately capture the present situation we are trying to represent by introducing a two – way effect that allows the intercept term to vary both by entity and by the adjustment period and a random coefficient model that allows the impact of the policy variable to vary by observation. Unit effects are controlled for by including both the lagged term of the dependent variable and a dummy variable if the entity experienced a price change in the immediate past period. Apart from controlling for these effects, the regression equation was structured with an eye towards the policy instrument of choice – changing price to influence parking occupancy. For one, we use a log-log relationship between parking occupancy and price that allows us to retrieve elasticity measures directly from the coefficient estimates generated. In addition, we model the parameter estimates of the price variable as consisting of two components – a slope that captures the overall fixed effects of the average observation or entity – defined here as a unique block for each unique time period and a random deviation from the slope that is observation specific. The random coefficient modification allows us to differentiate price elasticity measure for each block by time band.

3.6 RESULTS AND DISCUSSION OF FINDINGS

3.6.1 Coefficients and price elasticity estimates

Results for the regression analysis are shown in Table 3-1. Given the number of distinct block by time bands (233 in all), price elasticity estimates for each block by time period observation are excluded from the result. The *robust* option introduced for the Stata command (84) guarantees that standard errors are adjusted for heteroscedasticity though it bears mentioning that even when the standard errors were not adjusted for heteroscedasticity, all the predictors listed in Table 3-1 were all statistically significant at the 1% significance level except for the lagged dependent variable that was significant at the 10% significance level.

As could be seen from the Stata command below, the predictors include a lagged form of the dependent variable and a price dummy that captures whether a price change happened on the block the previous period. Both effect combine controls for block specific effects. In addition, price elasticities are allowed to vary across the blocks by using both an overall fixed effect (*lnp_p*) and another that is block specific (*c.lnp_p#i.blkid_tiba*) that collectively capture the variation in price elasticity of parking demand for the blocks. A formal Chow test reveals that the slopes are indeed significantly different from one another in a statistical sense. The Stata command is provided immediately below and a more detailed information on the calculated elasticity estimates is provided in Figure 3-2.

```
xtreg lny_f lny_p lnp_p pr_dum c.lnp_p#i.blkid_tiba lny_bar i.adj_per, re vce
(cluster blkid_tiba)
```

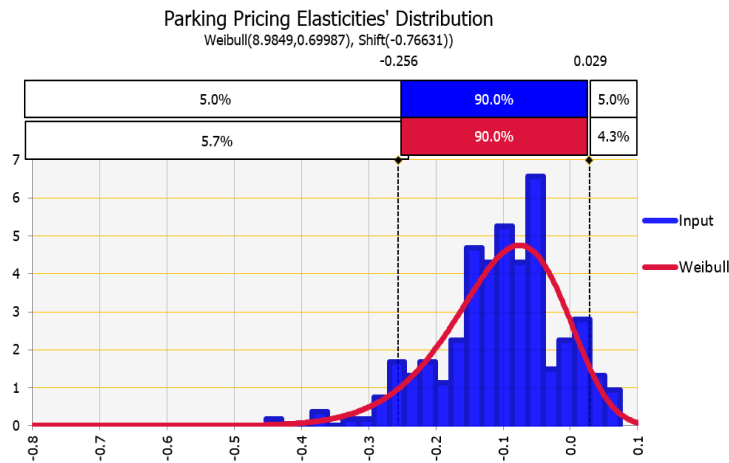
Table 3-1: Panel Data regression result

Independent Variables	Coefficient	Std. Err.	z	P> z	[95% Conf. Interval	
Natural Price Log	0.084	0.047	1.80	0.072	-0.008	0.176
Price Change Dummy	-0.049	0.015	-3.38	0.001	-0.078	-0.021
Mean Field Approx.	0.402	0.055	7.34	0.000	0.295	0.509
Adjustment Period						
2	-0.117	0.019	-6.14	0.000	-0.154	-0.079
3	-0.182	0.030	-6.15	0.000	-0.240	-0.124
4	-0.317	0.039	-8.19	0.000	-0.393	-0.241
5	-0.360	0.043	-8.41	0.000	-0.444	-0.276
6	-0.254	0.032	-7.85	0.000	-0.317	-0.191
7	-0.166	0.023	-7.18	0.000	-0.211	-0.120
8	-0.284	0.032	-8.76	0.000	-0.347	-0.220
9	-0.380	0.039	-9.84	0.000	-0.455	-0.304
10	-0.199	0.029	-6.89	0.000	-0.256	-0.143
11	-0.306	0.042	-7.38	0.000	-0.388	-0.225
12	-0.243	0.044	-5.57	0.000	-0.328	-0.157
13	-0.261	0.028	-9.32	0.000	-0.316	-0.206
14	-0.263	0.030	-8.72	0.000	-0.323	-0.204
15	-0.214	0.026	-8.12	0.000	-0.266	-0.162

What is particularly noteworthy is the potential impact of the mean factor approximation variable - *lny_bar*, the variable that captures the correlation across blocks for any specific parking period. No other variable has a comparable impact on the estimated parking

occupancy either when calculated through the *log-log* relationship that exists between the dependent variable and the regressors or in the *log-linear* relationship between parking occupancy and dummy regressors. *lny_bar* is calculated as the average occupancy of blocks in close proximity to the reference block for any one parking period. It captures a simple approximation of the mean field theory, where many interactions are replaced by just one effective interaction – here the average occupancy of blocks that share a boundary with the block of interest for a specific parking period.

Figure 3-2: Distribution of price elasticities of parking demand



Apart from the parameter estimates in Table 1, price elasticity measures were also estimated for each block for each time period. The estimated price elasticities are provided in Figure 1. The estimates obtained, with a range from -0.45 to 0.07 were fitted to a distribution using several information criteria and the one with the best fit is the Weibull distribution with 8.985 shape parameter value and a 0.7 scale parameter value. A third parameter that addresses the location is included in form of a shift value of -0.766 that shifts the domain of the distribution in which the function is included by that magnitude to the left.

Except for the positive values obtained, the elasticity measures fall within what was observed in previous rigorous empirical studies of parking demand (52). It also presents

more plausible findings compared to the works of Pierce and Shoup (48) where price elasticities values were derived not from regression but solely from price – occupancy tuples. The results from the present regression are more believable – for one, the negative values go without saying, second, the appreciable variation across the blocks monitored reflects what happens in reality with high parking demand blocks having lower (in absolute terms) elasticity measures and finally, the inelastic nature of the parking demand with respect to price changes conforms to what conventional wisdom dictates. The positive elasticity measures were observed only in five blocks – the Zero block of Grove Street; the 300, 400, 500 blocks of Hayes Street and the zero block of Oak Street and this is typically the case for all the parking time periods. The elasticities, excluding the positive ones were subsequently used for the optimization model in the next section using the Civic Center neighborhood as a test case.

3.6.2 Figures of merit and assessing SFpark’s effectiveness

While all the blocks participating in the SFpark program were covered in the panel data regression analysis, only a subset of the blocks were analyzed for the optimization module. We focused on Civic Center, a neighborhood with forty two blocks in the pilot program. Only a subset of these blocks by time band – 124 in total - featured in the optimization. Those with missing data and blocks by parking period with positive elasticities were excluded from the analysis.

The predicted occupancy per observation and the associated prices were forecasted one adjustment period into the future. To provide comparison as to the value added by our approach, the 16th adjustment period is employed as a hold-out sample dataset. The regression used a long form of the data from adjustment period one to 15. The fitted regression was subsequently used to predict what should be for the 16th period. The actual occupancy-parking rate tuple for each observation is then compared to the predicted values using two performance metrics – the number of blocks by time band that do not fall within the desirable 60 – 80% occupancy threshold and the cumulative magnitude of the

deviation across the 124 observations. The predictions for the 16th price modifications are calculated using the coefficient estimate for the 15th adjustment period. This is an acceptable practice as argued by Baltagi (85).

Table 3-2 compares the performance metrics – percentage of blocks that deviated from the 60 – 80% parking occupancy range and the cumulative magnitude of the deviation. In addition to these metrics, we look at the percentage of blocks with price changes. The optimization was carried out over multiple scenarios, with each using either SFpark or plausible price change/range combinations with the latter scenarios providing insight into feasible policy prescriptions. As shown in the table, improvements were observed with regards to the performance metrics when the predicted results were compared to the actual figures. In addition, the optimization facilitated modifying parking rates for a higher number of blocks – 53% compared to 37% observed for SFpark. Apart from the aforementioned, there is also an improvement in the number of blocks that deviate from the desired occupancy threshold – from the approximately 60% observed for SFpark’s approach to 50% for the optimization using the price range and changes specified by SFpark. It also relevant to mention that the magnitude of changes with these metrics might have been muted given the convergence observed across the adjustment period. Expressed differently, assuming we have enough data points, the impact of the optimization would have been greater at adjustment period “t-k” compared to “t” where $k \geq 1$.

Table 3-2: Comparison across key metrics of performance

	Deviated Blocks (%)	Cumulative Deviation	Blocks with price changes (%)
Actual Outcome	59%	946	37%
Predicted - SFpark Values			
GA - Continuous	50%	769	55%
OptQuest - \$0.25 price change	50%	769	53%
Predicted - Plausible Policy Values			
GA - Continuous	49%	765	50%
OptQuest - \$0.25 price change	48%	763	52%

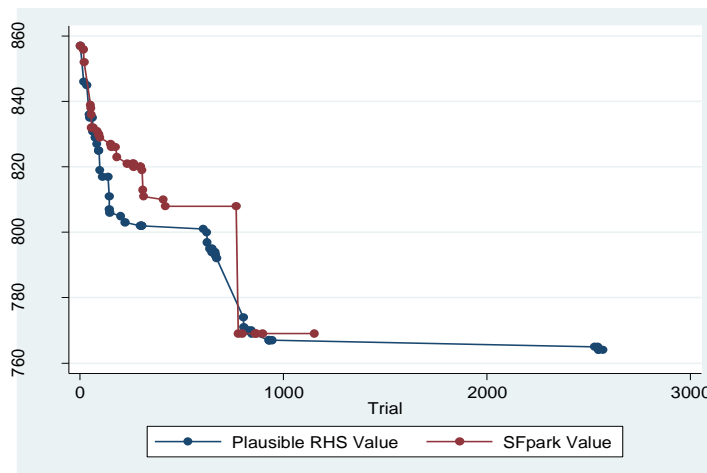
It bears mentioning that the predicted results underestimate the magnitude of the deviation from the desired 60 to 80% threshold for each block. The 16th adjustment period cumulative deviation, calculated by summing across the observations was 946 while the predicted cumulative deviation value was 857. The log transformation provide a plausible reason for these results – the log form essentially reduced the influence of outlier data points. In addition, the underestimation may be as a result of biased coefficient estimates arising from omitted variables or measurement errors for some of the regressors. The other assumptions violation or misspecification issues – homoscedastic assumption and the inclusion of irrelevant variable lead to efficiency losses – higher standard errors. Rarely is it the case that they lead to biased parameter estimates.

Apart from mimicking the price range and price changes specified by SFpark, the optimization was also carried out for simulated policy scenarios, using plausible price ranges and changes. In addition to changing prices by a quarter, discrete changes to parking rates were allowed in 1 and 5cent increment. Prices were also allowed to vary purely in a continuous form and the genetic algorithm was used to optimize this form of variation though the outcome was reported to the nearest integer. Virtually no difference was observed with regards to the figures of merit using these different price changes -the OptQuest engine yielded the same finding irrespective of the size of the parking rate change. Marginal changes were observed when the results of the GA are compared to the OptQuest's.

The optimizations highlight the imperativeness of moving towards a predictive regime that allows for planning and proactively managing the parking program compared to a purely reactive approach based on the prevailing parking occupancy. Even more important are the policy related scenarios. The scenarios enrich the analysis by providing an advocacy tool that the agency could use in making its case before City Hall and the public. In addition, insights are provided with regards to key policy levers that can point SFMTA in the right direction with regards to the most effective options available to it in furthering the agenda of the parking program.

Although it could be argued that the gains shown in Table 3-2 are modest, relative to the SFpark values, it should be mentioned that only subtle changes – the price range was changed from $-.5/$.25$ to $-.5/$.5$ and the acceptable parking rate was capped at $\$6.25$ compared to the prevailing $\$6$. No changes were made to the $\$.25$ floor price. Finally, we have provided detailed information that documents the progress log of the optimization for both SFpark values and the range specified above using a 25¢ change in price as shown in Figure 3-3.

Figure 3-3: Progress log of the $\$.25$ price change without the revenue constraint



Beyond the parking occupancy rate objective, revenue was also taken into consideration. We have chosen to reflect revenue as a soft constraint that could be violated but with penalties attached to the objective function. This approach allows us to model revenue as a secondary goal whose priority could easily be modified by the weight attached to the penalty function. The target goal is to maintain a revenue neutral price modification from one adjustment period to the other. On the one hand, this is driven in part by the fact that though the agency has flexibility with price given the price inelastic nature of parking demand, the public needs to be reasonably assured that the program is not being used as a cash cow. On the other, the public programs that it helps finance, for example, the city's public transit system that it helps subsidize, are not scaled down because of revenue loss.

Figure 3-4: Progress log of the \$.25 price change with revenue modelled as a soft constraint

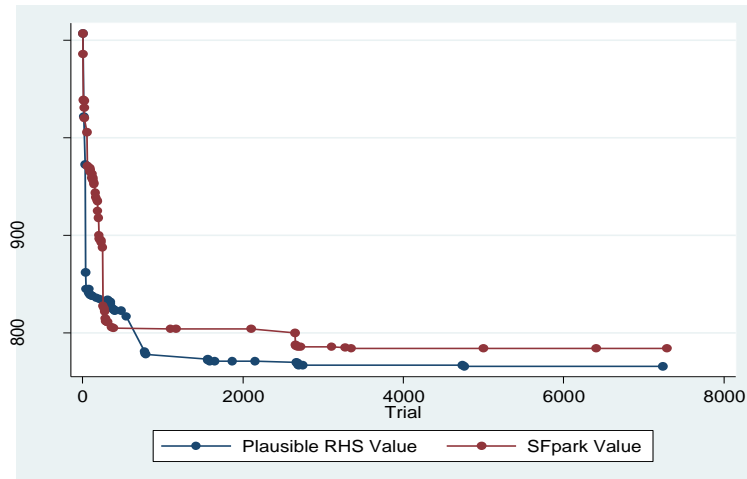


Figure 3-4 provides the detailed progress log of the optimization process with revenue reflected as a soft constraint with penalty added to the objective function when violated. Both upward and downward deviations from the revenue neutral

function are penalized thus addressing the concerns raised in the previous paragraph. To guarantee commensurability, percentage deviations from the target goals could be used or the dollar deviation could be rescaled, given some input from SFMTA before adding it to the objective function. We have chosen the latter using a hundred factor and obtain penalties ranging from 0.001 to 144.5. It must be the case that the optimized parking occupancy objective value is superior (lower) for a non-revenue constrained optimization compared to a revenue constrained one if the *ceteris paribus* assumption holds. When comparing across the solution set, it is obvious that a 763:95.4 occupancy/revenue penalty combination dominates a 763:144.5 solution option. A choice between a 774:0.45 and a 765:81.2 occupancy/revenue penalty tuple is however subjective.

We also looked for evidence of localized pattern of relatively high or low parking occupancy rates across the parking periods. The rationale for doing this is to ascertain the presence or lack thereof, of differences in average parking occupancy, which may provide a case for different priority structure. In doing this, we employed a Hirshmann-Herfindahl index (HHI), a measure similar to the participation ratio concept, to ascertain if there is any validity to this concern. An index value of 0.207- a 21% concentration ratio was obtained with the contribution across the six parking periods ranging from a low of 0.001

to 0.152. The proportional share could be used directly as weights for the next adjustment period optimization, effectively, getting rid of the need to use subjective weights.

3.7 ASSOCIATED CAVEATS AND POLICY IMPLICATIONS

The enriched information obtained from the optimized results provide invaluable policy insights not only in terms of present operational issues but more importantly by how SFMTA can proactively manage parking in the City of San Francisco. As a planning tool, the two-stage model will enable SFMTA to monitor the progress made with regards to the agency's core objective of parking occupancy thresholds, assess the sensitivity of results to changing operational guidelines and use these findings as an advocacy tool before the public or city council. Even discounting of the improvements from using a fitted linear prediction line, the optimization does appreciably better compared to SFpark's approach of modifying parking rates based solely on the latest parking occupancy figures. Excluding the reduction from the fitted prediction, there is an improvement of approximately 10% in systems performance when evaluated using the cumulative parking occupancy deviation across the blocks.

These findings however must be put in context. A key challenge of the present study is addressing assumption violations for the regression analysis. Millard Ball et al (50), (51) addressed endogeneity issues, albeit in a limited manner by focusing solely on the simultaneity between parking rate and parking demand. It turns out that the scope of possible endogeneity problems that the SFpark data may be exposed to is much broader than simultaneity. For one, omitted variables, such as event data that is not reflected in our regression analysis may lead to endogeneity problems. Secondly, measurement error, particularly in the explanatory variable such as the lag form of parking occupancy may lead to assumption violations that could trigger endogeneity concerns. The payment compliance issue is particularly relevant in this context given that parking occupancy is inferred from parking meters.

Apart from the aforementioned, our findings on the impact of flexible price changes on the performance metric seem counterintuitive. Only marginal improvements, where they exist, were observed. In actuality, the GA solution fares worse compared to any of the results obtained using the OptQuest method. This result suggests that having more flexibility on the price change gradation may not necessarily improve system performance. Running the policy scenarios with relaxed price modifications and range values however offers a richer solution space although the improvement to the performance metrics, relative to SFpark values, were only minimal. This is only to be expected given the token changes made to the parking rates and the inelastic nature of parking demand. This finding buttresses earlier observations made by Millard Ball et al (50) that made the case that higher magnitude price changes are needed to effect the changes desired. Apart from this observation, SFMTA planning staff (in conversation with the authors) voiced concerns on the optimal price pain points that will get patrons to modify their parking demand behavior. The \$6/hr. price ceiling was deemed not sufficiently high enough to induce a change in parking demand preferences. This is even more relevant given that the median income in the San Francisco area is arguably the highest in the country.

The fitted prediction model has the effect of minimizing the impact of outliers for reasons provided in the results section. As a result of this, our analysis effectively winsorizes the actual parking occupancy data. This is of concern given that exceptional situations, such as event overlays, are specifically the situations that the program is meant to address. Even assuming a pricing regime where price ceilings are not imposed, very little could be done to address these situations except the pricing strategy transitions from using average occupancy as a yardstick to marginal occupancy. A more granular parking period and adjustment period may be minimally effective assuming multiple price elasticity measures exist for each block time band. The optimization could subsequently be carried out for the extremes of the distribution.

Following up on average parking occupancy as a defective yardstick, Millard Ball et al (50) provide an explanation of the relationship between average occupancy and the probability

of finding a parking space. They pointed out that the marginal occupancy, and by extension, the probability of finding a block full, is influenced by the time window over which the occupancy is averaged. Borrowing from Jensen's inequality, they showed that the longer the duration, the higher the probability of a patron observing a full block. Thus, we may effectively be picking up more instances of subpar systems performance for a three month adjustment period compared to a six week one even though our analysis will not flag this. Our regression analysis does not address differences in the length of the adjustment period; essentially putting them all on an even keel.

3.8 CONCLUSION

We have demonstrated that the two-stage panel data optimization approach presents us a panoramic view of SFpark's operations and facilitates the exploration of different policy-driven operational scenarios – thus allowing for the appraisal of both current and potential SFpark's operational activities relative to the key performance metrics. This explorative process is particularly relevant for planning purposes in that it affords the use of a two-way street to iteratively examine the feasible space for goals and the attendant policy related constraints that could be modified to ascertain how relaxing them improves the program's objectives.

Apart from the aforementioned, we have generated plausible price elasticities of parking demand – the very first time such estimates will be derived using panel data methods in any empirical work. More importantly, the specification of the regression model was robust enough to reflect the spatio-temporal variation in parking demand. These estimates were subsequently used, via an optimization platform, to generate parking occupancy rate tuples that show approximately 10% improvement to systems performance when evaluated relative to SFpark's approach.

Having said this, we would like to reiterate that our modeling approach is not without caveats. These caveats stem from assumption violations not explicitly addressed by the

panel data methods. For one, the explanatory variables used in the regression model are only a subset of the relevant regressors. We have no information on cross elasticity of parking demand; neither are event data reflected in the analysis. Secondly, most of the occupancy data, post the pilot phase were obtained from parking meters. To the degree to which payment compliance is not uniform across space and time, errors with parking occupancy may occur across the observations in a non-random manner. Several examples could be provided as to how this is possible but it is often the case that some blocks are monitored more frequently compared to others. These issues present potential areas for further studies.

CHAPTER 4. ADDRESSING PARKING CHALLENGES IN DOWNTOWN PITTSBURGH

4.1 INTRODUCTION

The past decade has witnessed the resurgence of interest in cities. US Census data, show that Americans are increasingly choosing to move to cities with the wave in urban growth concentrated in and around city centers especially in the West and the South (86). This trend in increased urbanization is not peculiar to the United States (US). The United Nations Population Fund (UNFPA) observed that the year 2008 is the first time in history when more than half of all humans will be living in urban area and that the absolute number is expected to rise to 5 billion by 2030. It is especially noteworthy that within a span of 50 years, there has been a 20% increase in the global urban population - from 34% in 1960 to an estimated 54% in 2014 (87), (88).

However, as a result of weak finances, an appreciable number of cities have limped along for decades, delaying much needed investments in critical infrastructure. Urban infrastructure deficits are further exacerbated in cities' central business districts (CBDs) given the increased concentration of workers and visitors. The City of Pittsburgh provides a canonical example in this regard. Less than 25% of the 288,000 individuals who work in Pittsburgh reside within the city. In addition, the compact, 0.5 square mile area is becoming increasingly attractive as a residential area and continues to be a mecca for cultural events and sporting activities (89).

Similar issues are being witnessed in an appreciable number of cities across the US. The present work addresses the associated parking problems with an emphasis on Pittsburgh's downtown. Our research team developed ParkPGH, a smart parking information system that uses parking, event and weather data to provide information on the availability of parking within downtown Pittsburgh with the goal of reducing parking search time and search time variability. The development of the product includes a robust needs

assessment, an open source platform, a detailed evaluation component, the use of a modular design, and a predictive algorithm.

The balance of the paper discusses each of these features. Section 4.2, assesses the nature and extent of Pittsburgh's parking problem through a stakeholder analysis and a broader environmental scan. Insights from the needs assessment are used to inform Section 4.3, the systems development section. The systems development section identifies the intervention that we determined would be the most robust in tackling the deficit identified in the needs assessment section. Section 4.4, the decision analytics section, documents the prediction approach and presents findings from the analysis. Section 4.5, the evaluation section, focuses on post-deployment evaluation, and itemizes the management challenges encountered during the project's implementation. It also presents the evaluation framework, the data sources employed, and estimates of the project's impact using key outcome metrics and cost effectiveness measures. Section 4.6, the conclusion section, summarizes the work, provides insights on the value-added by the application and provides suggestions for further research.

4.2 NEEDS ASSESSMENT

Parking space is a premium in downtown Pittsburgh, an area of 0.5 square miles with a workforce strength of approximately 130,000 (90). Apart from the geographical limitations, a number of additional factors explain this situation. As a result of policy measures put in place during the 1990s that were motivated in part by the Pittsburgh Downtown Plan (91) there has been a noticeable decrease in the supply of available parking spaces over the last two decades. In addition, current and proposed developments are anticipated to further reduce the total available number of parking spaces. These developments, and the need to avoid a situation where parking becomes a binding constraint to the economic vitality and growth of the downtown area prompted both the Pittsburgh Cultural Trust (PCT) and the Pittsburgh Downtown Partnership (PDP) to seek solutions to these parking problems.

Addressing the parking problem demands a robust approach that can establish the nature of the problem, design and implement a program intervention to rectify the problem and provide an assessment of the degree to which the problem has been ameliorated. The needs assessment speaks to the first of these requirements. Our approach to the needs assessment involves a gap analysis where the difference between the desired state and the present state determines the need. Stakeholders' expectations are examined for feasibility by reflecting the broader environmental constraints in the analysis. The former is made up solely of primary data while the environmental scan was conducted using secondary data from US Census and document review.

Table 4-1: Stakeholders' Objectives and Measures

Stakeholder	Expectations & Objectives	Measures/Indicators
PCT/PCT Patrons	Decrease in late coming incidence to events	Percentage of patrons arriving after curtain
	Reduction in parking search time	Difference in pre/post average search time
	Reduction in search time variability	Mean parking search time deviation
	Reduced parking stress for patrons	Perception measures through surveys
Garage Owners & Management	Improved management of lease holders	Predicted demand estimates
	Increased flexibility in handling contingencies	Predicted demand estimates
Pittsburgh Downtown Partnership (PDP)	More positive perceptions on parking	Perception measures through surveys
	Reduced cycling time	Average parking search time outcome measure
	Decreased greenhouse gas emissions	Estimated value of congestion reduction

Key stakeholders were identified based on their roles, the resources they control or the responsibilities they have and the relationship they have within the parking community or greater transportation ecosystem. Table 4-1 presents the expectation and objectives of stakeholders. In addition to the detailed conversations with key stakeholders, intercept and online surveys were administered to 736 individuals in order to have an assessment of

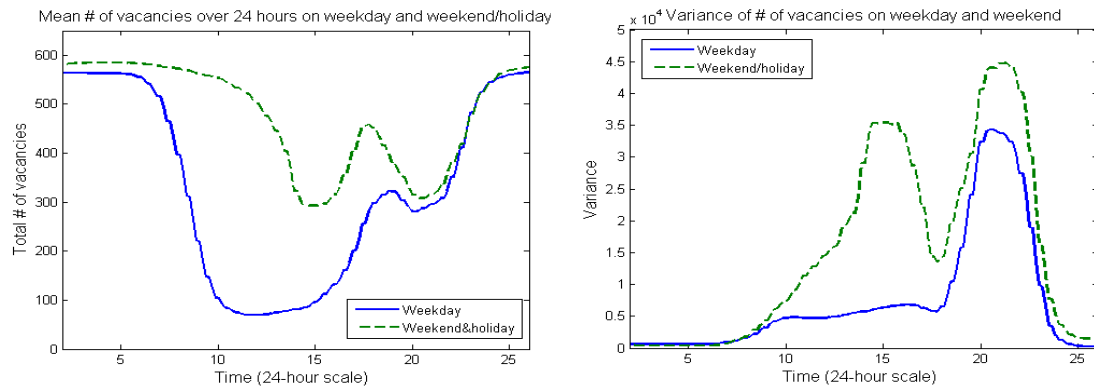
individuals' parking experiences and to gather baseline data on key indicators. Table 4-2 provides the summary results for the baseline survey data.

Table 4-2: Baseline Data on Key Program Objectives

Program Objectives	Data
Parking search time	7.3min
Search time variability	5.4min
Late coming incidence	27.0%
Parking perception (% indicates respondents without a positive response)	
Parking satisfaction	25.7%
Ease of finding a parking space	22.4%
Overall parking experience	22.7%

From the combined primary data, we were able to obtain key insights as to the state of parking within downtown Pittsburgh. We found that the problem is less an absolute shortage of parking spaces but more of a limited parking availability in close proximity to individuals' destinations. Secondly, there were numerous complaints about little turnover in parking spaces – a situation attributed to employees parking in prime parking spots. In addition, the need to emphasize off-street parking became clear given that within the downtown area only 13% of all available parking is on-street. Finally, there was a consensus on the need to focus on commuters, a viewpoint that kept resurfacing throughout the semi-structured interview sessions.

Figure 4-1: Theatre Square Garage Parking Vacancies Mean (left) and Variance



In light of these findings, the parking and travel demand pattern of commuters merits a careful look. Of the 282,841 individuals who work in the city, 211,185 live outside the city and only 55,483 live in the city who are non-city workers. This translates to a daytime population of 463,000 on any workday. This net gain of 155,000 is as a result of workers commuting to the city and is equivalent to more than 50% increase in the city's overall population. Figure 4-1, obtained from a representative parking garage, illustrates parking demand using the average available parking spaces for weekdays and weekend-holidays and their corresponding variances. The huge drop in the number of available spaces observed between 10 am and 3 pm on weekdays is considered to be work-related given that the number of spaces is relatively stable – i.e. low variance. In contrast, we consider the drops around 3 pm and close to 8 pm on weekends to be event-driven because the number of parking spaces fluctuates greatly (i.e. high variance) depending on event occurrences.

To assess parking needs it is important to understand the means of transportation and vehicle occupancy for commuters, both city and non-city residents. Using the US Census Public Use Micro-data Sample dataset (41), we generated population estimates of commuters' transportation modes and their associated margins of error at the 90% confidence interval level. The estimates were calculated from 3 years (2009-2011) of US Census Public Use Micro-data Sample (PUMS) population records. The analysis covered 10 public use micro-data areas (PUMAs) with each PUMA boundary drawn based on a population size of approximately 100,000 residents. Two types of weights were employed:

person weight and replicate weights. The person weight is required for the point estimates and both person weight and the replicate weights are needed to calculate the standard errors.

We employed the Fay's variant of the Balanced Repeated Replication (BRR) method in calculating the standard errors (42). Fay's approach, called the Modified Half Sample (MHS) improves on the BRR by addressing the problem of perturbed weights and decreased sample size using an adjustment factor called the Fay coefficient. This coefficient was set to 0.5 for the PUMS data.

Building on the sampling variance, \hat{V} where: $\hat{V} = \frac{\sum_{i=1}^N (X_i - X)^2}{N}$

Fay's MHS variance equals: $\hat{V}_{mhs} = \frac{1}{(1-m)^2} \hat{V} = \frac{1}{(1-m)^2} \frac{\sum_{r=1}^N (X_r - X)^2}{N}$

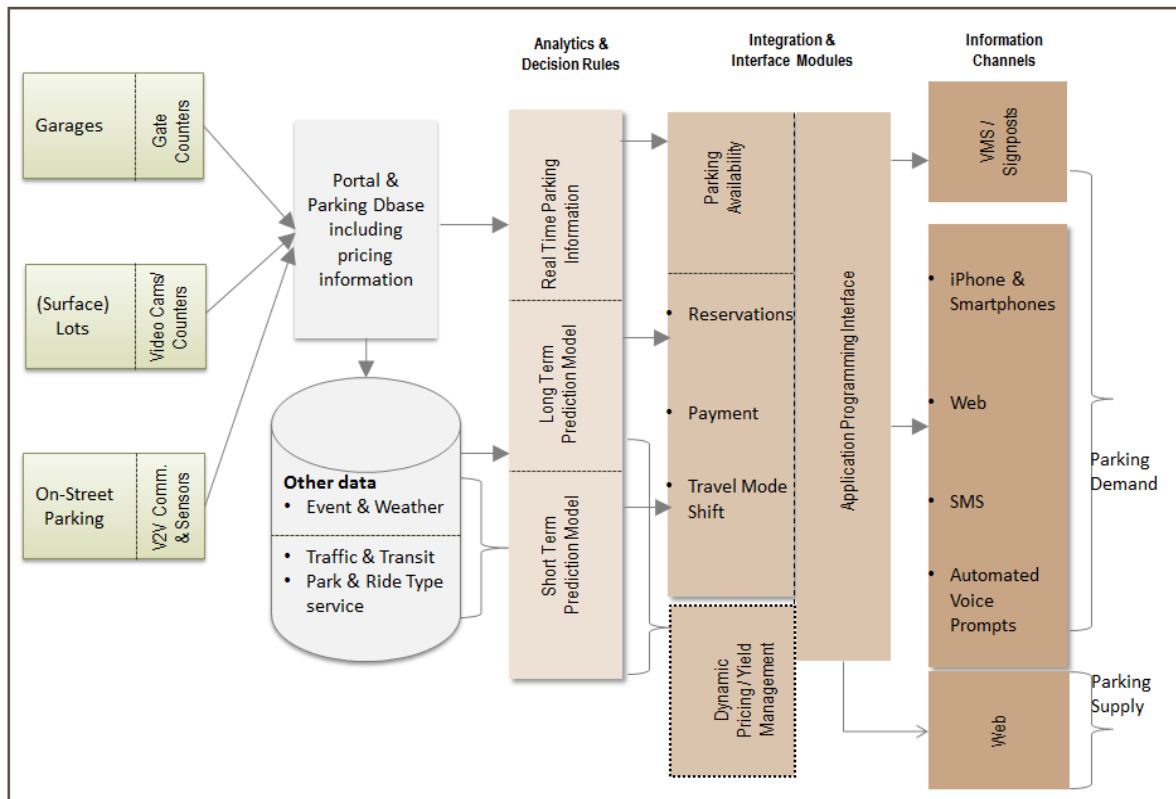
where m is the Fay's coefficient, X_r the replicate estimate and X , the full sample estimate. Since the PUMS person records file has $N = 80$ replicates and $m = 0.5$, the expression above reduces to:

$$\hat{V}_{mhs} = \frac{\sum_{r=1}^{80} (X_r - X)^2}{20}$$

A Stata code was used to implement the procedure that generated the point estimates and the associated margins of error. Of the total number of employees who live in Pittsburgh, 90,000 individuals commute to work either by car, truck or van, a figure equivalent to 64% of all city resident workers. A higher proportionate figure was obtained for Allegheny County, excluding the City of Pittsburgh. Using these results, it is estimated that the demand for parking spaces by workers in the City of Pittsburgh on any working day is 214,000. A 5% reduction in the absolute number of individuals who drive unaccompanied will translate to a reduction in parking demand of more than 5,000 spaces.

Apart from the travel mode patterns of commuters, the analysis enabled the research team take into consideration the systems' constraints in addressing the feasibility of achieving stakeholders' objectives. These insights were subsequently used in determining the most robust program intervention. They provided the rationale for framing the research question – *can a demand-side intervention, specifically providing real-time and predictive information address the (perceived) lack of parking availability in downtown Pittsburgh?* The rest of the study, starting with the schema in Figure 4-2, is geared towards answering this question.

Figure 4-2: Smart Parking Application Schema



4.3 SYSTEMS DEVELOPMENT

The program intervention is a pilot product called “ParkPGH,” a smart parking application providing real time and predictive information on the availability of off-street parking within downtown Pittsburgh. The app falls within the realm of initiatives classified into two broad sections: parking guidance systems, and real-time/prediction information.

Parking guidance systems (PGS) use variable message signs (VMS) to inform drivers about available parking spaces with the information being relayed through a series of channels and oftentimes, integrated with traffic information system. Much of the literature on parking guidance systems is concerned with transit and park-and-ride lots (92) with examples of the system described in Orski (93). Another stream of work on PGS explores their use inside of parking facilities (94). ParkPGH is distinct from the parking guidance system literature in two ways: ParkPGH is not coupled with transit, and it does not employ VMS. The parking availability information available only through mobile devices, interactive voice response (IVR) and the Internet.

Real-time and prediction information systems examine the display and use of information for finding parking spots. Information on parking availability is either provided during a trip or before the trip begins. The works of Caliskan et al. (95) and Teng et al. (96) are examples of systems that provide parking prediction models based on information exchanged between wirelessly connected vehicles for use *during* a trip. Other variations include the use of agent-based modeling in simulating drivers' behavior (97); an allocation model by Teodorovic and Lucic (98) that accepts parking requests via an inventory control system; utilizing sensors to provide real time parking data (99) and predicting the number of available parking spaces where the parking requests are routed to a number of competing parking facilities (100).

Our approach is to employ a predictive model that utilizes historical garage occupancy rates with detailed information on events and weather conditions. The approach improves on similar works that have been bedeviled with poor forecasts or the inability to determine the ideal number of neural units (101). The prediction model utilized for the present study is an event-based parking prediction model for use *before* a trip begins along with historical parking and event data to predict future parking availability. These predictions have been shown to reduce the uncertainty often related to parking in downtown areas and central business districts (102).

The application was implemented by combining systems development and integration with a parking prediction algorithm as shown in the schema. The system development and integration module collects real-time parking information from both public and privately held parking garages. This was made possible through the use of a web application programming interface (API) and infrastructure that collects, validates, and stores parking information in real time. The system integration also includes the development of an iPhone application, text message gateway, and an API that provides third party developers access to ParkPGH data.

The prediction model uses as inputs historical parking and event data that have occurred downtown and provides estimates of the available parking spaces for each garage. The prediction model is trained on a historical parking data set. This dual-pronged technological innovation was deployed through a pilot program that monitors eight parking garages totaling 5000 parking spaces, representing about 20% of the total parking supply in downtown Pittsburgh. Parking information is updated every minute and delivered through multiple channels that include websites, iPhone app, SMS text, voice and a mobile version of the website that provides the same information as the traditional website but is optimized for mobile devices such as Blackberries and Android phones. We have embraced the traffic sign colors in providing information to patrons looking for parking spaces. The green, yellow and red color coding is complemented with a numerical figure that shows the available number of parking spots, except in cases where the garage is deemed full or close to full capacity. A snapshot of the website showing destinations within the Cultural District, garages and the available spaces is provided in Figure 4-3.

Figure 4-3: Pittsburgh's Downtown Map with available parking spaces



Clicking on any of the garages reveals more information, including the facility address, map and pricing. In addition to parking garage information, popular destinations are displayed so that visitors can locate their targeted destination and find the closest available parking. A mobile version of the website, m.parkpgh.org, provides the same information as the traditional website but optimized for mobile devices, focusing on the Blackberry and iPhone. The exact number of available parking spots is not provided when the garage is deemed nearly full.

Figure 4-4: ParkPGH iPhone parking prediction output

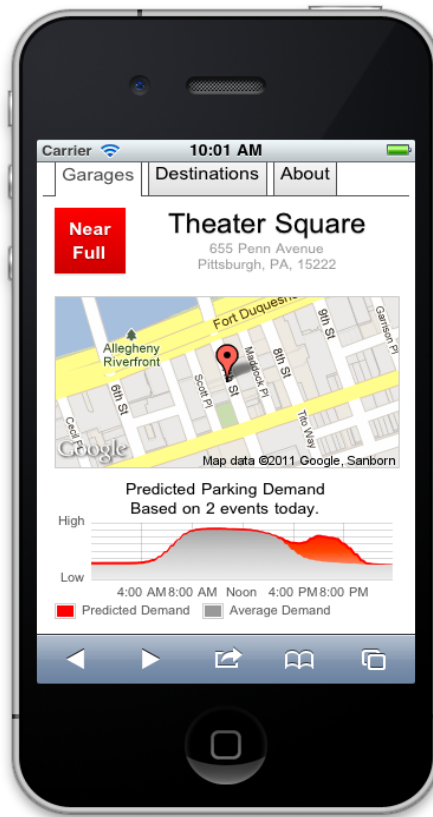


Figure 4-4 is a screenshot of the ParkPGH iPhone application. In the pictured scenario a popular garage, Theater Square, is currently designated as “Near Full.” In addition to this real-time information, a plot of predicted parking demand is provided on the lower half of the screen. The predicted parking demand plot shows the average or baseline parking demand for the garage based on historical data. Additionally, the demand exceeding the average is also provided. In this scenario, the excess demand is predicted based on two events occurring near Theater Square garage that influence future parking availability.

4.4 DECISION ANALYTICS

This section documents the steps involved in predicting parking availability, including information on the key assumptions. The prediction models are presented in two parts. First, we focus on predicting the number of available parking spaces at a given time, based on a set of events and weather data using neural network-based predictors. Given that the driver would like to know whether the garage is full, we provide a robust approach that reduces the possibility of Type II errors—a situation that occurs when the garage is full, but the application shows that parking spots are available. This motivates the classification methods presented in the second part of this section. A range of classification and prediction methods, including logistic regression, naïve Bayes classifiers, classification and regression trees (CART), and a neural network, complements the continuous prediction methods.

The dependent variable is the number of available parking spaces in the garage. The key assumption is that the time-varying arrival rate is a function of the events occurring in the garage's vicinity. We can observe the arrivals at the parking garage when it is not full by collecting the electronic counts from the garage entrance-gate system. However, time of day is available and critical for our model. The design rationale behind this decision is that the real-time information provided by ParkPGH is useful for immediate and short-term purposes while the prediction algorithm is useful for longer-term parking predictions.

4.4.1 Data Description

We describe the model for the Theater Square garage, the analysis for other garages is similar. The training set of parking data includes the number of available parking spaces for each 10-minute interval for 18 hours each day from November 9, 2008 to July 10, 2010. The prediction model estimates parking vacancy based on the variables itemized in Table 4-3 below. Events are categorized as (1) morning (before noon), (2) day (12:10 pm to 4 pm), or (3) night (after 4:10 pm).

Table 4-3: Variable list for the prediction model and their associated definitions

Variable	Definition
Categorical independent	
Theater Dummies: <i>ben2</i> , <i>ben3</i> , <i>byh2</i> , <i>byh3</i> , <i>or2</i> , <i>or3</i> , <i>hnz2</i> , <i>hnz3</i>	Dummies for theater events at the Benedum Center (<i>ben*</i>); Byham Theater (<i>byh*</i>); O'Reilly Theater (<i>or*</i>); and Heinz Hall (<i>hnz*</i>)
Sport Dummies: <i>pir2</i> , <i>pir3</i> , <i>hnzf2</i> , <i>hnzf3</i> , <i>pen2</i> , <i>pen3</i> , <i>stl2</i> , <i>stl3</i>	Dummies for sporting events - Pirates (<i>pir*</i>); Heinz Field (<i>hnzf*</i>); Penguins (<i>pen*</i>); and Steelers (<i>stl*</i>)
Day of the week: <i>Sunday</i> , <i>Monday</i> , <i>Tuesday</i> , <i>Wednesday</i> , <i>Thursday</i> , <i>Friday</i> , <i>Saturday</i> , <i>holiday</i>	Dummies for day of the week
Time of the day: <i>period</i>	Period of the day measured in 10-minute increments
Numeric independent	
Weather: <i>snow</i> , <i>rain</i>	Snow and rainfall measured in inches
Numeric dependent	
<i>Availability</i>	Number of available parking spots

4.4.2 Neural Network

During preliminary analysis, a multiple linear regression analysis proved to have low predictive power. Thus, we sought better prediction models using neural networks. The forms of neural network approaches employed include both the generalized regression neural network (GRNN) and the multilayer feed-forward network (MLFN). For the GRNN (103), how well the training sample does is provided in the equations below where d_i which is the difference between the training sample and the prediction point and σ , the standard deviation. Table 4-4 provides summary findings from the GRNN analysis that provided the best search.

$$Y(x) = \frac{\sum_{i=1}^n Y_i \exp(-d_i^2/2\sigma^2)}{\sum_{i=1}^n \exp(-d_i^2/2\sigma^2)}$$

and;

$$d_i^2 = (x - x_i)^T(x - x_i)$$

Table 4-4: Summary of the Neural Network Analysis

Summary	
Net information	
Configurations included in search	GRNN, MLFN: 2 to 6 nodes
best configuration	GRNN numeric predictor
Training	
Number of cases	24,418
Training time	0:56:01
Number of trials	31
Reason stopped	Auto-stopped
% bad predictions (30% tolerance)	9.23%
Root mean square error	56.44
Mean absolute error	21.89
Std. deviation of abs. error	52.02
Testing	
Number of cases	6,105
% bad predictions (30% tolerance)	11.40%
Root mean square error	60.84
Mean absolute error	25.28
Std. deviation of abs. error	55.33

The best net search from the architectures utilized is the GRNN numeric predictor. We carried out 31 trials for the training set. Measures of how close the predicted values are to the eventual outcomes are provided in the form of root mean square error (RMSE), mean absolute error (MAE), and standard deviation of the absolute error values for both training and testing trials. In addition, we provide measures for the percentage of bad predictions. We obtained a bad prediction value of 11.4 percent for the testing case, an indication that, on average, approximately 1 out of every 10 predictions will be wrong.

Figure 4-5: Bar Chart showing relative variable impact on GRNN predictions

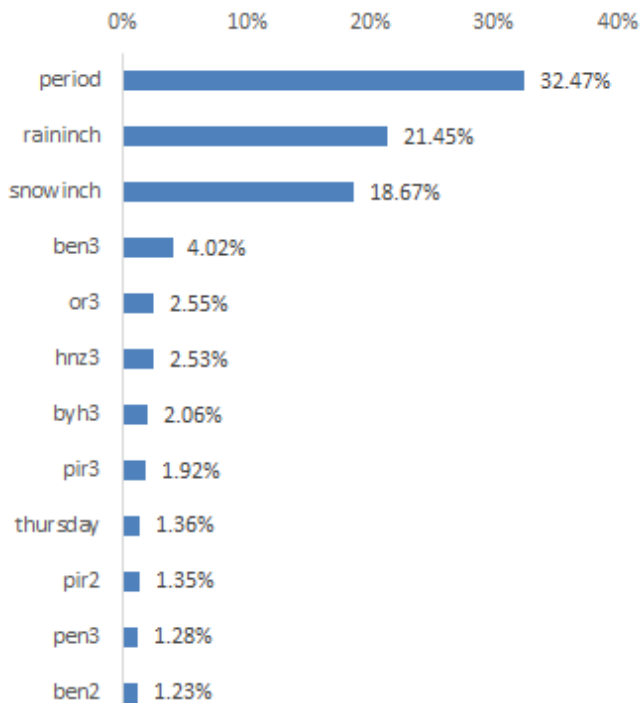


Figure 4-5 provides measures of the sensitivity of the GRNN predictions to changes in the nine most influential regressors. The cumulative value of the impact of all the regressors is normalized to 1, and the proportionate contribution of each regressor is provided as relative variable impact value. The explanatory variable that represents the specific time of the day in 10-minute increments has the most impact, followed by variables that capture the weather situation.

4.4.3 Measures of Predictive Accuracy

Table 4-5 provides a comparative analysis of predicted accuracy measures using RMSE values. We trained and tested six nets to identify the best one. The configuration of the MLFN includes the use of multiple nodes, ranging from two to six for the hidden layers. The MLFN (104) approach utilized a back propagation algorithm where the steepest descent minimization method is used. The weights and thresholds are adjusted using the expression below where ϖ_{ij} is the weight coefficient between the i th and the j th neuron, ϑ_i is the threshold coefficient, λ the rate of learning and E represents the sum of squared differences. As shown in Table 4-5, the GRNN performed best of all the nets we trained and tested.

and:

$$\omega_{ij}^{(k+1)} = \omega_{ij}^k - \lambda \left(\frac{\partial E}{\partial \omega_{ij}} \right)$$

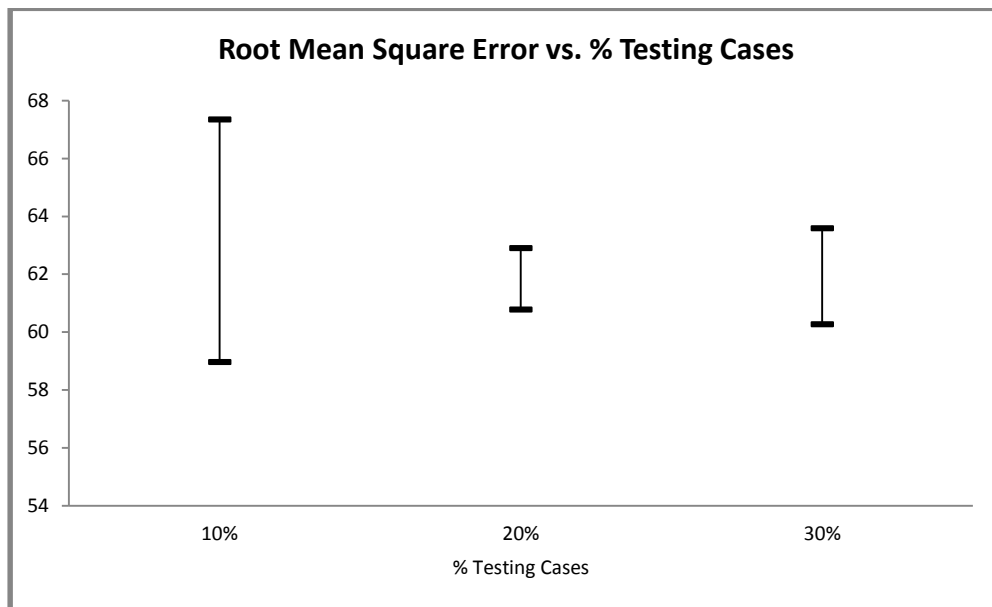
$$\mathcal{G}_i^{(k+1)} = \mathcal{G}_i^k - \lambda \left(\frac{\partial E}{\partial \mathcal{G}_i} \right)$$

Table 4-5: Net Search Result using RMSE measures

Net Explored	Root Mean Square Error
Linear predictor	170.68
GRNN	60.84
MLFN 2 nodes	144.62
MLFN 3 nodes	136.96
MLFN 4 nodes	139.81
MLFN 5 nodes	136.85
MLFN 6 nodes	138.72

The degree of improvement in the accuracy of predictions is particularly noteworthy when we compare the GRNN RMSE value to that of the linear predictor or any of the MLFN. Finally, we carried out sensitivity analyses to determine the reliability of the predictions for each testing case. Results from the sensitivity testing were invaluable in estimating reliability measures as a result of changing the size of the subset of data used for testing and in ascertaining the quality of the predicted values. The ideal percentage testing case is 20% and the RMSE ranges from a low of 60.78 to a high of 62.90 for this threshold as shown in Figure 4-6.

Figure 4-6: RMSE values for different sizes of the testing data set

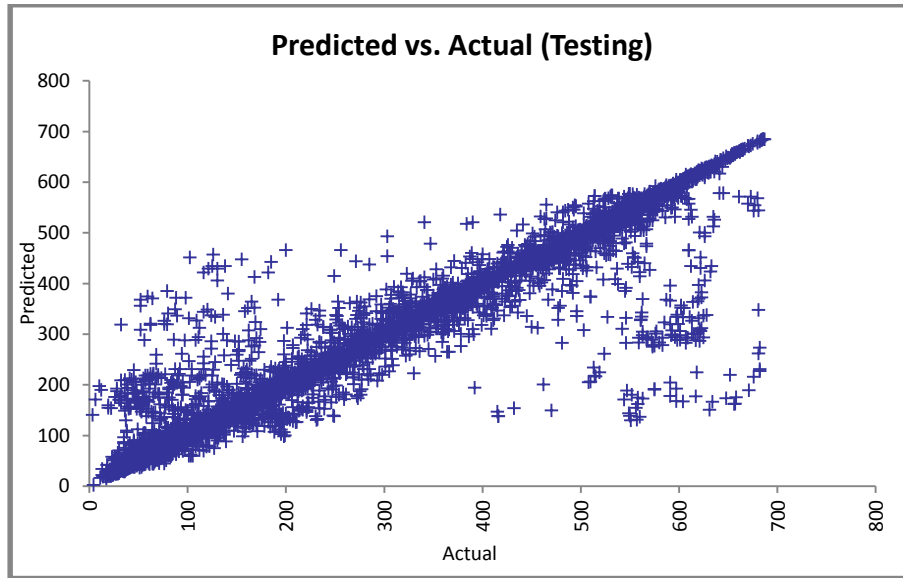


4.4.4 Rationale for using Predictive Classification Methods

Figure 4-7 provides the rationale for the use of classifiers. The 45° line represents the locus of points where the predicted and the actual values are of the same magnitude. The figure shows that the GRNN net provides a good fit in situations where the utilization of the garage parking spots is neither low nor high. However, the magnitude of the errors seem to increase at the extremes. The predicted values systematically overestimated the actual values for low values and consistently underestimated the actual values for high values. This

explains our rationale for switching to a categorical dependent variable, especially at high-capacity utilization where users of the application may be extremely sensitive to Type II errors—a false negative (not full) when it is indeed full. To be conservative, we selected a threshold level of 85% for dichotomizing the dependent variable.

Figure 4-7: Scatter plot comparing predicted values to actual parking utilization



4.4.5 Predictive Classification Methods

As a complement to the continuous variable prediction models, we explored classification methods based on machine learning. Using a binary dependent variable—full or not full. The garage is considered full when the availability of parking spaces is less than 15 percent of capacity.

We use the same data set for the availability prediction, which contains 36,949 observations of not full, representing 94 percent of the observations, and 2,409 observations of full. The independent variables remain the same as in the continuous variable prediction models. We report the prediction results of three classifiers, naïve Bayes, logistic regression and classification and regression tree (CART).

The naïve Bayes classifier (NBC) predicts the class, (full or not full), using an approach similar to an empirically driven maximum-likelihood estimator (105). Given the independence assumption, the Baye's rule:

$$P(Y = y_k | X_1, \dots, X_n) = \frac{P(Y=y_k)P(X_1, \dots, X_n | Y=y_k)}{\sum_j P(Y=y_j)P(X_1, \dots, X_n | Y=y_j)}$$

reduces to:

$$P(Y = y_k | X_1, \dots, X_n) = \frac{P(Y=y_k) \prod_i P(X_i, \dots, X_n | Y=y_k)}{\sum_j P(Y=y_j) \prod_i P(X_i, \dots, X_n | Y=y_j)}$$

The expression above is a tractable one given that the denominator does not depend on y_k . Just as we have for the NBC, the Logistic Regression approach (105) assumes a parametric distribution and the regression coefficients are estimated via maximum likelihood method. Given a learning function $f: X \rightarrow Y$ and a binary output variable for the dependent variable of class {full, not full}, we want to model the conditional probability of Y as a function of x . Assume $1 \equiv \text{"Full"}$ and $0 \equiv \text{"Not Full"}$, the parametric model utilized is of the form:

$$P(Y = 0 | X) = \frac{1}{1 + \exp(\beta_0 + \sum_{i=1}^n \beta_i X_i)}$$

and;

$$P(Y = 1 | X) = \frac{\exp(\beta_0 + \sum_{i=1}^n \beta_i X_i)}{1 + \exp(\beta_0 + \sum_{i=1}^n \beta_i X_i)}$$

The classification approach is to assign the value $y_k \in \{\text{full, not full}\}$ that maximizes $P(Y=y_k | \mathbf{X})$ using the expression below:

$$y_k = \begin{cases} full & \text{for } 0 < \beta_0 + \sum_{i=1}^n \beta_i X_i \\ not\ full & \text{otherwise} \end{cases}$$

However, in contrast to the NBC and the logistic regression, the classification and regression tree (CART) method is a nonparametric method, which employs a binary tree

for classification and prediction (106). Assuming X is the variable and S , a subset of the possible realizations of X , the search process is specified thus:

Start at the root node.

For each X , find the set S that minimizes the sum of the node impurities in the two child nodes and choose the split $\{X^ \in S^*\}$ that gives the minimum overall X and S .*

If a stopping criterion is reached, exit. Otherwise, apply step 2 to each child node in turn.

The classification methods were implemented using tenfold cross validation. Table 4-6 shows the results—average RMSE, *precision*, and *recall*. We note that the RMSE values obtained for the classification-based prediction methods are of much smaller magnitude compared to the numeric predictor because of the binary nature of the dependent variable.

Table 4-6: Summary result of the classification based prediction methods

	10-Fold RMSE	Precision	Recall	Binary outcome
Naïve Bayes	0.252			
		0.951	0.964	Not full
		0.299	0.237	Full
CART	0.129			
		0.984	0.991	Not full
		0.842	0.756	Full
Logistic	0.222			
		0.942	0.995	Not full
		0.449	0.059	Full

Precision is the ratio of the number of true positives divided by the sum of the number of true positives and false positives. Recall is the ratio of the number of true positives divided by the number of true positives and false negatives. As Table 6 shows, the performance of CART is superior to the other measures. This is not surprising given the independence assumption of both the naïve Bayes and the logistic regression. CART does not make this assumption and is free to build trees to exploit any correlation in the feature structure. The

resulting CART has 341 nodes with 171 leaf nodes. Each leaf node corresponds to a unique combination of features or scenarios.

The CART extension to the GRNN numeric predictor provides a robust prediction platform. The hybrid approach mimics the real-time feed that is currently running across multiple channels of the smart-parking application in which the “full” sign is displayed when the parking space utilization goes above a specific level. The frequency of updates to be made to the models is determined by the levels of RMSE, percentage of bad predictions, precision, and recall values associated with the models. These thresholds will be established using weekly live predictions and analyzing the residual values. Input will also be solicited from garage operators regarding the tolerable level of error.

4.5 EVALUATION: POST DEPLOYMENT MEASURES

The post-deployment or ex-post evaluation addresses both the formative and the summative evaluation. The formative evaluation is process driven and provides real time information as ParkPGH is being rolled out. The real time information has been invaluable in providing the development team crucial insight on how receptive the public is to the pilot program and alerts them to areas where users are experiencing difficulties. The summative evaluation focuses primarily on program outcomes and captures the state of parking within the Cultural District after ParkPGH deployment based on indicators that proxy the key objectives identified by stakeholders. In addition, we have provided estimates of program impact using outcome measures and benefit cost analysis (BCA).

4.5.1 Evaluation Approach

The evaluation framework was designed to be robust enough to address the inherent pitfalls when evaluating a program of this nature. Some of these include:

- *Correction for the oversampling of early comers surveyed during PCT events:*
The most constraining issue with the use of in-person surveys has to do with an

oversampling of early-comers to events, as those patrons arriving early were the ones with enough time to complete an in-person survey. This was addressed in two ways. The first was by using an online survey, which included a broad sampling of patrons which encompassed both early-comers and late-comers. The online surveys were administered to patrons using the PCT email list the day after the event. We were also able to selectively identify and survey those patrons who arrived close to the curtain call using time stamped information from their event tickets.

- ***Increasing the number of respondents for the Pre-Deployment phase:*** The Smart Parking application went live during the time baseline data was being collected. To increase the number of observations used for the baseline data, we identified the control group for the post-deployment phase; those individuals who did not know about the existence of the smart parking technology or who were yet to make use of it and treated this cohort as if they belonged to the pre-deployment phase.
- ***Addressing response-shift bias:*** The evaluation framework allows for the examination of the quality of self-reported measures provided by respondents in instances where multiple surveys are administered over a period of time. This response shift bias situation may arise when individuals have rated themselves at one time from one perspective and then change their responses later because these perspectives have changed.

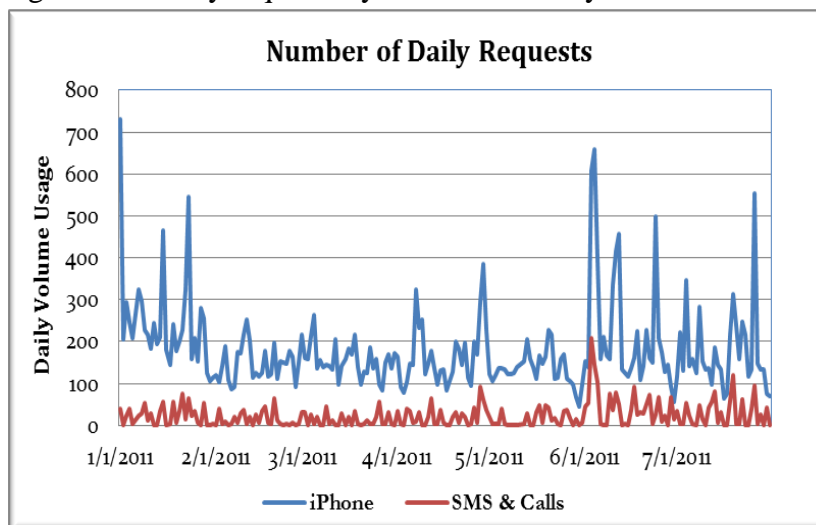
It is imperative to address the possible presence of response shift bias given that the assumption of an invariant measurement scale both before and after the program intervention may not hold (107). The typical approach in ascertaining the effect of a program is to administer identical questionnaires both before and after program implementation. However, for self-reported measures, an inaccurate impact measure may be obtained if the frame of measurement on which the pretest questionnaire is answered is altered when responding to the posttest survey.

We addressed this by utilizing a binary numbering system to identify subsets of the respondent population, for example, 01 (respondent completed only the post deployment survey) and 11 (respondent completed both pre and post deployment surveys). We could now subsequently investigate the response shift bias alluded to earlier. For example, are there systemic differences in the data provided by individuals who completed the outcome survey given that they did (or did not) complete the baseline questionnaire? We investigate this by looking at the respondents for the posttest questionnaire, classifying into two cohorts depending on whether they responded or not to the pretest survey. For each question of interest, we subsequently compare $\frac{\sum_{j=1}^j Q_i}{j} | z = 0$ and $\frac{\sum_{k=1}^k Q_i}{k} | z = 1$ to ascertain if the difference between them is statistically significant where Q_i represents the i th question, j and k represent the respondent population and z assumes a value of 1 if the pretest survey was completed and 0 otherwise. While a comparison of means showed some differences, we could not conclusively say that these differences are statistically significant given the small sample size and the need to control for population differences across the cohorts compared.

4.5.2 Post Deployment Measures

In order to improve upon and ascertain the value added by ParkPGH, a series of indicators were tracked. Count data was used to track output measures that include the weekly usage

Figure 4-8: Daily requests by selected delivery channels



volume for each of the delivery channels used to provide information by ParkPGH. This includes iPhone app, mobile and traditional website usage, number of text messages sent on request, number of unique views, number of automated phone responses, bounce rate and average duration of page views. Figure 4-8 shows the number of daily requests for selected delivery channels between January 1st 2011 and July 31st, 2011. The usage volume is typically higher during the weekdays compared to weekends except when events are scheduled. The noticeable spike in usage on the weekend of June 3rd to the 5th is attributed to the Pittsburgh JazzLive Festival, a weekend of music that includes outdoor stages, visual art shows, International musicians and a JazzLive crawl.

The usage pattern was driven, in part, by a wide range of marketing activities carried out to inform and educate PCT patrons about ParkPGH. Marketing activities targeted to PCT patrons include ticket inserts, email blasts to patron lists, playbill advertisements, website promotions and social media campaigns. Other activities address a broader population by utilizing posters in and around the Cultural District, advertising in regional magazines and jumbotron video advertisement. However, the evaluation of the effectiveness of these activities was not undertaken.

Process measures were also utilized for formative evaluation purposes. Information obtained from these measures was used to make modifications to the smart parking project. Ease of use, difficulties with design and accuracy of the information provided are some of the process related measures tracked. A negative response on the online survey to any of these measures prompts an open-ended question that allowed the respondent to provide detailed information as to the nature of the problem being encountered. Such information was subsequently relayed to the development team. Table 4-7 shows the data obtained for process related measures for a total sample size of 43 respondents. While only 3% of respondents reported having difficulties using ParkPGH, approximately 1 out of every 12 reported experiencing difficulties with the design or with the accuracy of the information provided. Some of the problems users experienced include difficulties telling a

landmark from a parking structure, recorded messages not being clear, the cursor hovering over the icon, and sending a text request too early.

Table 4-7: Process related measures (%)

PROCESS MEASURES	%
<i>Usability (Ease of Use)</i>	
% of respondents with positive response	77.1
% of respondents that are indifferent	20.0
% of respondents with negative response	2.9
<i>Usability (Difficulties with Design)</i>	
% of respondents that experienced difficulties with product design	8.8
<i>Information Accuracy</i>	
% of respondents that said information provided was not accurate	6.8

An appreciable degree of effort was expended on addressing the information accuracy issue, particularly in situations where garages are close to full capacity. Some of the challenges in providing accurate information stem from the fragmented ownership and diverse management structure of the garages that are featured in the pilot phase of the smart parking application. When the project was initially conceptualized, it was thought that there was a uniform method of determining the number of currently available parking spots in the garages, along with a way of determining when the garage could be identified as being “full.” However, each parking garage has its own “culture” of determining how and when to identify the garage as being “full.” Variables that factor into that decision include the number of leased spots to hold open, use of valet parking, the threshold level at which the “full” sign goes up and garages that distinguish between hard and soft full.

This lack of standardization has made it challenging to provide accurate information on parking availability across garages when garage utilization is close to maximum capacity. This issue was addressed in part by providing a web portal for garage managers. The platform allows the documentation of lease management strategies and process issues that shape the idiosyncratic features exhibited by some garages. The information is shared with the software development team with the objective that subsequent modifications to the

smart parking algorithm will reflect these garage specific traits. In addition, we have suppressed the information on the number of parking spaces available when the garage is deemed full or close to maximum capacity. The possible options by which information is relayed to the public were pilot tested to ascertain the ideal level of detail especially when garages are close to full capacity.

Outcome measures that document the impact of ParkPGH are provided in Table 4-8. Approximately one out of every two respondents reported that the application has reduced the time it takes them to find a parking space. The magnitude of the reduction in search time ranges from as little as a minute to more than 6 minutes with individuals reporting a 4-6 minute reduction in search time being in the majority.

Table 4-8: Outcome related measures

DOCUMENTED OUTCOME	%
<i>ParkPGH has made finding parking spaces easier</i>	
% of respondents with positive response	57.2
<i>Specific reduction in search time</i>	
% of respondents reporting a reduction in search time	48.6
% of respondents with 1-3min reduction	17.1
% of respondents with 4-6min reduction	22.9
% of respondents with > 6 min reduction in search time	8.6

4.5.3 Program's Estimated Impact

ParkPGH's estimated impact was calculated over a seven month period (1/1/2011 – 7/31/2011) using evaluation findings and count data from Google Analytics. 271,512 garage enquiries were generated over the seven month period. Of this figure, 63.4% are unique views. Statistics also exist for the bounce rate and average duration of page views. However, we have chosen not to adjust the effective number of requests by these figures given that virtually all the information a driver needs to find a parking spot is accessible on the first page of the app.

This information, coupled with the mean reduction in search time from Table 4-9

translates to a total of 5,746.6 hours saved. From the hours saved, we have estimated ParkPGH's impact as the cumulated savings from the time value of money and the reduced gas expenditure. Key assumptions used in calculating the impact include an average fuel economy of 20 miles per gallon, a \$3.50 cost for a gallon of gas, a downtown cruising speed of 10 miles per hour and a \$20.44 average hourly wage for the City of Pittsburgh (108). The use of the average hourly wage for the time value of trips is premised on the concept of opportunity cost. We have also assumed that the survey results are representative of the population. The calculated estimates are provided in Table 4-9.

Table 4-9: ParkPGH Estimated Impact

<i>Impact Type</i>	<i>Measure</i>	<i>\$/Unit</i>	<i># of Unit</i>	<i>Savings (\$'000)</i>
Time Value	Hour	20.44	5746.6	117.5
Gas Expenditure	Gallon	3.50	2873.3	10.1
Cumulative Savings				127.6

The annual impact is estimated at \$219,000 and the annual potential impact at planned full scale deployment is estimated at \$1.1 million. These figures are premised on a linearity to scale assumption. Secondly, it bears stating that we have not employed an adjustment factor for the number obtained for the unique garage inquiries. The rationale is that figures from the documented outcomes from the surveys were used to estimate the value added by the parking application and we have assumed that the individuals surveyed are representative of the population of ParkPGH users. Thus, it is expected that individuals who ended up not making the trip or were just checking the site out for novelty reasons will not be within the cohort of respondents who reported a reduction in parking search time.

In addition to the estimated impact, a Benefit Cost analysis was calculated using both the NPV and the Benefit Cost ratio (BCR) approach. Cost components taken into consideration include capital costs, operation and replacement costs. The analysis was performed with a 5% discount rate and a 5 year time horizon using $\frac{\sum_{t=0}^4 (B_t - C_t)}{(1+r)^t}$ for the

NPV and $\frac{PV_B}{PV_C}$ for the BCR where PV is the present value, r is the discount rate, t , the horizon in years, B the benefit and C , the cost. The analysis revealed a NPV of \$550,000 and a 2.34 BCR, an indication that for every dollar spent on the ParkPGH project, the society gains approximately \$2.34 in benefits.

4.6 CONCLUSION

This research introduces an innovative demand-side intervention to address parking challenges in downtown Pittsburgh. The intervention, a smart parking application, uses garage occupancy rates with detailed information on events and weather conditions to provide real-time and predictive parking information. The system has been effective in reducing parking search time and in changing the perceptions of patrons about the downtown parking situation. In 2011, more than 300,000 unique inquiries were made of the parking application.

Compared to other smart parking applications, ParkPGH is distinguished by a number of unique and distinct features. For one, a detailed stakeholders' analysis was carried out, which was complemented by a process and summative evaluation. This served multiple purposes; it allowed stakeholders to identify shortcomings in the parking situation within downtown Pittsburgh, and determined the most robust form of program intervention to address the deficits. This comprehensive stakeholders' analysis and the input from end users helped counteract the typically "silo" approach by which parking data is managed, an approach that precludes the exploration of synergies across different garages, IT platforms and ownership/management cultures, and often leads to sub-optimal outcome. In addition, involving stakeholders in the program design phase and having a robust evaluation platform yielded crucial insights on the design of product features and allowed modifications to be made to these features in real time. These provisions were invaluable in making the application user-friendly and effective.

The application's relatively low cost, its ease of retrofitting, modular structure and its open-source platform can enable other cities to lower the costs of implementing and managing similar smart-parking solutions; significantly shorten their learning curves and ensures that transitioning to a fully-fledged integrated parking application could be achieved at a relatively low cost. The design process placed emphasis on creating citywide solutions that span both public and private providers and made the case for a coordinated approach that ensures that the application provides standardized, platform-independent solutions that could be scaled up – an invaluable feature that could promote its buy-in across multiple cities. Apart from the aforementioned, the modular design of ParkPGH makes provision for product enhancements, ensuring that retrofitting can occur with relative ease.

The novelty of the application is the deployment of the first real-time and predictive parking analytics system with input from multiple stakeholders. While the prediction method used in ParkPGH is not new, the environment in which it is deployed is unique and the approaches used in addressing these challenges have been nothing short of innovative. More importantly, of all the components of the smart parking application, the predictive module holds the most promise given its potential influence on commuters' travel and parking demand patterns. Integrating parking information with traffic flow patterns is more effective if the parking information provided is not only real time but also predictive. Having this knowledge can encourage commuters to change the way they travel or how they schedule their trips. Apart from providing information on the demand side, garage operators could use the predictive information to better manage their facilities. These findings underscore the need to prioritize the implementation of a fully functional predictive module that encompasses both long- and short-term predictive modules.

CHAPTER 5. SUMMARY

5.1 SUMMARY OF FINDINGS AND CONTRIBUTIONS

US Census data highlights national trends in increased urbanization with more Americans choosing to move to cities. At the same time, urban areas are dealing with significant problems related to delayed investments in critical infrastructure. In addition, and compounding the problem, higher tiers of government; county, state, federal, with infrastructure responsibilities often pay little to no attention to these assets, which forces the burden back on financially stretched cities. Urban infrastructure deficits are compounded in cities' central business districts (CBDs) given the increased numbers of workers and visitors.

This thesis addresses the problems of this development primarily at it relates to urban parking, using a wide range of datasets that include parking availability from electronic gate counts, sensors, smart meters, events and weather data for specific localities and individuals' commuting pattern using the US Census Public Use Microdata Sample (PUMS) dataset. The main focus of the research effort is examining the use of demand side interventions in influencing the parking and/or travel mode options individuals exercise and how such interventions could be implemented in such a way that incentivize individuals to exercise parking options that are socially beneficial. The summary of the key findings is provided for each of the three empirical studies carried out.

5.1.1 Travel Mode Preferences

The study in Chapter 2 yields insights on the key determinants of the travel mode choices individual make, and how commuting behavior varies either across distinct population segments or across geographical areas. Study results show that statistically significant differences on travel mode options were observed across distinct population segments, and that not only were these differences significant, but the magnitude of the variation was substantial. Looking at gender and education, for example, our analysis revealed that on

average, women have a higher probability of using the public transit system compared to men. However, women with more than 12 years of education use public transit less than men. A similar education threshold is also observed for driving solo, although higher education levels compared to the threshold is associated with higher probability of women driving solo relative to men. Regardless of gender, there is a significant preference for using the public transit system, relative to driving alone, once the commute time goes beyond an hour. In addition, on average, carpooling to work is used much less by women compared to men. For commuters with children, non-motorized and carpooling options are rarely used. Other than driving alone, the probability of any mode choice for individuals with children trail those without children at every level of travel duration.

When we examine commuting profiles based on race, we see appreciable differences. We find that White respondents drive solo or carpool more than other races regardless of whether the probabilities were calculated based on education attainment or travel duration. In addition, once an average White commuter has the equivalent of a high school diploma, the probability of her taking public transit becomes almost nil. For Whites, the preference for utilizing the public transit systems occurs when the commute duration is an hour or more, while other ethnic groups prefer the public transit system much sooner. Among Asians there is almost a ten-fold difference in the probability of carpooling between the most and the least educated commuters. Overall, commuters are highly resistant to carpooling given the minimal spatial variation observed in the percentage of commuters who carpool, but they may be open to using a functional public transit system.

Spatial differences in travel behavior are also observed. Of the five PUMS area with the least ratio of commuters who drive alone relative to the total commuting population, the top four are from San Francisco County and the fifth is from Alameda County, all from the state of California. And population density does not explain all of the differences, as there is variability even across densely populated PUMS areas. For example, LA County has four of the five worst performing PUMS area when measured by the ratio of commuters

driving solo as a fraction of all private vehicle users. In the City of San Francisco, only 35% of the workforce drive to work solo compared to 74% for the three states analyzed, and the average San Francisco resident is more than seven times (34.4% to 4.8%) likely to use the public transit system compared to a representative commuter across the three states. In addition, a much higher percentage of San Franciscans (23.7%) either work close to home or use some form of non-motorized or bi-pedal travel option compared to 10.7% for the balance of the commuting population. As seen with demographic attributes, very little spatial variation is observed between the two cohorts with regards to carpooling probability.

5.1.2 Parking rate

Chapter 3 uses a two-stage panel data regression optimization method to build on the findings from Chapter 2. It is focused on the pricing experiment conducted by the City of San Francisco with the aid of a grant from the Department of Transportation (DOT), to change parking rates via elasticity measures in order to influence demand for parking spaces. The research has generated plausible price elasticities of parking demand for the first time using panel data methods – a first for any empirical study using panel data regression methods. In addition, the specification of the regression model was robust enough to reflect SFpark unique situation and the spatio-temporal variation in parking demand.

The findings from this study highlight the importance of creating a predictive pricing regime that allows for proactively managing the parking program, as opposed to a purely reactive approach based on the prevailing parking occupancy. In contrast to SFpark's figures, our model yielded approximately a 10% improvement in systems performance when measured by the cumulative deviation from the 60 to 80% occupancy target. What is most important are the directional guidelines provided by the policy related scenarios. These scenarios also provide an advocacy tool and help to identify key policy levers that can direct the agency managing the program towards the most effective options available to it

in order to further the program's goals. The two-stage model, when used as a planning tool, will enable SFMTA to monitor the progress towards the agency's parking occupancy threshold goals, assess the sensitivity of the results to changing operational guidelines and be able to use these findings as an advocacy tool before the public or governmental entities.

5.1.3 Parking information

Chapter 4 introduces an innovative demand-side intervention to address parking challenges in downtown Pittsburgh. The intervention, a smart parking application called ParkPGH, uses garage occupancy rates with detailed information on events and weather conditions to provide real-time and predictive parking information to parking consumers. The system has been effective in reducing parking search time and in changing the perceptions of patrons about the parking situation in downtown Pittsburgh. In 2011, more than 300,000 unique inquiries were made of the parking application.

The novelty of the application is the implementation of the first real-time and predictive parking analytics system with input from multiple stakeholders. The prediction method used in ParkPGH is not new, but the environment in which it is deployed is unique and the approaches used in addressing these challenges very innovative. More importantly, of all the components of the smart parking application, the predictive module has the potential to have the most influence on commuters' travel and parking demand patterns. Integrating parking information with traffic flow patterns is more effective if the parking information provided is not only real time but also predictive. This knowledge can encourage commuters to change the way they travel or how they schedule their trips. In addition, apart from providing information on the demand side, garage operators could use the predictive information to better manage their facilities. These findings highlight the need to prioritize the implementation of a fully functional predictive module that encompasses both long- and short-term predictive modules.

ParkPGH's relatively low cost, the ease of retrofitting, its modular structure and its open-source platform allow other cities to lower the costs of implementing and managing similar smart-parking solutions. Understanding this application can significantly shorten their learning curves and ensure that transitioning to a fully-fledged integrated parking application could be achieved at a relatively low cost. The design process emphasized the creation of citywide solutions that span both public and private providers and made the case for a coordinated approach that ensures that the application provides standardized, platform-independent solutions that could be scalable – an invaluable feature that could promote its buy-in in many other cities.

5.2 FUTURE RESEARCH DIRECTION

Given the data driven nature of our research, it is not immune to the data issues that typically bedevil most empirical studies. Except for ParkPGH that is related to systems development and the use of non-parametric regression methods, the travel mode and the parking rate studies suffer from endogeneity problems emanating from omitted variables, measurement errors and simultaneity concerns. Instrumental variables could be used to address simultaneity concerns but the effectiveness depends crucially on the quality of the instrument thus in specifying the direction for future research, we focus primarily on omitted variables and measurement errors. Methodological limitations that are data related are also discussed.

5.2.1 Measurement errors

Measurement errors may manifest in different forms and are often triggered by absence of data from relevant regressors. With regards to travel mode preferences for example, there is an assumption that the individual is cognizant of the full range of the travel options available to her and that she is sophisticated enough to tradeoff attributes and costs associated with one alternative for the other. The travel choices are assumed made with perfect information, thus any heterogeneity in choice must be related to taste variation, latent or otherwise, with regards to her sensitivity to the attributes of the travel modes. This is often not the case in reality. New movers may not have information on all the available

commuting options and even individuals who are not new to the environment, may exhibit her choices in a myopic manner, a trait that most of us have been guilty of given that humans are creatures of habit. Consequently, the heterogeneity we may be observing may have less to do with taste variation but driven more by asymmetric information.

Given perfect information on commuting options and quality, the travel mode choices utilized by commuters may be appreciably different compared to what is revealed in our dataset. This has a host of implications starting from the theoretical basis of the empirical strategy. The RUM theory is anchored on commuters making choices that maximize their utility and implicit in this theory is the assumption of perfect information by decision makers. It is obvious that commuters will not choose travel modes that they do not know exist, or where they have misconception about the quality of the service offered and often times, they stick with an option they have utilized in the past without exhausting the set of the travel mode options. However, this poses a conundrum – are mode choices influenced by habit formation or the lack of knowledge of the existence of a comparable service.

This could be attenuated by the length of time commuters have been living within a specific locality assuming we have information on the length of tenure. Reflecting this variable in the regression should control for this. In the same vein, including payment compliance data in the panel data regression for SFpark parking rate should produce a superior fit. In fact, the measurement error associated with parking occupancy is directly related to this. The ability to reflect these variables should in theory, produce a better fit and reveal more effective policy insights.

5.2.2 Omitted variables

Issues related specifically to omitted variables include the paucity of supply-side transportation related data or alternative specific data that are relevant to the travel mode discrete choice regression model or cross elasticity of parking demand measures that are relevant to the SFpark study. A supply side data challenge, for example, is the lack of data

on the quality of the public transit service - - frequency, routes covered, cost, timeliness, etc. This data paucity precludes us from establishing the direction or magnitude of travel mode switch assuming a change in transit fare, service improvement, or the elimination of some routes. In fact it could be argued that the possible heterogeneity in travel mode options is being driven in part by the differences in the quality of the transportation service, particularly public transit, of which we have no data.

From existing literature, travel cost is arguably the most important covariate of travel mode choice that was not explicitly taken into consideration in our analysis as no information exist on the costs associated with each travel mode. However, we posit that this does not invalidate the research results given that oftentimes, public transit fares do not reflect the true value and when it does, it is not unusual that the travel cost is not wholly borne by the commuter. Even when market rates are used, transit fares could still be priced differently for dissimilar segments of the population through the use of vouchers, transit subsidies, seniors' rate, and employers' assistance program to mention just a few.

Where alternative specific data exists, Commuters may exhibit different sensitivities to the sub-attributes of the travel mode options. The PUMS dataset has data on travel time, an alternative specific data. However the variable captures the total travel time irrespective of the number of travel mode options utilized given that the question on the survey simply asked "how many minutes did it usually take this person to get from home to work last week?" Even within a specific alternative, finer segmentation of the travel time is possible. For public transit, a significant portion of the total travel time may be out-of-vehicle time, say time spent waiting for a connecting ride. To the degree that commuters may find one more odious compared to another, ideally the time associated with each activity of a trip should be measured. These concerns provide promising challenges to be tackled in the near future.

5.2.3 Methodological challenges

Over and above the aforementioned are methodological challenges. Ideally, the policy instrument in changing parking demand via rate changes should be tied to marginal occupancy, not average occupancy. As argued, a finer time window may ameliorate this but at best, the value added will only be incremental. To achieve innovative changes in travel mode behavior, a policy instrument based on marginal occupancy data has to be deployed – effectively transitioning from a variable pricing to a dynamic one using marginal occupancy data. Assuming we have access to marginal occupancy data, one way to implement this is to employ an adaptive pricing mechanism that uses parking rate and occupancy data from the predictive model as the baseline with these figures revised based on present marginal occupancy data. Each time the marginal occupancy data is updated, the adaptive pricing mechanism compares the current rate to the base values for each block time band. If the block for that specific time has a higher parking occupancy compared to the expected value suggested by the predictive model, rates are adjusted downwards. The reverse is the case if a higher occupancy is observed. The process continues until the desired marginal occupancy level is reached. Achieving parking occupancy within the 60 to 80% threshold across space and time hinges crucially on this pricing approach.

5.3 LOOKING FORWARD

The rationale for this thesis is the inordinate number of individuals who drive alone and the attendant implications of this travel behavior on parking demand. We have examined the effectiveness of policy levers in influencing individuals' travel preferences and the impact of parking rates and advance information on the demand for parking spaces. Looking forward, we pose the question as to what the near future portends for travel behavior and the demand for parking spaces, particularly in urban areas in light of recent disruptive technologies.

It is safe to assume that the advent of peer-to-peer mobility juxtaposed with driverless vehicles will upend the prevailing landscape. As these innovations become more mainstreamed, the reduction in car ownership and increased car-pooling could lead to tens of millions of parking

spaces being repurposed for totally different economic activities, significantly altering the urban landscape. Uber, for example, is investing in these frontier technologies while also partnering with cities and towns to conduct pilot programs that couple car sharing with first and last mile services.

As a critical mass of vehicles become self-driving and the public increasingly embraces travel behavior that is peer-to-peer mobility driven, we leave it to the reader's imagination how this transformation could influence the demand for parking spaces and the menu of options and the quality of offerings afforded to consumers. It also underscores the need to revisit the present thesis in the near future. The literal use of the phrase - may you live in interesting times, could not be more apt.

References

1. **McFadden, D.** The Measurement of Urban Travel Demand. *Journal of Public Economics*. 1974, Vol. 3, 4, pp. 303–328.
2. **McKenzie, Brian.** *Who Drives to Work? Commuting by Automobile in the U.S. in 2013*. American Community Survey Report, ACS-32 Bureau of US Census, 2015.
3. **United Kingdom Department of Transport.** 2008 Annual Report. [Online] [Cited: March 28, 2015.] www.gov.uk/government/publications/departments-for-transport-annual-report-2008.
4. **Frey, W.** Population and Migration. *State of Metropolitan America - On the Frontlines of Demographic Transformation*. Washington D.C., 2010.
5. **Cortright, Joe.** Surging City Center Jobs. *Cityobservatory.org*. [Online] February 2015. [Cited: February 18, 2015.] <http://cityobservatory.org/wp-content/uploads/2015/02/Surging-City-Center-Jobs.pdf>.
6. **Ben-Akiva, M and Lerman, S.** *Discrete Choice Analysis: Theory and Application to Travel Demand*. Boston : MIT Press, 1985.
7. **Ben-Akiva, M, et al.** Integration of Choice and Latent Variable Models. MIT Working Paper. 1999.
8. **Manski, Charles F and Lerman, Steven R.** The estimation of choice probabilities from choice based samples. *Econometrica: Journal of the Econometric Society*. 1977, Vol. 45, 8, pp. 1977-1988.
9. **Rosenbloom, S and Burns, E.** Gender differences in commuter travel in Tucson: Implications for travel demand management programs. *Transportation Research Record of the National Academies*. 1404, 1993, pp. 82-90.
10. —. Why working women drive alone: Implications for travel reduction programs. *Transportation Research Board*. 1459, 1994, pp. 39-45.
11. **Weinberger, R.** Men, Women, Job Sprawl and Journey to Work in the Philadelphia Region. *Public Works Management and Policy*. 2007, Vol. 11, 3, pp. 177–193.
12. **Blumenberg, E and Smart, M.** Brother can you spare a ride? Carpooling in immigrant neighbourhoods. *Urban Studies*. 2014, Vol. 51, 9, pp. 1871-1890.
13. **Blumenberg, E and Evans, A.** Planning for Demographic Diversity: The Case of Immigrants and Public Transit. *Journal of Public Transportation*. 2010, Vol. 13, 2, pp. 23-45.

14. **Kim, S.** Immigrants and Transportation: An Analysis of Immigrant Workers' Work Trips. *Cityscape: A Journal of Policy Development and Research*. 2009, Vol. 11, 3.
15. **Giuliano, G.** Travel, location and race/ethnicity. *Transportation Research Part A: Policy and Practice*. 2003, Vol. 37, 4, pp. 351-372.
16. **Blumenberg, E and Smart, M.** Travel in the 'hood: Ethnic neighborhoods and mode choice. *88th Annual Meeting of the Transportation Research Board*. 2009.
17. **Ferguson, E.** The rise and fall of the American carpool: 1970–1990. *Transportation*. 1997, Vol. 24, pp. 349–376.
18. **Bhat, C and Sen, S.** Household Vehicle Type Holdings and Usage: An Application of the Multiple Discrete-Continuous Extreme Value (MDCEV) Model. *Transportation Research Part B*. 2006, Vol. 40, 1, pp. 35-53.
19. **Barbour, E.** Time to Work: Commuting Times and Modes of Transportation of California Workers. *Public Policy Institute of California*. 2006, Vol. 7, 3.
20. **Thakuriah, V, et al.** Activity and Travel Changes of Job Access Transportation Service Users: Analysis of a User Survey. *Transportation Research Record: Journal of the Transportation Research Board*. 1927, 2005, pp. 55-62.
21. **Waddell, P, et al.** Design of an Integrated Land Use and Activity-Based Travel Model System for the Puget Sound Region. *Transportation Research Record: Journal of the Transportation Research Board*. 2002, Vol. 1805.
22. **Cervero, R and Duncan, M.** Walking, Bicycling, and Urban Landscapes: Evidence From the San Francisco Bay Area. *American Journal of Public Health*. 2003, Vol. 3, 9, pp. 1478–1483.
23. **Zhang, M.** Travel Choice with No Alternative: Can Land Use Reduce Automobile Dependence? *Journal of Planning Education and Research*. 2006, Vol. 25, 3, pp. 311-326.
24. **Blumenberg, E and Shiki, K.** Transportation Assimilation : Immigrants, Race and Ethnicity, and Mode Choice. *86th Annual Meeting of the Transportation Research Board*. 2007.
25. **Krizek, K, Johnson, P and Tilahun, N.** Gender Differences in Bicycling Behavior and Facility Preferences. *Conference Proceedings 35 Research on Women's Issues in Transportation*. Nov 18-20, 2004.
26. **Hu, L and Giuliano, G.** Beyond the Inner City: A New Form of Spatial Mismatch. *Proceedings of the 90th Annual Meeting of the Transportation Research Board*. Jan 23-27, 2011.

27. **Haas, P, et al.** Housing and Transportation Cost Trade-Offs and Burdens of Working Households in 28 Metros, Center for Neighborhood Technology, Chicago, IL. July 2006.
28. **Donnelly, R, et al.** NCHRP Synthesis 406: Advanced Practices in Travel Forecasting. *Transportation Research Board of the National Academies, Washington D.C.* 2010.
29. **Deal, B, Kim, J and Pallathucheril, V.** Growth Management and Sustainable Transport: Do Growth Management Policies Promote Transit Use? *Proceedings of the 88th Annual Meeting of the Transportation Research Board, Washington D.C.* 2009.
30. **Huang, H, Yang, H and Bell, M.** The models and economics of carpools. *Annals of Regional Science.* March 2000, Vol. 34, 1, pp. 55-68.
31. **Charles, K and Kline, P.** Relational Costs and the Production of Social Capital: Evidence from Carpooling. *The Economic Journal.* 2006, Vol. 116, 511, pp. 581-604.
32. **Blumenberg, E and Smart, M.** Getting by with a little help from my friends... and family: immigrants and carpooling. *Transportation.* 2010, Vol. 37, 3, pp. 429-446.
33. **Ben-Akiva, M, et al.** Hybrid Choice Models: Progress and Challenges. *Marketing Letters.* 2002, Vol. 13, 3.
34. **Van Acker, V and Witlox, F.** Car ownership as a mediating variable in car travel behaviour research using a structural equation modeling approach to identify its dual relationship. *Journal of Transport Geography.* 2010, Vol. 18, 1.
35. **Atasoy, B, Glerum, A and Bierlaire, M.** Mode choice with attitudinal latent class: a Swiss case-study. *2nd International Choice Modeling Conference, Leeds UK.* 2011.
36. **Walker, J and Li, J.** Latent Lifestyle Preference and Household Location Decisions. *Journal of Geographic Systems.* 2007, Vol. 9, 1, pp. 77-101.
37. **Chorus, C and Kroesen, M.** On the (im-) possibility of deriving transport policy implications from hybrid choice models. *Transport Policy.* 2014, Vol. 36.
38. **Vij, A and Walker, J. L.** Preference Heterogeneity in Discrete Choice Models. *Transportation Research Part B.* 2014, Vol. 64, pp. 90-105.
39. **Garrido, R and Mahmassani, H.** Forecasting freight transportation demand with the space-time multinomial probit model. *Transportation Research Part B: Methodological.* 2000, Vol. 34, 5, pp. 403-418.

40. **United States Census Bureau/American FactFinder.** American Community Survey (ACS). *B01003: Total Population. 2009-2013.* [Online] 2013. [Cited: March 28, 2015.] <http://factfinder2.census.gov>.
41. **United States Census Bureau .** Public Use Microdata Sample (PUMS) Dataset. [Online] 2009-2011. [Cited: May 13, 2013.] http://www.census.gov/acs/www/data_documentation/pums_data/.
42. **Fay, R.** *VPLX: Variance Estimation for Complex Surveys, Program Documentation.* Bureau of US Census. 1995.
43. **McGuckin, N and Srinivasan, N.** National Summary, Journey to Work Trends in the United States and its Major Metropolitan Areas 1960-2000. *Federal Highway Administration, #FHWA -EP-03-058.* 2003.
44. **Blumenberg, E and Song, L.** Travel Behavior of Immigrants in California: Trends and Policy Implications. *87th Annual Meeting of the Transportation Research Board, Washington D.C.* Jan 2008.
45. **Colby, S and Ortman, J.** Projections of the Size and Composition of the U.S. Population: 2014 to 2060. *U.S. Census Bureau Current Population Reports.* 2014, pp. 25-1143.
46. **San Francisco Municipal Transportation Authority.** SFpark: Putting theory into practice. [Online] 2011. Retrieved from http://sfpark.org/wp-content/uploads/2011/09/sfpark_aug2011projsummary_print-2.pdf.
47. —. SFpark Pilot Project Evaluation. [Online] 2014. Retrieved from http://sfpark.org/resources/docs_pilotevaluation/.
48. **Pierce, G and Shoup, D.** Getting the Prices Right. *Journal of the American Planning Association.* 2013, Vols. 79:1, 67-81.
49. **Chatman, D. G and Manville, M.** Theory versus implementation in congestion-priced parking: An evaluation of SFpark, 2011-2012. *Research in Transportation Economics.* 2014, Vols. 44, 52-60.
50. **Millard-Ball, A, Weinberger, R and Hampshire, R.** Is the curb 80% full or 20% empty? Assessing the impacts of San Francisco's parking pricing experiment. *Transportation Research Part A.* 2014, Vols. 63, 76-92.
51. **Millard-Ball, A and Weinberger, R: Hampshire, R.** Comment on Pierce and Shoup: Evaluating the impacts of performance-based parking. *Journal of the American Planning Association.* 2013, Vols. 79 (4):330–336.

52. **Ottosson, D. B, et al.** The sensitivity of on-street parking demand in response to price changes: a case study in Seattle, WA. *Transport Policy*. 2013, Vols. 25, 222-232.
53. **Jakle, J and Sculle, K.** *Lots of Parking: Land Use in a Car Culture*. Charlottesville : University of Virginia Press, 2004.
54. **Shoup, D.** *The high cost of free parking*. Chicago, IL : Planners Press, 2011.
55. **Arnott, R and Inci, E.** An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics*. 2006, Vols. 60, 418-442.
56. **Inci, E.** A review of the economics of parking. *Economics of Transportation*. 2015, Vols. 4, 50-63.
57. **Vickery, W. S.** The economizing of curb parking space. *Traffic Engineering*. 1954, 62-67.
58. **Arnott, R.** Spatial competition between parking garages and downtown parking policy. *Transport Policy*. 2006, Vols. 13, 458-469.
59. **Froeb, L, Tschantz, S and Crooke, P.** Bertrand competition with capacity constraints: mergers among parking lots. *Journal of Econometrics*. 2003, Vols. 113, 49-67.
60. **Chone, P and Linnemer, L.** A treatment effect method for merger analysis with an application to parking prices in Paris. *Journal of Industrial Economics*. 2012, Vols. 60, 631-656.
61. **De Nijs, R.** The price discrimination effect of a large merger of parking garages. *Economic Letters*. 2012, Vols. 117, 928-931.
62. **Kobus, M.B.W., et al.** The on-street parking premium and car drivers' choice between street and garage parking. *Regional Science and Urban Economics*. 2013, Vols. 43, 395-403.
63. **Calthrop, E and Proost, S.** Regulating on-street parking. *Regional Science and Urban Economics*. 2006, Vols. 36, 29-48.
64. **Vickerman, R.** Evaluating the Wider Economic Impacts of Congestion Charging Schemes: the Limitations of Conventional Modeling Approaches. *Paper presented at the 45th ERSA Congress, Amsterdam, The Netherlands*. August 2005.
65. **Clinchant, S, et al.** Using analytics to understand on-street parking; the impact of special permit use and the benefit of demand based rates over zones. *Proceedings of 22nd ITS World Congress*. 2015.

66. **Lin, T.** Smart Parking, Network Infrastructure and Urban Services. *Computer science doctoral thesis, INSA, Lyon*. 2015.
67. **Russo, A.** Voting on road congestion policy. *Regional Science and Urban Economics*. 2013, Vols. 43, 707–724.
68. **van Ommeren, J.N, Wentink, D and Rietveld, P.** Empirical evidence on cruising for parking. *Transportation Research Part A*. 2012, Vols. 46, 123–130.
69. **Small, K and Verhoef, E.** *The Economics of Urban Transportation*. Abingdon, UK : Routledge, 2007.
70. **Roth, G.** *Paying for parking*. London : Hobart Papers, Institute of Economic Affairs, 1965.
71. **Shoup, D. C.** An opportunity to reduce minimum parking requirements. *Journal of American Planning Association*. 1995, Vols. 61, 14-28.
72. —. Evaluating the effects of cashing out employer-paid parking: eight case studies. *Transport Policy*. 1997, Vols. 4, 201-216.
73. —. The trouble with minimum parking requirements. *Transportation Research Part A*. 1999, Vols. 33, 549-574.
74. —. *Parking Cash Out*. Chicago : American Planning Association, 2005.
75. —. *The High Cost of Free Parking*. Chicago : American Planning Association, 2005.
76. **Vaca, E and Kuzmyak, J. R.** *Parking Pricing and Fees - Traveler Response to Transportation System Changes*. Transit Cooperative Research Program (TCRP) Report 95. Washington, D.C. : Transportation Research Board, 2005.
77. **Litman, T., 2011.** **Transportation Elasticities: How prices and Other Factors Affect.** *Transportation Elasticities: How prices and Other Factors Affect Travel Behavior*. Victoria, B.C. : s.n., 2011.
78. **Kelly, J. A and Clinch, J. P.** Temporal variance of revealed preference on-street parking price elasticity. *Transport Policy*. 2009, Vols. 16 (4), 193-199.
79. **Parisi, G.** *Statistical Field Theory*. Redwood City, CA : Addison-Wesley, 1988.
80. **Kadanoff, L. P.** More is the Same: Phase Transitions and Mean Field Theories. *Journal of Statistical Physics*. 2009, Vols. 137 (5–6): 777–797.
81. **Wooldridge, J.** *Econometric Analysis of Cross Sectional and Panel Data*. Cambridge, MA : MIT Press, 2002.

82. **Clarke, P, et al.** Revisiting fixed- and random-effects models: some considerations for policy-relevant education research. *Education Economics*. 2015, Vols. 23 (3), 259-277.
83. **Snijders, T and Berkhof, J.** Diagnostic checks for multilevel models. [book auth.] J. & Kreft, I. (eds) de Leeuw. *Handbook of Quantitative Multilevel Analysis*. New York : Sage, 2004.
84. **Colin, C. A and Trivedi, P. K.** *Microeconometrics Using Stata*. College Station, TX : Stata Press, 2010.
85. **Baltagi, B. H.** Forecasting with Panel Data. *Center for Policy Research Working Paper*. 2007, Vol. 91.
86. **Frey, W.** Population and Migration. *State of Metropolitan America: On the Frontlines of Development Transformation*. Washington D.C. : Brooklyn Institution, 2010, 4, pp. 36-49.
87. **United Nations Department of Economic and Social Affairs, Population Division.** *World Urbanization Prospects: 2014 Revision, Highlights (ST/ESA/SER.A/352)*. 2014.
88. **United Nations Population Fund (UNFPA).** State of the World Population 2007: Unleashing the Potential of Urban Growth. [Online] 2007. [Cited: November 18, 2014.] <http://www.unfpa.org/publications/state-world-population-2007>.
89. **Fabusuyi, T. and Hampshire, R.** *Needs Assessment for an Integrated Parking Application Project*. Report submitted to Carnegie Mellon University/University of Pennsylvania University Technology Center (UTC). 2013.
90. **Pittsburgh Downtown Partnership.** State of Downtown Pittsburgh 2012. [Online] 2012. [Cited: February 11, 2014.] http://www.downtownpittsburgh.com/_files/docs/pdp-sod_report_lowres.pdf.
91. **Strada, L.** Redeveloping downtown Pittsburgh — The last 20 years. [Online] 2009. [Cited: May 24, 2012.] http://www.stradallc.com/files/07.28.09_white_paper_1.pdf.
92. **Shaheen, S and Kemmerer, C.** Smart Parking Linked to Transit: Lessons Learned from Field Test in San Francisco Bay Area of California. *Transportation Research Record of the National Academies*. 2008, Vol. 2063, pp. 73-80.
93. **Orski, K.** Best Space Scenario. *Traffic Technology International*. 2003, pp. 54-56.
94. **Caicedo, Felix.** The use of space availability information in 'PARC' systems to reduce search times in parking facilities. *Transportation Research Part C*. 2009, Vol. 17, 1, pp. 56-68.

95. **Caliskan, M., et al.** Predicting Parking Lot Occupancy in Vehicular Ad Hoc Networks. *Vehicular Technology Conference (VTC2007-Spring IEEE 65th)*. April 2007, pp. 277-281.
96. **Teng, H., Qi, Y. and Martinelli, D. R.** Parking difficulty and parking information system technologies and costs. *Journal of Advanced Transportation*. 2008, Vol. 42, 2, pp. 151–178.
97. **Martens, K. and Benson, I.** Evaluating urban parking policies with agent based model of driver parking behavior. *Transportation Research Record: Transportation Research Board of the National Academies*. 2008, Vol. 2046, pp. 37-44.
98. **Teodorovic, D. and Lucic, P.** Intelligent parking systems. *European Journal of Operational Research*. 2006, Vol. 175, pp. 1666-1681.
99. **Vlahogianni, E. I., et al.** Exploiting new sensor technologies for real-time parking prediction in urban areas. *Transportation Research Board 93rd Annual Meeting Compendium of Papers*. 2014, Vols. 14-1673.
100. **Caicedo, F., Blazquez, C. and Miranda, P.** Prediction of parking space availability in real time. *Expert Systems Applications*. 2012, Vol. 30, pp. 7281-7290.
101. **Yang, Z., Liu, G. and Wang, X.** The research on the key technologies for improving efficiency of parking guidance system. *Proceedings of IEEE intelligent transportation systems, Shanghai, China*. 2003.
102. **Bos, I., Ettema, D. and Molin, E.** Modeling Effect of Travel Time Uncertainty and Traffic Information on Use of Park-and-Ride Facilities. *Transportation Res. Record: J. Transportation Res. Board*. 2004, Vol. 1898, pp. 37-44.
103. **Specht, D.** A General Regression Neural Network IEEE Transaction on Neural Networks. *IEEE Transaction on Neural Networks*. 1991, Vol. 2, 6.
104. **Svozil, D. et al.** Introduction to multi – layer feed forward neural networks. *Chemometrics and Intelligent Laboratory Systems*. 1997, Vol. 39, pp. 43-63.
105. **Mitchell, T.** Generative and Discriminative Classifiers: Naive Bayes and Logistic Regression. [Online] 2010. [Cited: March 3, 2013.] <http://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf>.
106. **Loh, W. Y.** Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining Knowledge Discovery*. 2011, Vol. 1, 1, pp. 14-23.
107. **Howard, G. S.** Response-shift Bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*. 1980, Vol. 4, 1, pp. 93-106.

108. **Bureau of Labor Statistics.** Occupational Employment and Wage Estimates for Pittsburgh, PA. *US Department of Labor, Bureau of Labor Statistics.* [Online] May 2010. [Cited: April 7, 2012.] http://www.bls.gov/oes/oes_ques.htm.