DEVELOPMENT OF PREDICTION MODEL FOR REAL-TIME PARKING AVAILABILITY FOR ON-STREET PAID PARKING

by

Chang Liu

B.S. Engineering, Beijing Institute of Technology, 2014

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science in Civil Engineering

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This thesis was presented

by

Chang Liu

It was defended on

November 18th, 2015

and approved by

Mark Magalotti, PhD, P.E. Senior Lecturer, Department of Civil and Environmental Engineering

Konstantinos Pelechrinis, PhD, Assistant Professor, School of Information Sciences

Cynthia Jampole P.E. Principal Trans Associates

Thesis Advisor: Mark Magalotti, PhD, P.E. Senior Lecturer, Department of Civil and Environmental Engineering

DEVELOPMENT OF PREDICTION MODEL FOR REAL-TIME PARKING AVAILABILITY FOR ON-STREET PAID PARKING

Chang Liu, M.S.

University of Pittsburgh, 2015

This research is based on the need for a parking information system to provide on- street parking availability information in the neighborhoods in the City of Pittsburgh has been established. This project intends to provide a methodology to determine the real-time parking availability information for on street parking operators.

As a part of this project, this research models the parking availability with no specialized hardware other than purchased parking time from a kiosk type pay system. The prediction model developed is based on the sample data collected with no real-time data required other than paid time information.

Because paid parking systems only record the time of arrival and paid parking time, the actual departure time and thus the availability of the parking space is unknown. This research determined the relationship between over-paid on-street parking time, which means that the owner pays more than the actual parking time needs and how this relates to parking space availability for next vehicle based on actual on-street parking time. Many variables may influence over-paid parking time, such as trip purpose types of users, weather conditions and temporal distribution. All of these variables were explored to develop the predictive model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

This chapter introduces the background, hypothesis, objectives and methodology of this research.

1.1 BACKGROUND

This research is based on the need for a parking information system to provide on-street parking availability information in the neighborhoods in the City of Pittsburgh has been established. This project intends to provide a methodology to determine the real-time parking availability information for on street parking operators.

As a part of this project, this research intends to model the parking availability with no specialized hardware used as availability counter for real-time parking information, which means the prediction model is predicted based on the data collected with no real-time data provided.

1.2 HYPOTHESIS

The hypothesis of this research is that a real time parking availability information systems for on street parking that uses pre-paid parking could be increase occupancy if a prediction model was developed to provide information on availability based a predicted parker's departure time.

Currently on street pre-paid systems can only report arrival time and payment period but have no information on the actual departure time for parkers. It is hypothesized that many parkers overpay because of the uncertainty of their parking needs and more space hours of supply are availability than may be reported.

This research will determine the relationship between over-paid on-street parking time, which means that the owner pays more than the actual parking time needs and how this may make the same parking space available for next vehicle based on actual on-street parking time. Many variables may influence over-paid parking time, such as trip purpose types of users, weather conditions and temporal distribution.

The research mainly conducted in Oakland area of the City of Pittsburgh, Pennsylvania and four different neighbors were selected to develop and test a methodology to predict this relationship. This was done by data collection and analysis of information from the Pittsburgh Parking Authority (PPA) Field data was collected on actual pay durations, this information was used to create a predictive model and then the model was used to verify new field data to verify validity of the mode.

1.3 OBJECTIVES

The objectives of this thesis were to identify the relationship between the parking availability and the overpaid parking situations; to explore the influence factors that have impacts on the parking availability; and to produce a prediction model for the real-time parking information system.

1.4 METHODOLOGY

The software R was to be used to analyze the linear regression model between the parking time ratio, which is defined as the result of actual parking time divided by paid parking time, and all the variables that have impact on overpaid parking time. The test of model was done by using k-fold cross validation to predict the data collected to measure if there were any needs to adjust the model and how to adjust it.

2.0 LITERATURE REVIEW

This chapter describes practices and research on parking studies relevant to this thesis.

2.1 INTRODUCTION

This literature review describes practices and current research relevant to this the hypothesis. The current research mainly discusses the parking reservation and estimation of available on street parking spaces. The sensor or video system technologies being developed for real-time parking information are also mentioned in some of the research. This chapter also covers the evaluation of current parking practice and research in this topic area.

2.2 CURRENT RESEARCH BACKGROUND

The hypothesis of the research is to define the time gap between the time the driver pays for on-street parking and the actual on-street parking time used. Developing a prediction model, which is a part of the development of a model, will be completed after the data is collected and analyzed. Thus, the search for relevant research included the topics of on-street parking availability prediction and pay duration studies, which were considered as relevant.

2.2.1 Parking information system

Parking information systems are designed to minimize the cruising time for finding the available on street parking spaces. As previously stated, the time gap, which involves the time purchased for parking and the actual parking time, needs defined in order to determine when parking spaces are available as compared to what time was purchased. With the help of a parking information system, the data of the time purchased for parking is easy to report to the public. Also, it improves the parking system and helps manage the whole system to be more efficient.

Zhibin Chen et al. (2014) discussed a smartphone-based parking reservation system that manages a finite number of curbside parking spaces.[[1]]The parking social cost, as the efficiency factor of the system, is measured in the research as a weighted sum of the cruising time for drivers to find the available parking spaces and the walking time from the parking to destination. After the establishment of the parking social cost, Chen et al. presents a simple model to test the efficiency of the systems and finally an adjustment is applied to the cost model to make the model correspond to reality. The model mainly targets on how to optimize parking information provided to users to give the whole system an optimal solution, which means a balance of all the parking users travel time to keep the total time the shortest possible to find a parking space thus resulting in the lowest total parking social cost.

Diana Carvalho e Ferreira et al. (2012) also simulates a system of in-advance online parking space reservations for on-street parking to reduce the cruising time.[[2]] A model of parking and cruising behavior is presented, and also a stochastic discrete-event queuing micro-simulator is described to compare scenarios of variable allocations of parking spaces for reservations. The purpose of this research is to eliminate the congestion caused by cruising.

2.2.2 Estimation of parking spaces

The estimate of available parking spaces is also very relevant to the hypothesis. Systems that provide this information could also potentially determine when parkers leave as compared to what the duration was for the paid time.

Abdoulaye Dialllo (2012) developed indicators on parking spaces by using the OD (Origin and destination) survey data within a given area. [3] The vehicle accumulation profile (VAP) is used to estimate the theoretical parking capacity. Also, Diallo illustrates that by using data from OD surveys, it is possible to determine the vehicle accumulation profile by type of parking, by trip purpose and by region of origin of the movement for a typical day or weekday of fall. After the theoretical analysis, the field survey was conducted for testing and error analysis. This work thus demonstrated that it is possible to estimate the parking capacity from OD survey data.

Brain Maleck et al. (2014) developed another way to identify the individual parking spaces using GPS by developing space-by-space parking inventory maps.[4] This research explores the use of GPS to inventory individual parking space data. The authors also indicated that GPS data were ported into the GIS using longitude and latitude coordinates (WGS84 Datum). Another edit tool, ArcGIS, was used to correct the positional errors from data collection. With the help of this system, Clemson University now has an accurate count of spaces, which helps a lot in university parking management.

2.2.3 Sensor systems for real-time parking

Sensor systems, reflecting the real-time parking data, are used when predicting parking spaces availabilities, which relates to this research.

Eleni I. Vlahogianni et al. (2014) proposed a methodological framework to provide parking availability forecasts for extended prediction horizons. Two different types of predictions are provided: i. the probability of free space to continue free in subsequent time intervals, and ii. The short-term parking occupancy prediction in selected regions of an urban road network.[[5]] Two models were found to best describe the aforementioned situations, Weibull parametric models for the first situation and Multilayer Perceptions for region parking occupancy prediction. The authors concluded that the duration of free parking space follows a Weibull distribution. Moreover, the neural networks adequately capture the temporal evolution of parking occupancy and may accurately predict occupancy up to one hour ahead.

Orhan Bulan et al. (2013) presented a new video-based on-street parking detection system. The system accounts for the inherent challenges that exist in on-street parking settings, including illumination changes, rain, shadows, occlusions and camera motion.[[6]] Also, Orhan claims that the method they used utilizes several components from video processing and computer vision for motion detection, background subtraction, and vehicle detection. This system has a higher accuracy under various weather conditions among others. For improving accuracy, the authors claimed that a verification procedure based on a machine learning approach can be effectively used for refinement of candidate regions and localizing parked vehicles within these regions.

Lattunen Ali et al. (2013) tried to find an advanced solution for collecting car-parking information. In the paper, the authors introduce an information system for on street parking, which is based on real-time event-based data collection.[7] The information is collected from mobile payment providers and pay-and-display machines, and is analyzed by using parking data broker to take care of integration of various parking-related data in a flexible and highly maintainable way. The authors concluded that use of a single permission identifier for on-street parking, the plate number, combined with availability of real-time information systems for data collection, management and online lookup of payments and permissions, makes it possible to realize novel solutions for parking guidance and enforcement.

2.3 SUMMARY

In this section, the current research information involving parking reservation systems, the estimation of available on street parking spaces and the sensor systems for data collection of this information was reviewed.

In summary, there is little current research relating to the determination of the relationship between over-paid on-street parking time and actual on-street parking time. The Lauttuen research did collect real time payment information but did not establish actual departure times for parkers. This research attempts to build a model of parking prediction after the establishment of the parking time gap, which is the gap between the time bought for parking and the actual parking time.

Based on this literature review, it was concluded that real-time parking prediction model is needed to provide actual parking availability information as compared to paid parking times and could be established without adding any other complex equipment to system. This model is not only good for parking management, but also has benefits to mitigating the congestion caused by cruising for parking spaces after the further study of online parking system, increasing occupancy of on street parking and corresponding revenues.

3.0 DATA COLLECTION AND ANALYSIS

This chapter mainly discusses the data collection and analysis, which includes data collection plan, introduction of a new concept the real time parking gap and the preliminary development of a prediction model.

3.1 DATA COLELCTION

From the hypothesis, this research targets the prediction of parking time gap between time bought for parking and actual parking time. To reflect the real situation and analyze the prediction model, a data collection plan was needed. The data collected included the two parking times (paid vs. actual use), users' characteristic and other information that may influence the relationship and results. The researcher used the data collected as variables and analyzed the relationship between user characteristics and the parking time gap.

3.1.1 Location of data collection

For collecting data to support the hypothesis and perform the analysis, 4 locations were selected in the City of Pittsburgh, Pennsylvania. They were Forbes Avenue in Oakland, Thackeray

Avenue and Tech Street in Oakland and East Carson Street in Southside. These locations are sections of the City of Pittsburgh that represent varying on street parking characteristics.

As stated in the hypothesis, the purpose of this research is to predict the time associated over-paid parking spaces in the City of Pittsburgh. For this purpose the researcher selected the Oakland area and three locations within Oakland with different land uses characteristics, Thackeray Avenue and Tech Street for university parking characteristics as well as working and Forbes Avenue for shopping and restaurant parking characteristics. Also, the researcher selected another location as contrasting neighborhood in another sections of the city, which is the Southside neighborhood in the City. Figure 3-1 shows the general data collection locations of each neighborhood. The details are in the next section.

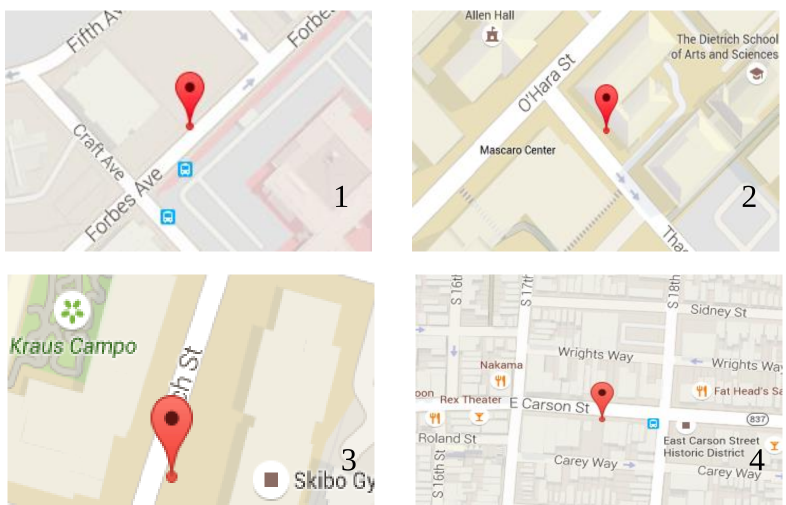


Figure 3-1. Data collection location of each neighborhood

(1. Forbes Ave; 2. Thackeray Ave; 3. Tech St; 4. E Carson St)

Table 3-1 shows the summary of information on each section where data was collected and studied. All of these study areas currently has paid on-street kiosk type parking payment systems. Also shown are the current parking rates, maximum parking duration and hours of enforcement.

Table 3-1. Detail information of Study Area locations

Location	Node	Rate
FORBES AVE	Oakland	$3/hr./2hr max/M-SAT,8AM-6PM
THACKERAY AVE	Oakland	$3/hr./4hr max/M-SAT,8AM-6PM
TECH ST	Oakland	$2.25/hr./10hr max/M-SAT,8AM-6PM
E CARSON ST	Southside	$1.5/hr./2hr max/M-SAT,8AM-6PM

3.1.2 Data collection plan

The data collection was conducted from September 8th, 2015 for two weeks. For every location, during two 8-hour durations data was and used for data analysis. One kiosk is contained in each of the data collection areas. Data was collected on pay durations, actual and paid through interviews with parkers, however not every car's duration of the parking could be collected. This is because parker's were present at the beginning and end of the data collection period. Also the drivers' trip purpose of parking was included as a part of the interview survey.

3.1.2.1 Study Area Characteristics

Forbes Avenue, Oakland

Figure 3-1 shows the kiosk locations of Forbes Ave, Oakland.

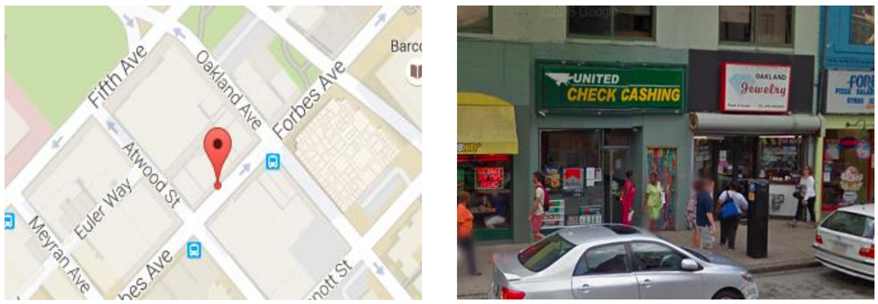


Figure 3-2. Kiosk location on Forbes Ave

Forbes Avenue is one of the busiest streets in the City of Pittsburgh. Through Oakland, Forbes Avenue continues eastward past Carnegie Mellon University and Schenley Park, and through the neighborhood of Squirrel Hill. Because it contains parts of both universities and there are some retail stores and restaurants along this road, it attracts many parkers and therefore on-street parking is in high demand. Figure 3-3 shows the land uses along the Forbes Avenue.

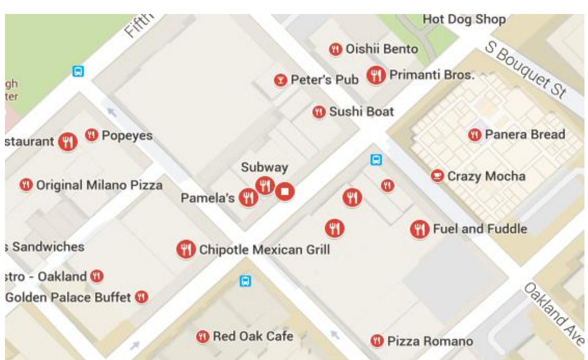


Figure 3-3. Land uses along the Forbes Ave

Thackeray Avenue, Oakland

Figure 3-4 shows the kiosk locations on Thackeray Avenue, Oakland.

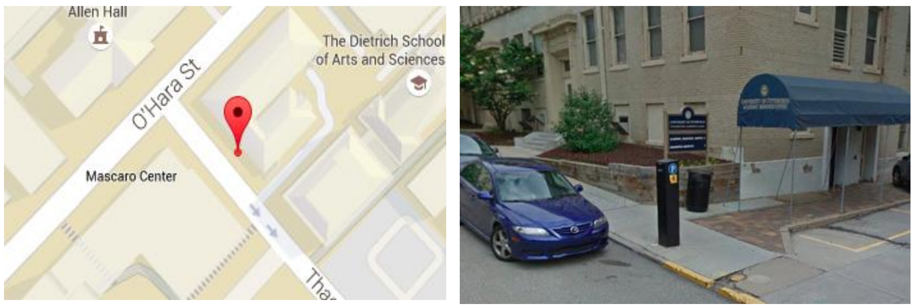


Figure 3-4. Kiosk locations on Thackeray Avenue, Oakland

On both sides of the street, all of the buildings belong to the University of Pittsburgh or associated uses. As the locations show in Figure 3-4, there are university buildings on both sides. For this University location with many university buildings around, on-street parking is always in demand by students and visitors. Land use type can be seen in Figure 3-5.

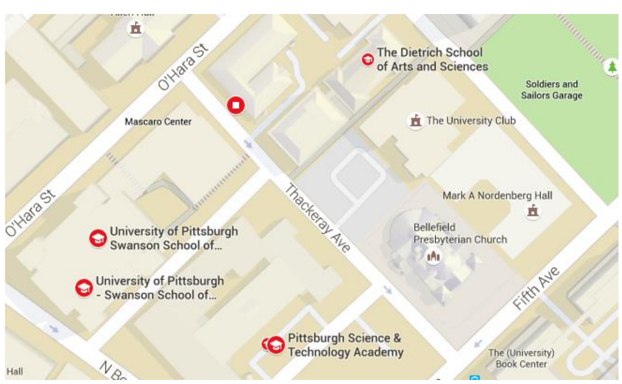
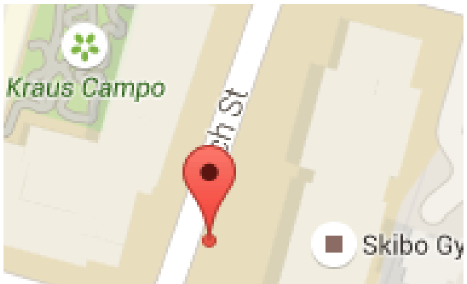


Figure 3-5. Land uses along Thackeray Avenue

Tech Street, Oakland

Figure 3-6 shows the kiosk locations on Tech Street, Oakland.



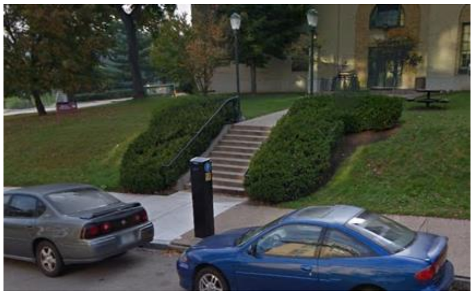


Figure 3-6. Kiosk location on Tech St, Oakland

Tech Street is located near Carnegie Mellon University. It has a classroom building on one side and a gym center on another, which makes on street parking in high demand for students and visitors. Figure 3-7 shows the land use around this area.

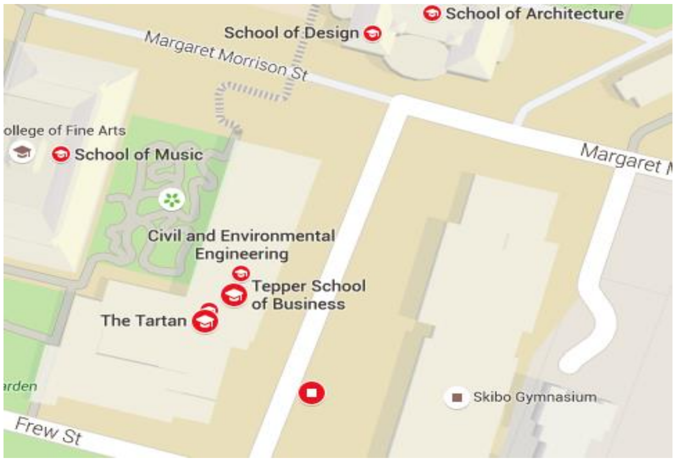


Figure 3-7. Land uses along Tech Street

East Carson Street, Southside

Figure 3-8 shows the kiosk locations on East Carson Street, Southside.

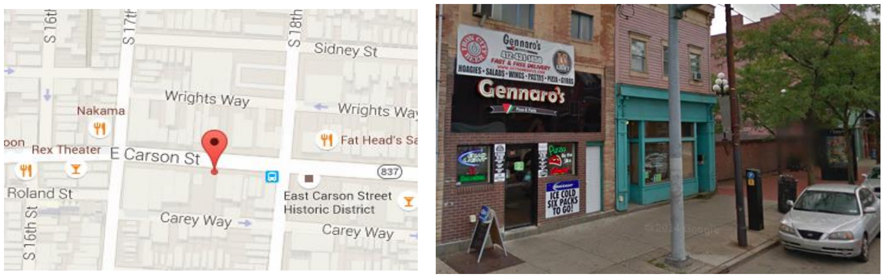


Figure 3-8. Kiosk locations on E Carson Street, Southside

East Carson Street is another well-known business district in Pittsburgh for its fine dining and private boutiques. The street has hundreds of restaurants with high quality, which attracts a large quantity of parkers every day. Figure 3-9 shows the land use of E Carson Street.

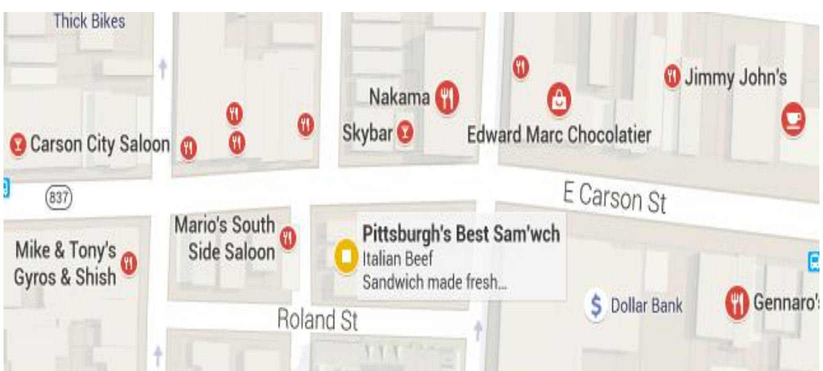


Figure 3-9. Land use of E Carson Street

3.1.2.2 Data collection plan details

A detailed data collection plan was completed before the actual data collection. Data collection details included a data collection schedule, data collection personal assignments and data collection forms.

Data collection schedule: The data collection for each location was conducted for one weekday and one weekend day for each business area (Forbes Avenue and East Carson Street) and two weekdays for each university area (Thackeray Avenue and Tech Street).

Data collection personal assignment: Two persons were assigned to collect data at each location. This required that each collector take charge of one kiosk and conduct the interviews and parking data collection for the area near the kiosk.

Data collection forms:

Table 3-2 provides the parking inventory summary sheet.

Table 3-2. Parking Inventory Summary Sheet

PARKING INVENTORY			
SUMMARY SHEET			
AREA OF INVENTORY____			
DATE OF INVENTORY____			
LOCATION	KIOSK#	COVER AREA IN LENGTH(FEET)	TOTAL COVER SPACES(20' PER VEHICLE)
		DATE______________ COMPLED BY_______________________	

After the data collection was completed, a parking inventory summary sheet was needed for summarize the parking data.

The Individual interviews contained 3 questions. The interview questions are provided in Table 3-3.

Table 3-3. Individual Interview

Individual interview	
I am a student from the University of Pittsburgh conducting research on how this parking system operates. Could you please answer a few brief questions to help with my research project concerning your parking experience today?	
1. Could you provide your car's license plate number or the location of your vehicle, this information will only be used to record how long you parked in the space:	
2. How much parking time did you purchase?	
3. What's your trip purpose to be here today? *A. for work B. for school* *C. for shopping D. others________*	
Thank you for your cooperation.	

The parking data collection form contains a simple map of the kiosk area. A field survey was done before the final data collection to measure the parking spaces for data collection map. These measurements were used to determine the number of available on street parking space. Table 3-4 is the data collection form.

Table 3-4. Data Collection Form

DATA COLLECTION FORM											
AREA OF DATA COLLECTION_______________											
DATE OF DATA COLLECTION_______________											
COLLECTOR_____________											
TIME	ACTUAL	LICENSE PLATE # PER PARKING TIME									
		1	2	3	4	5	6	7	8	9	10
09:00											
09:15											
………………											
16:30											
16:45											
17:00											

As stated above, the detailed data collection plan needed to consider every possible datum needed for analyzing and collect data as detailed as possible. Besides the parking time, locations and date, the data collection also needed to include the weather conditions for further study. Also, a field survey was needed before the final data collection to see if the plan needed any additional considerations to conduct the data collection.

3.2 DATA ANALYSIS

This section details the proposed methodology for data analysis and modeling. The data analysis included a variables analysis and modeling analysis after determining the most important variables.

3.2.1 Data analysis variables

For testing the model reliability, the data collected was divided into two parts. One part, which includes data of Forbes Avenue and Thackeray Avenue, was for analyzing the prediction model and another part, which includes data of Tech Street and E Carson Street, was for testing and adjusting this model.

As the locations for data collection were previously described, there are two business districts and two University districts. Thus, the research considered the parking location types as one of the variables that may impact the real time parking gap, which is defined as the difference between actual parking time and paid parking time.

In addition to the parking location type, parking hour limit, parking rate and trip purposes (school, work, shopping or others) may also be variables in the predictive model.

The duration of parking (0-1, 1-2, 3-4, 4+ hours) also may influence the parking occupancy and the parking time gap, which was the target object for this research.

The day of week may also count as one variable. The data were collected on weekdays and weekends and there may be some relationships between the weekdays parking and weekend parking characteristics.

In summary, the variables involved in the modeling formula were parking location types, parking hour limit, trip purposes, duration of parking and the day of week.

3.2.2 Data analysis methodology

After the data collection and summary, the relationships between each variable and the parking time gap were needed to be analyzed by creating the relationships between them and then find the possible parametric or other possible ways to connect them.

The target object for this research, parking time gap, is defined by the paid/actual parking time ratio. The ratio is calculated as the actual parking time duration measured divided by the paid parking time. Formula 3-1 defines this relationship.

$$\boldsymbol{{}^{actual}/_{paid}\ parking\ time\ ratio(r) = \frac{actual\ parking\ time(t_1)}{paid\ parking\ time(t_2)}} \quad \textbf{(3-1)}$$

The influences of the variables, for the data collected, may not impact the ratio one by one, but may impact the ratio as an accumulation process of all the variables measured. The accumulation process needed to be defined after the data collection. The original hypothesized prediction formula is shown as formula 3-2.

$$\boldsymbol{t_1 = t_2 F(a_0 x_0 + a_1 x_1 + \cdots + a_n x_n)} \quad \textbf{(3-2)}$$

where t_1 is the actual parking time, t_2 is the paid parking time, F(x) is the accumulation process of the variables, a_i is the influence coefficient and x_i is the variables factor.

3.2.3 Data analysis

A linear relationship between parking time ratio and all the variables was hypothesized to exist. The initial attempt to establish the relationship was to use formula 3-2 for all of the variables for which data was collected. The following provides the results of this analysis.

A regression analysis of evaluating the relationship between the parking time ratio and all the variables, which includes trip purpose, parking location type, pay duration, parking limit hour, parking rate and the day of the week, was conducted. All the variables are categorical data, except pay duration, which is defined by paid parking time.

The initial analysis using all of the variables resulted in an R^2 of .04069 which is too small to establish a relationship using all of the variables, which means the relationship between parking time ratio and all the variables is not significant. R^2 means how the model reflects the situation in the real world application. The acceptable value in analysis is typically larger than 0.8 in transportation analysis.[8]

After the analysis of all the variables did not yield a strong relationship, a process of analysis between parking time ratio and each variable individually was conducted. Table 3-5 is a summary table of the results of this analysis.

Table 3-5. Summary of analysis results

Variables	R-Square
The day of week(weekday or weekend)	0.001
Pay duration	0.018
Parking location type	0.005
Parking limit time	1.67E-05
Parking rate	0.005
Trip purpose	0.021

The results again did not provide a strong relationship between any of the individual variables and the parking time ratio. The reason for this was hypothesized that the overpaid parking time has too large of a time increment, one hour, to establish a relationship to any of these variables. Overpaid time is defined as formula 3-3.

$$\boldsymbol{overpaid\ time = paid\ parking\ time - actual\ parking\ time} \tag{3-3}$$

Therefore the overpaid parking time data was refined and divided into fifteen minutes intervals. The analysis of comparing the paid time ratio to all of the variables together (the overpaid time is 0-15 minutes) was again performed.

This analysis resulted in a higher R^2 of .3289 and also the variable pay duration has a small P value of 0.0003, which means the pay duration has a significant impact on parking time ratio while other variables have little impact on parking time ratio.

The P value is defined as the probability of obtaining a result equal to or “more extreme” than what was actually observed, assuming that the hypothesis under consideration is true. Before the test is performed, a threshold value is chosen, called the significance level of the test, traditionally 5% or 1%. In this research, 1% was selected as the threshold value. If the P value is equal to or smaller than the significance level, it suggests that the observed data are inconsistent with the assumption that the null hypothesis is true and thus the hypothesis must be rejected.[9]

Because of these results the variable pay duration variable was focused on for further analysis. Then the linear regression analysis between parking time ratio and pay duration was also conducted.

The results showed that although the p value of 2.074e-07 is reasonable for it is smaller than 1%, which is acceptable. Also a small R^2 of 0.297 still results which cannot provide a significant linear relationship between parking time ratio and pay duration.

The existence of abnormal data was then explored which maybe impacting the results. Figure 3-10 shows the abnormal data, which means the data had a large scatter with all the other data and would influence the model accuracy. According to the analysis results in Figure 3-10 some of the data needed to be rejected.

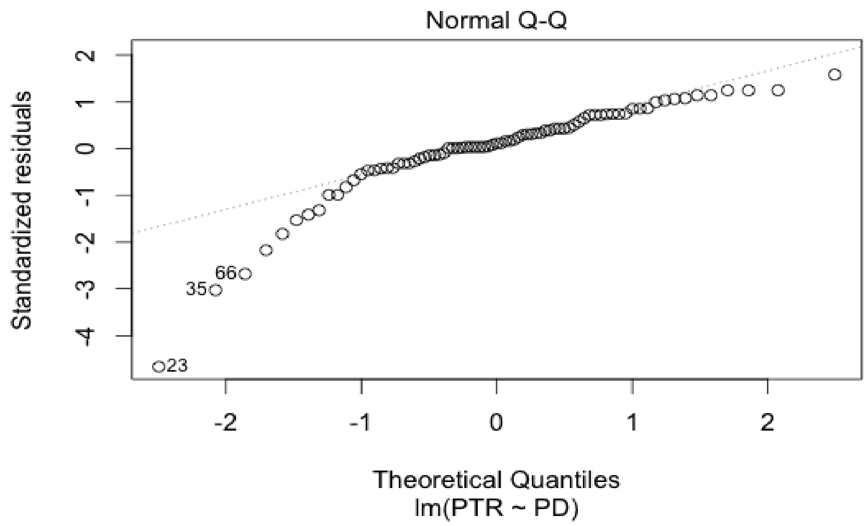


Figure 3-10. Normal Q-Q analysis results of abnormal data

Normal Q-Q plot is a probability plot, which is a method for comparing two probability distributions by plotting their quantiles against each other.[10] If the points all lie on one line, then the two distributions being compared are similar. The analysis process software R marked the abnormal data by displaying the row number by the side of the data points. The abnormal data were eliminated based upon the results shown in the figure 3-10 , they are data points 23, 35 and 66 rows in the summary data sheet(Abnormal data were highlighted in the Appendix A). These data points were eliminated from the data set for further analysis.

A major concern in the application of regression analysis, heteroscedasticity can invalidate statistical tests of significance that assume that the modeling errors are uncorrelated and uniform-hence that their variances do not vary with the effects being modeled.[11] Heteroscedasticity should also be considered and has an impact in this result.

After avoiding the influence of the abnormal data and heteroscedasticity, the results as shown in Figure 3-11 shows the resulting relationship between PTR and PD.

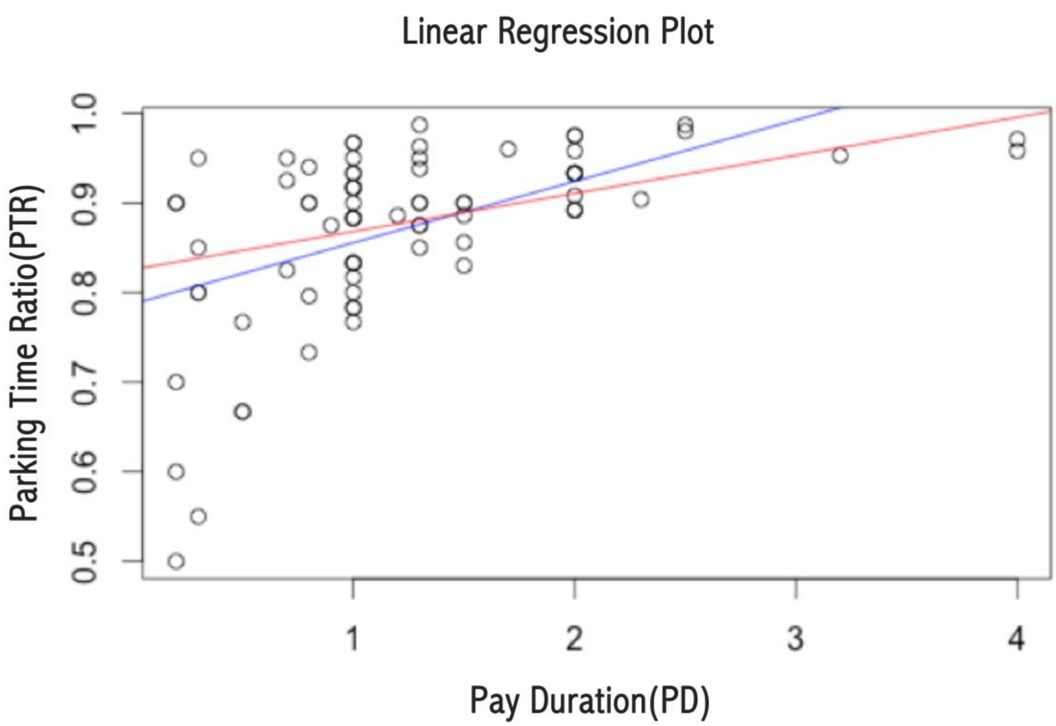


Figure 3-11. Linear regression diagram (original model in blue, adjustment model in red.)

From the result it can be can seen that, the R-square is .336, which means the linear relationship between parking time ratio and pay duration is still not significant.

The same model was applied to the data with overpaid time(15-30min) and overpaid time(30-60). Table 3-6 shows the results of this analysis using the overpaid time increments of (0-15min), (15-30 min), and (30-60 min). There is a small part of data in which the overpaid time is more than 1 hour, which should be considered as special data case and was not used.

Table 3-6. Adjustment results of each kind of overpaid parking time to the PTR

Overpaid time	R-square	Ad R-square	p-value
0-15 min	0.297	0.336	4.008e-08
15-30 min	0.544	0.643	3.215e-08
30-60 min	0.532	0.592	0.001289

In summary, the R^2 is again too small to prove that the overpaid parking time has a linear relationship with pay duration and all the other variables. Under this circumstance, another method of analyzing the data was considered.

The underpaid parking part of the data collection was then added to then data analysis and was considered in the impact on parking occupancy. At this point in the process it was hypothesized that the parking occupancy is not only influenced by the overpaid parking, but also by the underpaid part for this part influences the accuracy of prediction.

The same analysis method was again used to for the linear regression model and figure 3-12 provides the results.

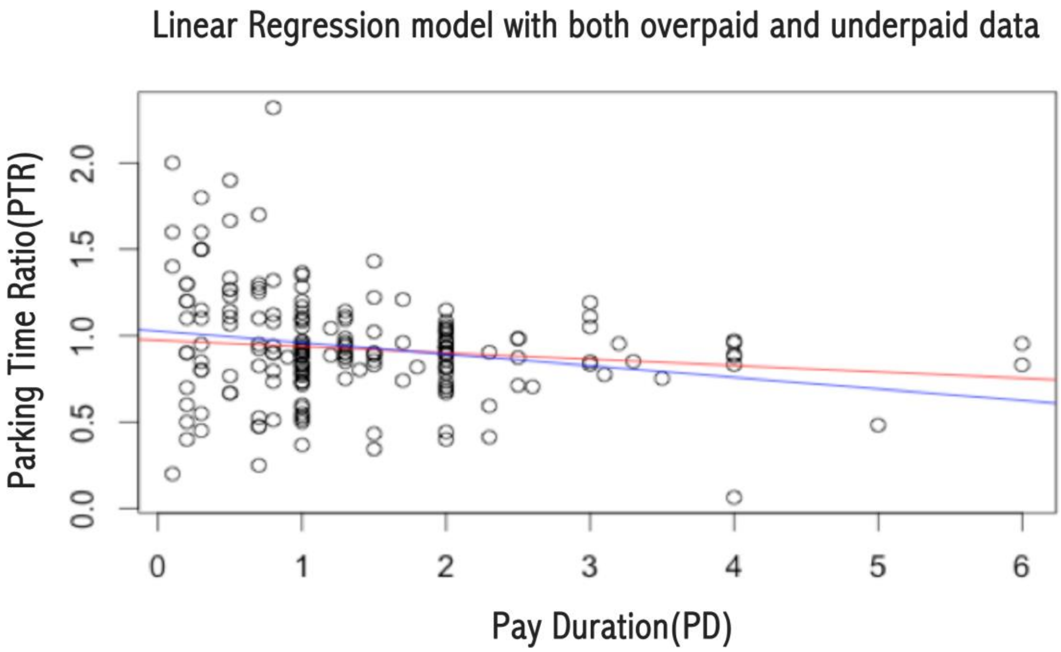


Figure 3-12. Linear regression plot using both overpaid and underpaid data

For the involvement of human behaviors, this R-square of .0485 is reasonable if the p-value is small enough. But a further adjustment is still needed because of a small R-square. Based upon this analysis of multiple variables and data refinement the relationship between the parking time ratio and paid parking duration was determined to be the most significant from the data collected. The linear regression formula recommended for further testing was as follows.

$$\boldsymbol{parking\ time\ ratio = -0.0369 * pay\ duration + 0.9737} \qquad \textbf{(3-4)}$$

Although the model in 15 minutes interval with the pay duration as the variable has a higher R-square, it was still considered not efficient enough not only for the small R-square, but also for that the data did not contain the underpaid part which is also a main consideration in analysis and prediction. This model has a small R-square, but the p-value is small enough for further adjusting. Thus, this model was selected for further study.

Some conclusions while analyzing the data and developing the model relationship are as follows.

1. The human behavior is much more complicated than what we expected. The data of a scheduled event, such as a class or a scheduled appointment, may have a better result for the events have a certain ending time. The data of an unscheduled event, such as shopping or some other events that have no certain ending time, should be too random to predict.
2. From the table 3-6 we can see, the overpaid time (15-30 min) has the highest R^2 than the other time intervals in the model (parking time ratio ~ pay duration). In this case, overpaid time (15-30 min) is more predictable then the other cases.
3. The data of the overpaid time interval of (0-15 min) was more prevalent than the other cases. Most of parkers demonstrated that they purchased their parking time to match the anticipated needed duration well.

3.3 SUMMARY

In this section, the data collection and analysis was presented. The data collection plan description included the details of the data collection locations and the data collection plan execution. The data analysis and modeling part included the variables that may influence the parking time ratio and the estimated prediction formulas that were tested. Parking time ratio defines the target object, parking time gap, and is defined as the actual parking time duration measured divided by the paid parking time. The analysis of the various variables for which data was collected established a relationship between the parking ratio and the paid duration. This relationship was then tested to determine if it could reliably predict the parking duration. The following is a flow chat of analysis process performed.

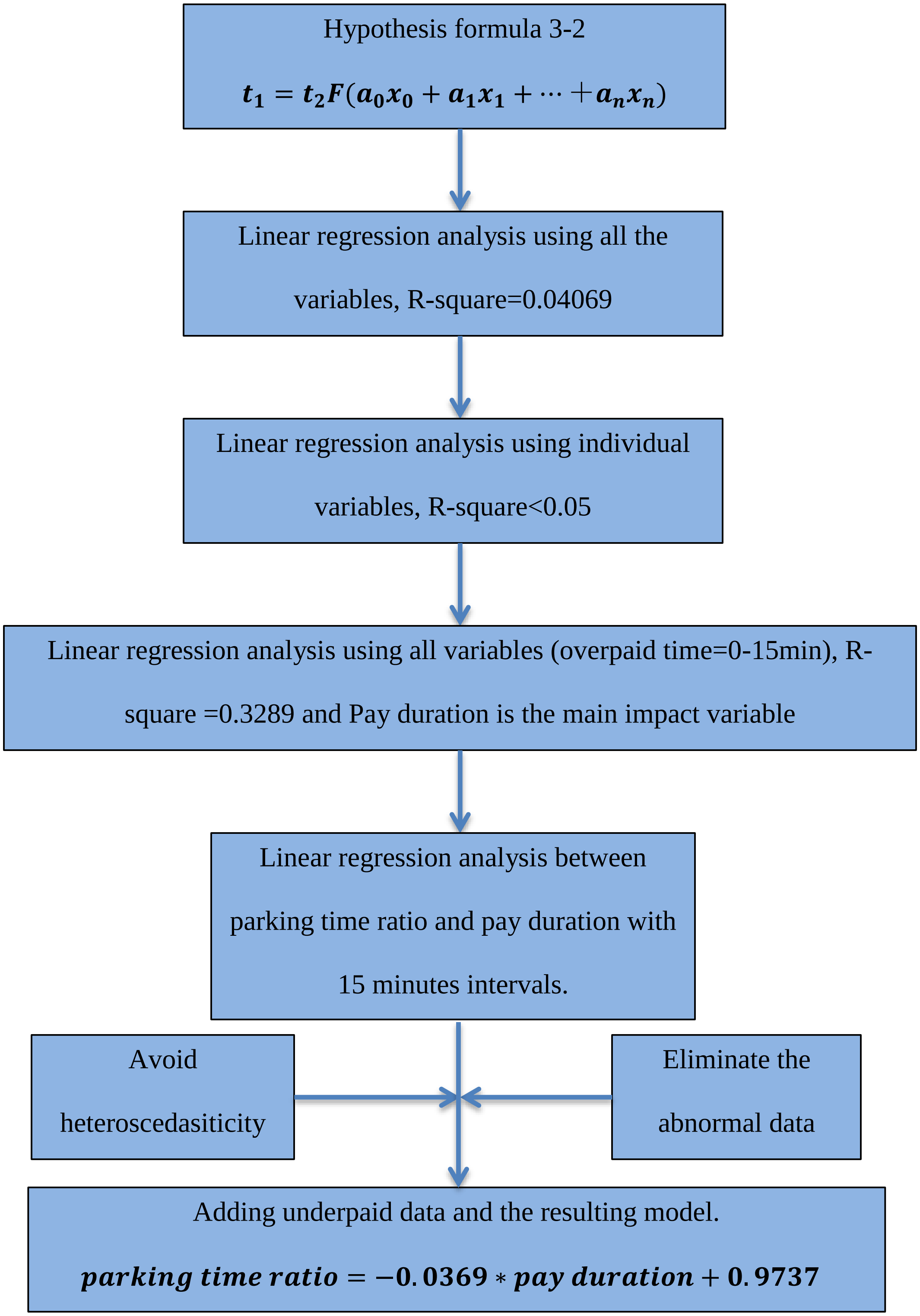


Figure 3-13. Data analysis flow chat

4.0 DEVELOPMENT OF PREDICTION MODEL

This chapter discusses the development of the final prediction model, which includes the establishment of the prediction formula, model testing and model adjustment.

4.1 PREDICTION MODEL AND TEST

After analyzing the data in chapter 3, the prediction model was finally determined as formula 3-3. Here again is the linear regression model formula to be tested.

$$\boldsymbol{parking\ time\ ratio = -0.0369 * pay\ duration + 0.9737} \qquad \textbf{(3-3)}$$

The method of K-fold cross validation was used to test the model. K-fold cross validation is a methodology that randomly partitions the original data sample into k equal sized subsamples and uses k-1 subsamples as training data to predict the rest part. This kind of methodology is suitable for small database testing, thus this method was adopted in this research. The 5-fold validation was used to test the model, which means the original data sample was randomly partitioned into 5 equal sized subsamples and the process used four subsamples without replacement as the training part of the data. Figure 4-1 shows the fit results between predicted value and the actual value.

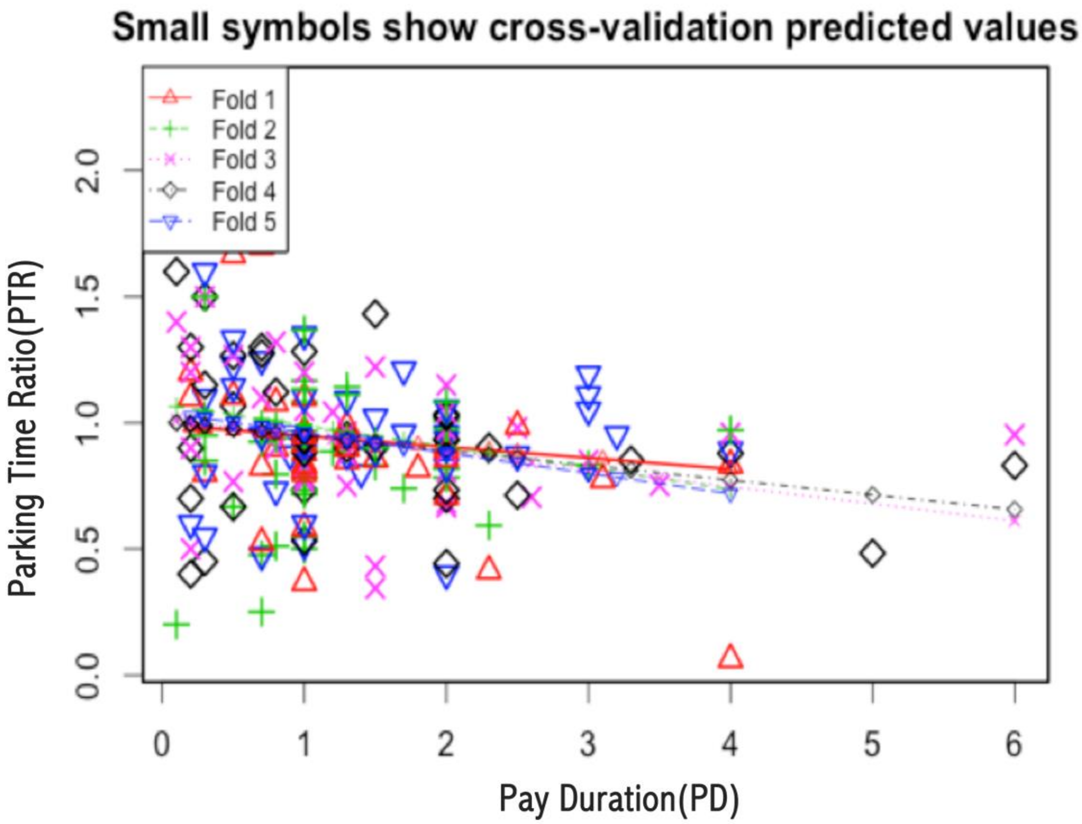


Figure 4-1. 5-fold cross validation results using pay duration and parking time ratio data

In the figure 4-1, PTR is parking time ratio and PD is pay duration. One fold means the process used 4/5ths of the whole data as training part, predicted the rest using 1/5th of the data and compared the results. The process repeated the analysis section 5 times to use all the data.

In the testing results the cvpred value, which is defined as the prediction value of cross validation testing, and the PTR value, which is the actual parking time ratio value. In the testing, the process used the Mean Square (MS), which is defined by formula 4-1 to evaluate the model accuracy.

$$MS = \frac{1}{n}\sum_{i=1}^{n}(predicted_i^2 - observed_i^2) \quad \textbf{(4-1)}$$

The MS value, which is a measure of how close a fitted line is to data points and the smaller the value is the closer the fit is, is 0.0976, which means the prediction result is acceptable with a small MS value. To avoid the reason of the data influence, which means the data is really small that may relate to a small MSE value, the research also calculated the Root Mean Square Error(RMS errors) to test the accuracy of prediction.

$$\boldsymbol{RMSerrors = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(predicted_i - observed_i)^2}} \qquad \textbf{(4-2)}$$

The RMS error is 0.3124 in this case, which means about 69% of the data was predicted accurately.

The results of testing were acceptable which means that formula 3-3 predicts the parking ratio based upon the parking duration using some of the data to establish the relationship and using the remaining data to test the relationship, but for a more accurate model of prediction, a further adjustment method was applied to the model.

4.2 MODEL ADJUSTMENT

After the testing of the model, the research found the testing results were acceptable, but the model still needed an adjustment to improve the prediction accuracy for the R-square of the model because it is too small to prove the relationship between parking time ratio and pay duration.

When the other linear regression models were adopted to test the data and the process found that the model without a Y intercept gave better results. The model was changed as figure 4-2 shows.

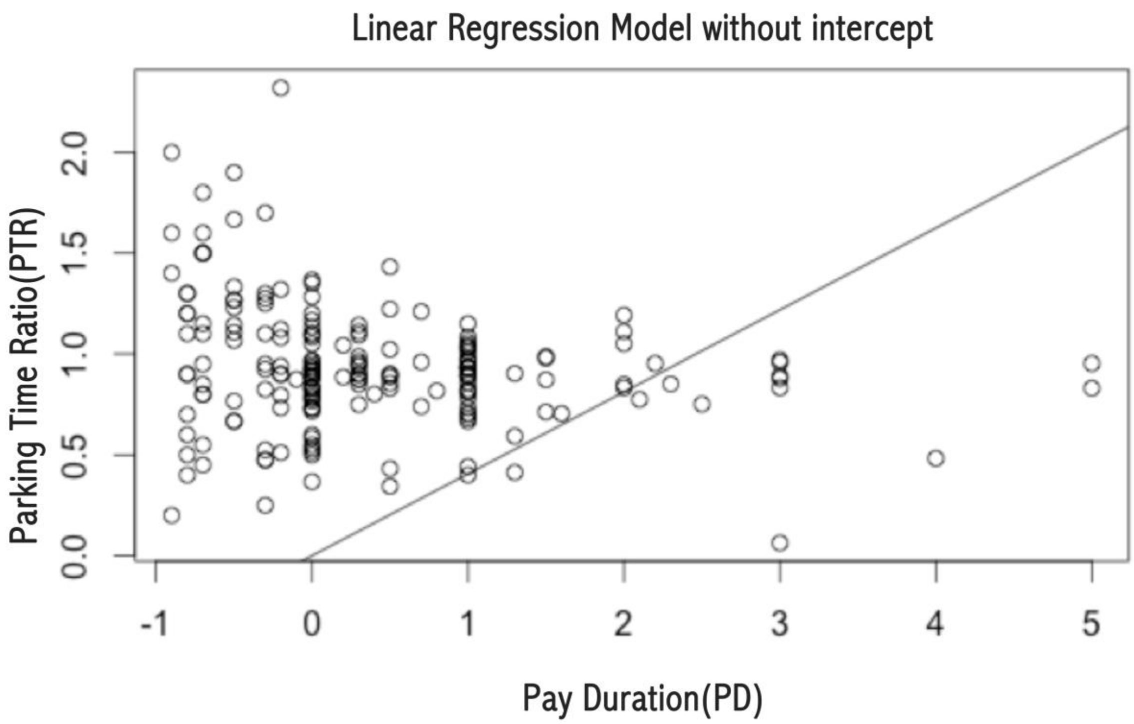


Figure 4-2. Linear regression without intercept results

The linear regression formula was then adjusted to formula 4-3.

$$\boldsymbol{parking\ time\ ratio = 0.4062 pay\ duration} \qquad \textbf{(4-3)}$$

These results showed that, the model after adjustment had a higher R-square of 0.509 and a smaller p-value.

The k-fold cross validation was again applied to determine if this model predicts the results better. The Mean Square(MS) was then0.484, while 0.0976 resulted before the adjustment. This may be because of the data influence as stated in section 4.1. To avoid this influence, the research evaluated the result by calculating Root Mean Square Error (RMSerror) also and a value of 0.2347 resulted, while the model before adjustment had a value of 0.3124.

Figure 4-3 shows the results of 5-fold cross validation of the model after adjustment.

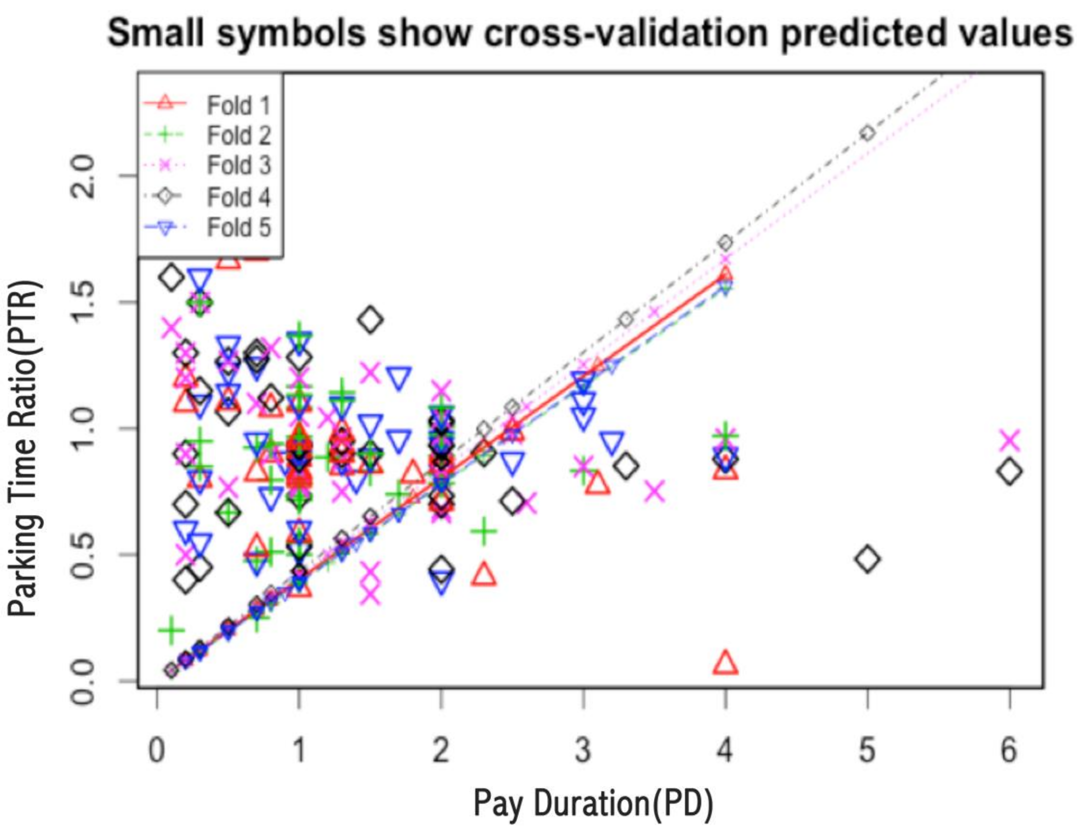


Figure 4-3. 5-fold cross validation of the model after adjustment

Table 4-1 is a summary of the differences between two models.

Table 4-1. Differences between the model before adjustment and the model after adjustment

	Model before adjustment	Model after adjustment
MS	0.0976	0.484
RMSerror	0.3124	0.2347
R-square	0.0485	0.509
P-value	<0.00221	<2e-16

In summary, the model after adjustment improves the accuracy of prediction and relates the parking time ratio with pay duration better than before. Therefore the adjusted formula was accepted at a more accurate predictor.

Based on the success of this analysis, the other variable relationships were revisited individually with the parking ratio. The purpose was to determine if there was any other variables had impacts on the parking ratio and could they be used as the variables in the new model. Table 4-2 is the R-square and P-value results of the variables analyzed separately.

Table 4-2. R-square and P-value results of each variable

Variables	R-square	P-value
Parking limit hour	0.898	<2e-16
Parking rate	0.898	<2e-16
The day of the week	0.898	<2e-16
Parking location type	0.898	<2e-16
Trip purpose	0.901	<2e-16
Pay duration	0.509	<2e-16

These results confirm that there was a strong relationship between the parking time ratio and all the variables. The analysis of all the variables together was again performed.

From the results it can be concluded that the R-square of .906 and p-value are good acceptable but only the trip purpose and parking duration have passed the t-test. The further analysis then isolated these two variables.

From the results, it was concluded that the final model should use trip purpose and pay duration as variables. Because of the R-square of .905 and p-value of smaller than 2e-16, the results were acceptable to illustrate the relationship between parking time ratio and two variables.

The revised and improved model formula is formula 4-4.

$$\boldsymbol{parking\ time\ ratio = -0.0770 pay\ duration + \begin{cases} \begin{cases} if\ trip\ purpose\ is\ A, use\ 1.0735 \\ if\ trip\ purpose\ is\ B, use\ 1.0908 \end{cases} \\ \begin{cases} if\ trip\ purpose\ is\ C, use\ 1.1700 \\ if\ trip\ purpose\ is\ D, use\ 1.0133 \end{cases} \end{cases}} \quad \textbf{(4-4)}$$

For this formula trip purpose A is for work, trip purpose B is for school, trip purpose C is for shopping and trip purpose D is others.

The method of K-fold cross validation was also applied to test if the final model has an accurate prediction result. Figure 4-4 shows the results.

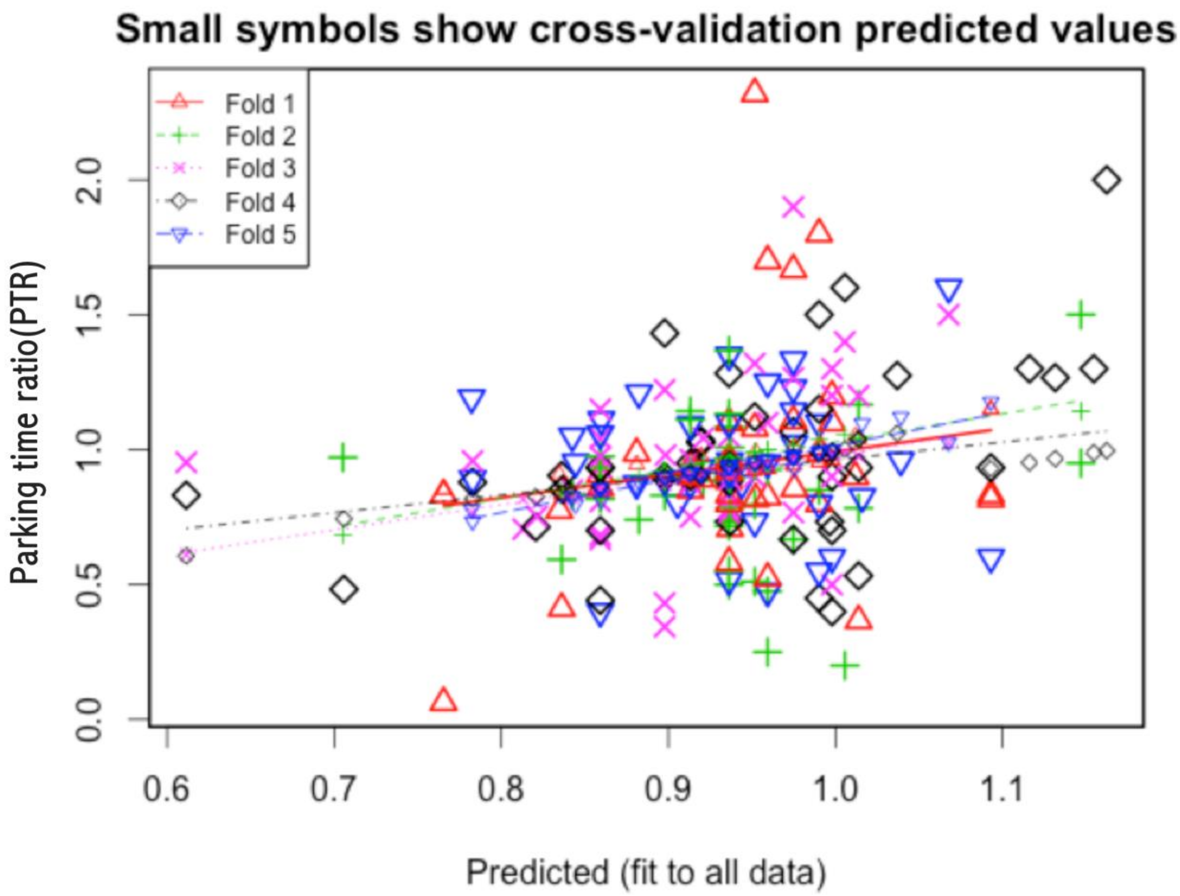


Figure 4-4. 5-fold cross validation results of the final model

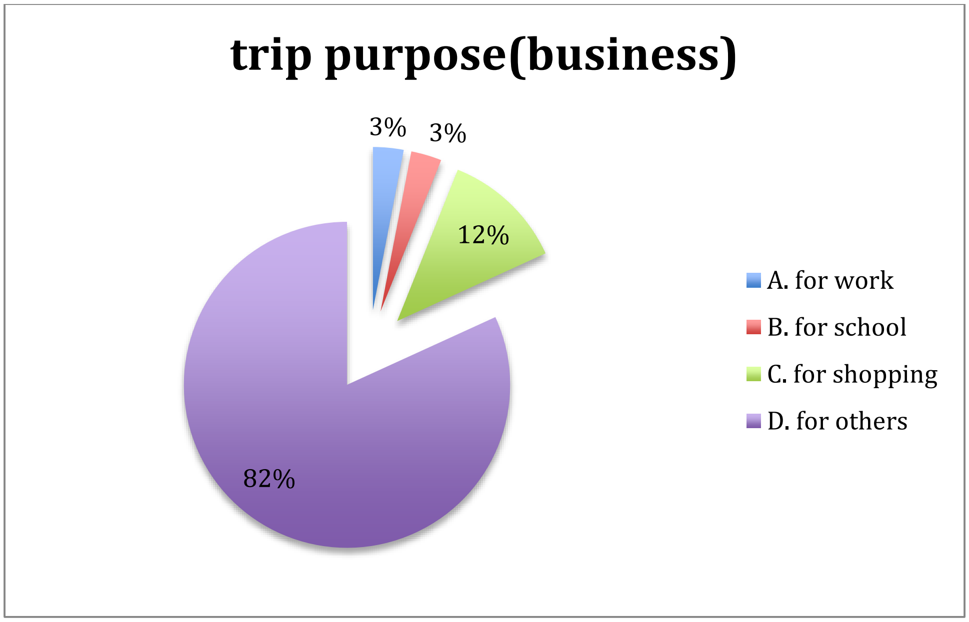
Table 4-3 is the summary results of all the models illustrated before.

Table 4-3. Comparison of the model results

	Model before adjustment	Model after adjustment	Final model
MS	0.0976	0.484	0.101
RMSerror	0.3124	0.2347	0.3186
R-square	0.0485	0.509	0.901
P-value	<0.00221	<2e-16	<2e-16

In summary, the final model has a reasonable R-square to relate the variables trip purpose and pay duration to parking time ratio, but the RMS error (0.3186) is larger than the model after adjustment (0.2347).

As a variable, trip purpose can be obtained related to the parking location type as to where the kiosk is located. Figure 4-5 shows the proportions of each kind of trip purpose from the data collected by data collection areas.



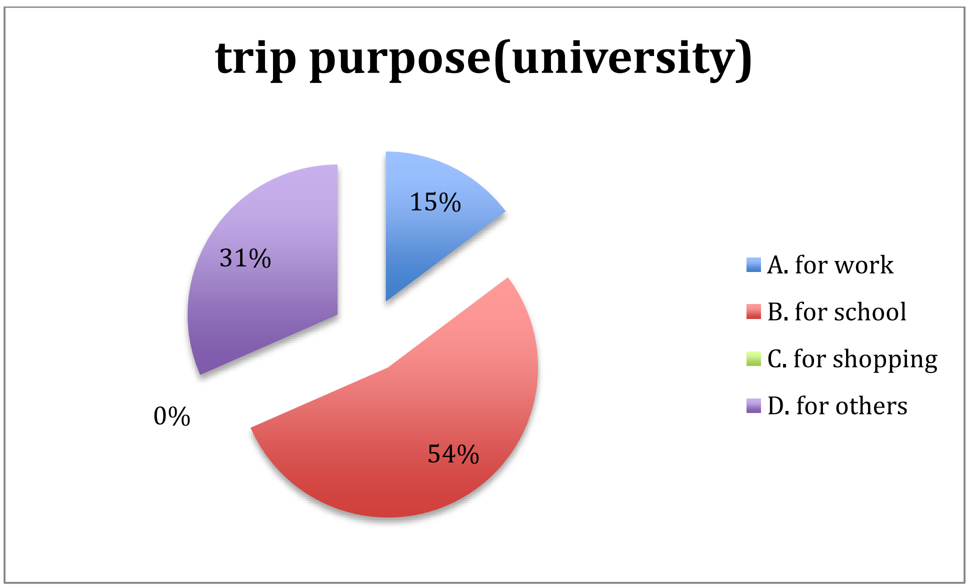


Figure 4-5. Proportions of each kind of trip purpose in different parking location type.

Table 4-4 is a summary of the amount of each kind of trip purpose in different parking location types.

Table 4-4. The number of users of each trip purpose in different parking location type

Trip purpose type	Business(capita)	University(capita)
A. for work	3	14
B. for school	3	51
C. for shopping	12	0
D. for others	81	30

For working and for school trips it they can be categorized as a similar types of trip purposes for the model. For shopping and for other trips it can be concluded that they are also similar types as well. It can be concluded that there are some trip purposed D in the university areas, most of them are meeting or appointment type trips and in some situations these are also the same kind of trip purpose as for school.

In summary, the formula can be simplified to define the business area as trip purpose D and university area as trip purpose B, which simplifies the model as below.

$$\boldsymbol{parking\ time\ ratio = -0.0770 pay\ duration + \begin{cases} if\ business\ area, 1.0133 \\ if\ university\ area, 1.0908 \end{cases}} \quad \textbf{(4-5)}$$

Thus, with the variables of pay duration and parking location type, the formula can easily predict the parking time ratio.

4.3 SUMMARY

In this chapter, the research established the prediction model, tested the model and adjusted the model to improve the accuracy of prediction. The model before adjustment had a smaller Mean Square value, but the other parameters, such as R-Square, p-value and RMS errors, were not as acceptable as the model after adjustment. The model after adjustment improved the accuracy from 69% to 77% and also had a more acceptable R-square value to relate parking time ratio with pay duration. After adding the other variables into the model, the R-square became more reasonable as the final model showed that the R-square was 0.905. Finally, the research defined how to use the variable trip purposes and redefined the final model as .

$$\boldsymbol{parking\ time\ ratio = -0.0770 pay\ duration + \begin{cases} if\ business\ area, 1.0133 \\ if\ university\ area, 1.0908 \end{cases}} \qquad \textbf{(4-5)}$$

The flow chat below shows the process of the model testing and adjusting.

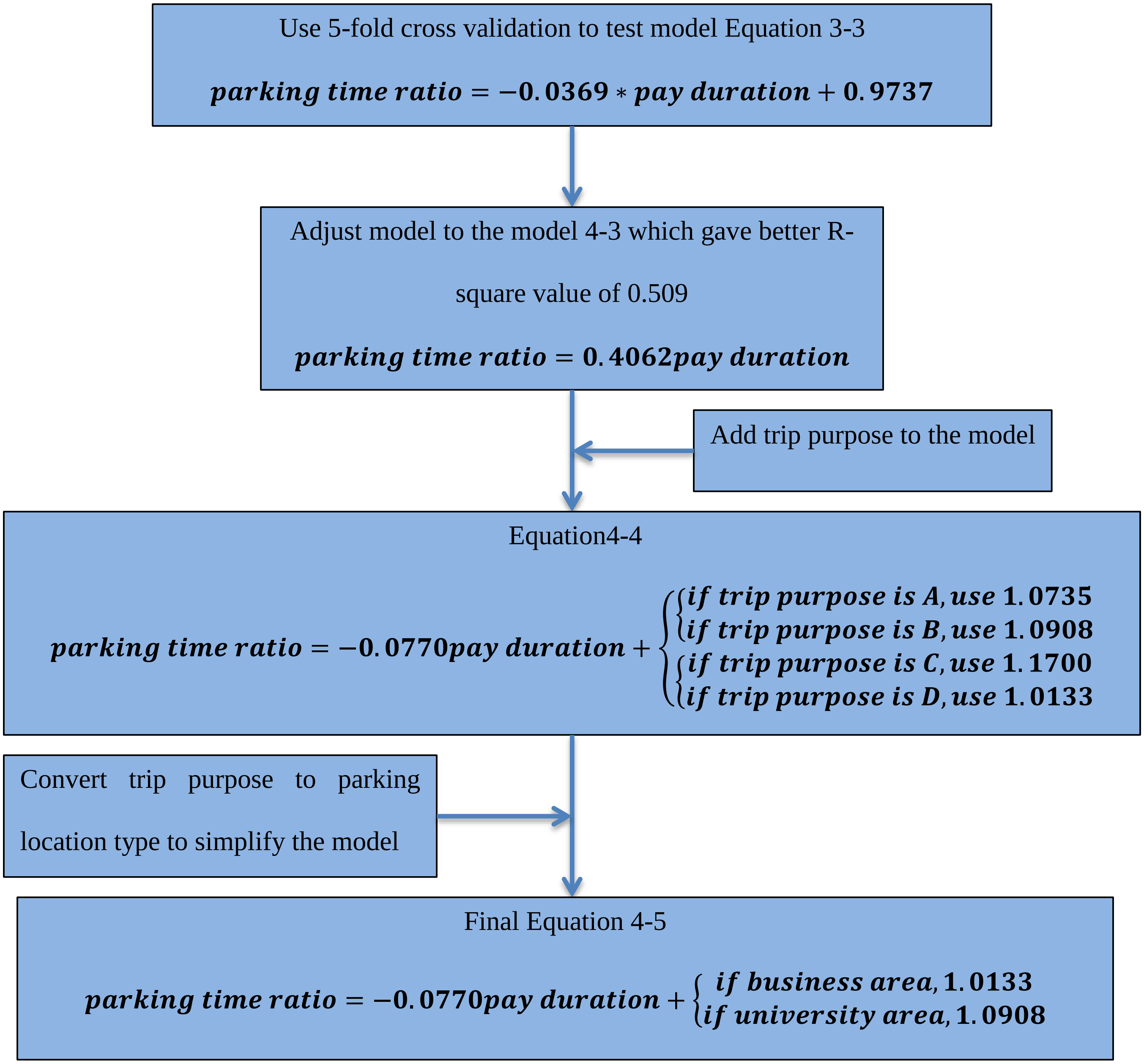


Figure 4-6. Model test and adjustment flow chat

5.0 CONCLUSIONS AND SUMMARY

This chapter summaries the results, determines whether the results match the hypothesis and gives the author's opinions on future research.

5.1 SUMMARY OF THE RESUTLS

In order to test the hypothesis, the researcher collected the related data and established a model to predict the parking time ratio, the ratio of actual parking time and pay parking time. This is the key information that would be used to predict the number of available parking spaces based on the time purchased and trip purpose, which can be deduced from the land use characteristics surrounding the kiosk area.

As stated before, the model before adjustment had a smaller Mean Square value, but the other parameters, such as R-Square, p-value and RMS errors, were not as acceptable as the model after adjustment. The model after adjustment improved the accuracy from 69% to 77% and also had a more acceptable R-square value to relate parking time ratio with pay duration. After adding the other variables into the model, the R-square became more reasonable as the final model showed that the R-square was 0.905. At last, the research defined how to use the variable trip purpose and obtain the final model as formula 4-5.

$$\boldsymbol{parking\ time\ ratio = -0.0770 pay\ duration + \begin{cases} if\ business\ area, 1.0133 \\ if\ university\ area, 1.0908 \end{cases}} \quad \textbf{(4-5)}$$

With the final prediction model, the methodology can use the variables of pay duration and trip purpose to predict the parking time ratio. Using parking time ratio, methodology can obtain the actual parking time by using formula 5-1.

$$\boldsymbol{actual\ parking\ time = pay\ parking\ time * parking\ time\ ratio} \quad \textbf{(5-1)}$$

The following is an example illustrating how to use the model.

Example: A parking user purchased 2 hours for parking in a business area. Using the model to predict the actual parking time provides the following result:

Result: Use formula 4-6 obtains the parking time ratio.

From the example, the pay duration is 2.0 hours and the trip purpose incremental time from the formula in business area is 1.0133

$$parking\ time\ ratio = -0.0770 * 2 + 1.0133 = 0.8593$$

Then, using formula 5-1 the actual parking time is calculated as the following:

$$actual\ parking\ time = 2h * 0.8593 \approx 1.7\ h$$

As can be seen in the result, the actual parking time is 18 minutes less than the pay parking time.

Another result of the research is that the model predicts a general actual parking time for a parking area instead of an individual actual parking time for a specific space, which can be applied to the area surrounding the kiosk instead of individual parking spaces. In this case, the answer to the example problem shows that most of the parking users, who buy 2 hours for parking in a business area, use 1.7 hours and leave 18 minutes for the next parking users.

And also with the time saved table, we can estimate the additional revenue. We assume that with this information, the parking space can be used right after the previous user and we use 1 dollar for 30 minutes as example.

Table 5-1 and table 5-2 are summaries of the time saved and additional revenue in 15 minutes intervals within 2 hours in different parking location type.

Table 5-1. Summary of time saved in business area

Pay parking time	Parking time ratio	Actual parking time	Time saved(min)	Additional revenue(dollar)
00:15	0.994	00:14	1	0.03
00:30	0.975	00:29	1	0.03
00:45	0.956	00:42	3	0.10
01:00	0.937	00:56	4	0.13
01:15	0.917	01:08	7	0.23
01:30	0.898	01:20	10	0.33
01:45	0.879	01:32	13	0.43
02:00	0.860	01:43	17	0.57

Table 5-2. Summary of time saved in university area

Pay parking time	Parking time ratio	Actual parking time	Time saved(min)	Additional revenue(dollar)
00:15	1.072	00:16	-1	-0.03
00:30	1.052	00:31	-1	-0.03
00:45	1.033	00:46	-1	-0.03
01:00	1.014	01:00	0	0
01:15	0.995	01:14	1	0.03
01:30	0.975	01:27	3	0.10
01:45	0.956	01:40	5	0.17
02:00	0.937	01:52	8	0.27

Here we assume one parking space is used from 8:00 am to 6:00 pm. All the users purchased 2 hours for parking. Thus, we get additional revenue per day per parking space as

In business area

$$0.57 * 5 = 2.85$$

In university area

$$0.27 * 5 = 1.35$$

The reason for the parking time ratio to be larger than 1 is that there are some underpaid parking situations in the database as well, which means the actual parking time is longer than paid parking time, especially the short use of parking spaces as table 5-2 shows. The results show that the parking users in university area underpaid their parking time more often. We can see from the results that they only use one minute longer than paid time. The reason for this is that the parking users in university area have a certain schedule, for example the class schedule, and thus the users know how much time should be purchased for parking.

The researcher used 2 hours as a limit because 2 hours is the shortest parking limit hours in data collection areas. From the summary table, it can be seen that 17 minutes in business area and 8 minutes in university area are saved.

With the time saved information, we can convert this to parking occupancy calculation for provide real-time parking occupancy information.

The model predicts a general behavior of parking users. For example, the users who purchased 2 hour in university area all intend to leave 8 minutes earlier.

The prediction is based on the real-time variable pay duration, which can be obtained by kiosks system. Here we use an example to illustrate how to convert this information.

At 10:00 am, a parking user parked in university area and purchased 2 hour. With the model, we know, the user intend to leave 8 minutes earlier from table 5-2. Thus, we assume this parking space can be available at 11:52 am.

5.2 CONCLUSIONS

The hypothesis that a real time parking availability information systems for on street parking that uses pre-paid parking could increase occupancy if a prediction model was developed to provide information on availability based a predicted parker's departure time is confirmed by the establishing the prediction model in this research. The model can predict the actual parking time of the parking users using the variables that includes the pay parking time information and parking location areas types.

5.3 RECOMMENDATIONS FOR FUTURE RESEARCH

This section provides the researcher’s recommendations on future study of this topic.

5.3.1 Relate the model to the occupancy prediction

Researchers could focus on the occupancy prediction in the future study for practical use of the model. For now, the process only has the saving time in the model from the data that was collected and a further analysis of how this information can be related to the occupancy prediction is needed before adopting this model into practical use. For example, the process can predict the saving time, but it cannot determine how many spaces are left in the parking area. As a real-time parking information system for parking users, the system needs to provide the occupancy information to users within a specified area. The current kiosk system does not have a defined area associated with payment process. Thus, the future studies can focus on the occupancy prediction using the saving time model in this research.

5.3.2 Reconsider of variable trip purpose

In this research, the trip purpose was converted to parking location type. For it is kind of special variable that can only be obtained by individual interview and hard to category in too much details. Thus in this research, we used others to illustrate other types of trip purpose. But for a more rigorousness analysis, using a more detailed trip purpose type should be developed and be studied of their impact on overpaid parking. There are also some variables that may influence the trip purpose, such as parking location type, which means in some cases, the trip purpose can also be predicted.

5.3.3 The impact of parking enforcement

The parking enforcement should be a reason for overpaid parking, but this kind of data is difficult to obtain, for too random. And this is also based on psychology of parking users. For example, how long I should purchase to avoid ticket, someone may pay just 25 cents while others may pay 1 dollar or more. The impact of this variable is difficult to be illustrated by the simple model and a big database is needed for this study. But a further study is needed.

5.4 IMPLEMENTATION OF THE MODEL

The model is based on the real time variable, pay duration. While we have this variable, the model can be used in a real- time occupancy information system. From the parking inventory, the occupancy data can be obtained.

Here, we use an example to illustrate.

In a business area, we assume there are 4 spaces for parking at 10:00 am. And here are two tables to illustrate the parking availability. We provide users parking information in a 15 minutes interval.

Table 5-3. Parking users real time parking information

Parking user#	Arrive time	Pay duration	Predict leave time
1	10:00	2	11:43
3	10:30	1	11:26
		1.5	11:50
		2	12:13

Table 5-4. Real-time parking occupancy information

Time interval	Available
10:00-10:15	3
10:15-10:30	3
10:30-10:45	0
10:45-11:00	0
11:00-11:15	0
11:15-11:30	1
11:30-11:45	2
11:45-12:00	3
12:00-12:15	4

And the detailed trip purpose should be revised for adopting the model to the big area.

Based on the detailed land use type of the data collection area, we should test every kind of trip purpose to category them. For example, if the influence factors of medical care and restaurant are very similar, which means the difference between these two factors is no larger than a value, like 0.01, we should consider these two as one category. We should consider trip purpose as detailed as possible, for it is a main influence variable to the model.

With the detailed trip purpose, we can break one area into small observation area to obtain the trip purpose variables for model use.

APPENDIX A

ORIGINAL DATA

In this section, all the data that was collected are presented.

Parking Inventory

PARKING INVENTORY			
SUMMARY SHEET			
LOCATION	KIOSK#	COVER AREA IN LENGTH(FEET)	TOTAL COVER SPACES(20' PER VEHICLE)
Forbes AVE	2	300	12
Thackeray AVE	2	400	15
Tech St	2	500	20
E Carson St	2	500	20
		DATE__9/4/2015___	COMPLED BY___Chang Liu______

Business area

Forbes Avenue-09/18/2015, Friday

CAR#	Arrive	Leave	Actual	Paid	Trip purpose	Overpaid time
1	10:59	12:54	01:55	02:00	B	00:05
2	11:20	11:39	00:19	00:20	C	00:01
3	10:21	10:30	00:09	00:10	D	00:01
4	11:03	12:01	00:58	01:00	D	00:02
5	09:00	09:37	00:37	00:40	D	00:03
6	11:33	11:50	00:17	00:20	D	00:03
7	10:18	10:25	00:07	00:10	D	00:03
8	09:30	09:31	00:01	00:05	D	00:04
9	09:02	10:18	01:16	01:20	D	00:04
10	12:07	12:30	00:23	00:30	D	00:07
11	11:57	12:50	00:53	01:00	D	00:07
12	10:00	11:12	01:12	01:20	D	00:08
13	12:06	12:17	00:11	00:20	D	00:09
14	11:09	11:42	00:33	00:45	D	00:12
15	11:52	12:38	00:46	01:00	D	00:14
16	10:50	11:50	01:00	01:20	D	00:20
17	10:19	10:38	00:19	00:40	D	00:21
18	09:47	10:10	00:23	00:45	D	00:22
19	09:15	09:50	00:35	01:00	D	00:25
20	10:50	12:04	01:14	01:40	D	00:26
21	11:52	12:59	01:07	01:35	D	00:28
22	09:55	10:25	00:30	01:00	D	00:30
23	11:20	11:30	00:10	00:40	D	00:30
24	09:40	10:13	00:33	01:20	D	00:47
25	09:30	10:00	00:30	00:20	D	00:10
26	09:40	09:48	00:08	00:05	D	00:03
27	11:50	12:13	00:23	00:20	D	00:03
28	09:50	12:13	02:23	02:00	D	00:23
29	10:20	11:11	00:51	00:40	B	00:11
30	10:36	11:30	00:54	00:50	D	00:04
31	09:42	11:45	02:03	02:00	A	00:03
32	11:48	13:09	01:21	01:00	D	00:21
33	10:15	12:05	01:50	01:30	D	00:20
34	10:35	11:41	01:06	00:50	D	00:16
35	11:17	11:53	00:36	00:20	D	00:16
36	11:43	11:50	00:07	00:05	D	00:02
37	11:59	12:12	00:13	00:10	C	00:03

Forbes Avenue-09/26/2015, Saturday

CAR#	Arrive	Leave	Actual	Paid	Trip purpose	Overpaid time
1	10:17	11:13	00:56	01:00	B	00:04
2	09:17	09:33	00:16	00:20	D	00:04
3	13:17	13:33	00:16	00:20	D	00:04
4	09:58	11:13	01:15	01:20	D	00:05
5	12:33	13:27	00:54	01:00	D	00:06
6	12:41	14:33	01:52	02:00	D	00:08
7	11:38	12:59	01:21	01:30	D	00:09
8	09:43	10:33	00:50	01:00	D	00:10
9	09:33	10:23	00:50	01:00	D	00:10
10	11:59	13:07	01:08	01:20	D	00:12
11	09:23	10:07	00:44	01:00	D	00:16
12	11:05	12:27	01:22	02:00	D	00:38
13	09:35	09:45	00:10	00:05	C	00:05
14	09:59	11:07	01:08	01:00	D	00:08
15	09:45	09:58	00:13	00:10	D	00:03
16	09:55	10:33	00:38	00:30	D	00:08
17	11:23	12:31	01:08	00:40	D	00:28
18	11:49	12:33	00:44	00:40	D	00:04
19	12:17	13:17	01:00	01:00	D	00:00
20	12:37	12:59	00:22	00:20	D	00:02

E Carson Street-09/25/2015, Friday

Car #	Arrive	Leave	Actual	Paid	Trip purpose	Overpaid time
1	09:49	09:58	00:09	00:10	D	00:01
2	11:59	12:56	00:57	01:00	D	00:03
3	11:51	13:08	01:17	01:20	D	00:03
4	13:51	14:47	00:56	01:00	C	00:04
5	11:48	12:33	00:45	00:50	D	00:05
6	13:59	14:52	00:53	01:00	D	00:07
7	09:21	10:23	01:02	01:10	D	00:08
8	09:07	10:59	01:52	02:00	A	00:08
9	12:07	13:17	01:10	01:20	D	00:10
10	10:01	11:11	01:10	01:20	D	00:10
11	10:07	10:27	00:20	00:30	D	00:10
12	11:38	12:27	00:49	01:00	C	00:11
13	09:10	09:58	00:48	01:00	D	00:12
14	12:55	14:38	01:43	02:00	D	00:17
15	11:09	12:48	01:39	02:00	C	00:21
16	12:47	13:23	00:36	01:00	C	00:24
17	09:11	09:23	00:12	00:10	D	00:02
18	12:17	13:23	01:06	01:00	D	00:06
19	12:33	13:23	00:50	00:40	D	00:10
20	10:10	11:07	00:57	00:30	D	00:27
21	09:11	09:23	00:12	00:10	D	00:02
22	11:11	11:49	00:38	00:30	C	00:08
23	09:26	11:30	02:04	02:00	A	00:04
24	09:29	10:01	00:32	00:30	D	00:02

E Carson Street-09/19/2015, Saturday

Car #	Arrive	Leave	Actual	Paid	Trip purpose	Overpaid time
1	11:13	12:33	01:20	01:20	D	00:00
2	12:09	12:49	00:40	00:40	D	00:00
3	11:09	12:07	00:58	01:00	D	00:02
4	10:30	11:08	00:38	00:40	D	00:02
5	11:49	13:46	01:57	02:00	D	00:03
6	10:20	12:17	01:57	02:00	D	00:03
7	13:47	15:23	01:36	01:40	C	00:04
8	12:45	12:51	00:06	00:10	D	00:04
9	10:41	10:46	00:05	00:10	D	00:05
10	09:52	09:56	00:04	00:10	D	00:06
11	09:57	10:30	00:33	00:40	D	00:07
12	12:43	14:35	01:52	02:00	D	00:08
13	10:05	11:17	01:12	01:20	D	00:08
14	11:07	12:59	01:52	02:00	D	00:08
15	13:07	14:59	01:52	02:00	D	00:08
16	14:17	15:07	00:50	01:00	C	00:10
17	13:13	13:33	00:20	00:30	D	00:10
18	10:48	12:37	01:49	02:00	D	00:11
19	10:05	10:14	00:09	00:20	D	00:11
20	10:48	11:07	00:19	00:40	D	00:21
21	10:22	11:59	01:37	02:00	D	00:23
22	10:14	11:07	00:53	02:00	D	01:07
23	10:07	10:55	00:48	02:00	D	01:12
24	14:07	14:37	00:30	00:20	C	00:10
25	12:13	13:19	01:06	01:00	D	00:06
26	13:21	14:13	00:52	00:40	C	00:12
27	10:30	10:41	00:11	00:10	D	00:01
28	10:50	12:03	01:13	01:10	D	00:03

University area

Thackeray Avenue-09/09/15, Wednesday

Car #	Arrive	Leave	Actual	Paid	Trip purpose	Overpaid time
1	09:35	12:03	02:28	02:30	A	00:02
2	12:05	12:52	00:47	00:50	D	00:03
3	10:07	14:00	03:53	04:00	D	00:07
4	10:49	13:50	03:01	03:10	B	00:09
5	14:00	14:47	00:47	01:00	B	00:13
6	08:56	09:40	00:44	01:00	A	00:16
7	10:20	11:03	00:43	01:00	D	00:17
8	08:30	14:13	05:43	06:00	A	00:17
9	12:13	14:24	02:11	02:30	A	00:19
10	14:33	14:54	00:21	00:40	D	00:19
11	14:08	14:40	00:32	01:00	B	00:28
12	14:05	14:27	00:22	01:00	B	00:38
13	09:30	11:17	01:47	02:30	D	00:43
14	09:07	09:22	00:15	04:00	A	03:45
15	14:26	14:56	00:30	00:20	B	00:10
16	09:08	09:40	00:32	00:20	B	00:12
17	11:05	13:10	02:05	02:00	D	00:05

Thackeray Avenue-09/17/15, Thursday

Car #	Arrive	Leave	Actual	Paid	Trip purpose	Overpaid time
1	14:33	15:33	01:00	01:00	B	00:00
2	11:11	13:38	02:27	02:30	D	00:03
3	12:13	13:08	00:55	01:00	D	00:05
4	13:18	14:13	00:55	01:00	D	00:05
5	12:08	13:01	00:53	01:00	D	00:07
6	13:08	16:58	03:50	04:00	B	00:10
7	09:51	11:38	01:47	02:00	A	00:13
8	10:11	11:58	01:47	02:00	B	00:13
9	09:49	10:33	00:44	01:00	D	00:16
10	11:38	15:12	03:34	04:00	B	00:26
11	09:38	12:11	02:33	03:00	A	00:27
12	14:28	14:59	00:31	01:00	D	00:29
13	10:48	13:18	02:30	03:00	A	00:30
14	09:08	14:07	04:59	06:00	A	01:01
15	09:11	10:33	01:22	01:15	D	00:07
16	10:18	11:08	00:50	00:30	D	00:20
17	09:42	11:43	02:01	02:00	B	00:01
18	09:44	10:47	01:03	01:00	D	00:03
19	12:28	13:33	01:05	01:00	D	00:05
20	12:51	14:08	01:17	01:00	D	00:17
21	14:38	15:18	00:40	00:30	D	00:10

Tech Street-09/10/2015, Thursday

Car #	Arrive	Leave	Actual	Paid	Trip purpose	Overpaid time
1	09:55	11:09	01:14	01:15	B	00:01
2	11:10	12:05	00:55	01:00	D	00:05
3	09:50	10:35	00:45	00:50	A	00:05
4	10:36	11:25	00:49	00:56	D	00:07
5	13:17	14:38	01:21	01:30	D	00:09
6	09:39	11:41	02:02	02:15	D	00:13
7	13:01	14:18	01:17	01:30	B	00:13
8	13:35	14:40	01:05	01:21	D	00:16
9	09:50	11:20	01:30	01:50	B	00:20
10	01:01	02:35	01:34	02:00	B	00:26
11	09:44	13:15	03:31	04:00	B	00:29
12	10:23	13:10	02:47	03:16	B	00:29
13	09:47	11:15	01:28	02:00	B	00:32
14	09:43	11:07	01:24	02:00	D	00:36
15	11:30	12:50	01:20	02:00	D	00:40
16	09:35	12:00	02:25	03:07	A	00:42
17	10:02	10:40	00:38	01:28	D	00:50
18	09:22	12:00	02:38	03:30	B	00:52
19	01:17	02:40	01:23	02:20	D	00:57
20	09:16	09:47	00:31	01:30	D	00:59
21	09:43	12:08	02:25	05:00	B	02:35
22	paid all day(10:44-18:00)				B	
23	paid all day(13:50-18:00)				D	
24	paid all day(12:03-18:00)				B	
25	paid all day(11:25-18:00)				B	
26	paid all day(11:32-18:00)				A	
27	09:05			10:27	01:22	01:00
28	12:08	13:40	01:32	01:30	D	00:22
29	12:47	14:48	02:01	01:40	B	00:02
30	12:17	13:45	01:28	01:17	D	00:21
31	10:11	12:07	01:56	00:50	D	00:11
32	12:22	12:53	00:31	00:28	D	01:06
33	09:55	11:25	01:30	01:21	D	00:03
34	09:13	12:33	03:20	03:00	D	00:09
35	09:54	11:06	01:12	01:00	B	00:20
36	09:55	12:05	02:10	02:00	B	00:12
37	11:10	13:28	02:18	02:00	B	00:10

Tech Street-09/16/2015, Wednesday

Car #	Arrive	Leave	Actual	Paid	Trip purpose	Overpaid time
1	09:20	11:20	02:00	02:00	D	00:00
2	09:53	11:11	01:18	01:28	D	00:10
3	09:12	09:51	00:39	00:49	D	00:10
4	09:45	10:32	00:47	01:00	D	00:13
5	12:51	14:38	01:47	02:00	B	00:13
6	12:48	14:01	01:13	01:28	D	00:15
7	09:32	11:10	01:38	02:00	D	00:22
8	10:28	11:01	00:33	01:00	D	00:27
9	13:08	14:33	01:25	02:00	B	00:35
10	09:51	13:11	03:20	04:00	A	00:40
11	11:08	pay all day			B	
12	13:48	pay allday			B	
13	10:11	10:43	00:32	00:28	D	00:04
14	10:39	13:48	03:09	03:00	A	00:09
15	09:37	10:32	00:55	00:49	D	00:06
16	11:05	13:21	02:16	01:35	D	00:41
17	10:01	12:08	02:07	02:00	D	00:07
18	12:55	13:38	00:43	00:35	D	00:08
19	12:08	13:18	01:10	01:00	B	00:10

APPENDIX B

ANALYSIS RESULTS

The following graphs are the analysis results from software R.

```
Call:
lm(formula = PTR ~ factor(TP) + factor(D) + factor(PLT) + PD +
    factor(PLH) + factor(PR))

Residuals:
     Min       1Q   Median       3Q      Max
-0.76049 -0.06868  0.05760  0.13275  0.22573

Coefficients: (2 not defined because of singularities)
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.773646   0.091267   8.477 6.61e-14 ***
factor(TP)2     0.023420   0.070445   0.332    0.740
factor(TP)3     0.046809   0.104698   0.447    0.656
factor(TP)4    -0.023960   0.071459  -0.335    0.738
factor(D)1      0.016208   0.046943   0.345    0.730
factor(PLT)1   -0.006374   0.059321  -0.107    0.915
PD              0.021544   0.019277   1.118    0.266
factor(PLH)4   -0.029960   0.047748  -0.627    0.532
factor(PLH)10  -0.038295   0.071438  -0.536    0.593
factor(PR)2.25        NA         NA      NA       NA
factor(PR)3           NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1904 on 121 degrees of freedom
Multiple R-squared:  0.04069,	Adjusted R-squared:  -0.02274
F-statistic: 0.6415 on 8 and 121 DF,  p-value: 0.7415
```

Figure A. Linear regression result between parking time ratio and all the variables

Figure A shows the original data linear regression results analysis using all the variables includes trip purpose(TP), day of the week(D), parking location type(PLT), pay duration(PD), parking limit hour(PLT) and parking rate(PR) to predict parking time ratio(PTR).

```
Call:
lm(formula = PTR ~ factor(TP) + factor(D) + factor(PLT) + PD +
    factor(PLH) + factor(PR))

Residuals:
     Min       1Q   Median       3Q      Max
-0.55456 -0.04099  0.01637  0.06841  0.17929

Coefficients: (2 not defined because of singularities)
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.73478    0.07581   9.692 1.44e-14 ***
factor(TP)2     -0.03625    0.07295  -0.497    0.621
factor(TP)3      0.06117    0.08697   0.703    0.484
factor(TP)4     -0.01505    0.06721  -0.224    0.823
factor(D)1      -0.02430    0.03393  -0.716    0.476
factor(PLT)1    -0.03881    0.05397  -0.719    0.474
PD               0.10756    0.02146   5.011 3.90e-06 ***
factor(PLH)4     0.02408    0.03468   0.694    0.490
factor(PLH)10    0.05599    0.06373   0.879    0.383
factor(PR)2.25        NA         NA      NA       NA
factor(PR)3           NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1219 on 70 degrees of freedom
Multiple R-squared:  0.3289,    Adjusted R-squared:  0.2522
F-statistic: 4.288 on 8 and 70 DF,  p-value: 0.0003097
```

Figure B. Linear regression result between parking time ratio and all the variables (overpaid time 0-15 min)

Figure B is the results of the model used the data that the overpaid time is between 0 and 15 minutes to do linear regression results analysis using all the variables includes trip purpose(TP), day of the week(D), parking location type(PLT), pay duration(PD), parking limit hour(PLT) and parking rate(PR) to predict parking time ratio(PTR).

```
Call:
lm(formula = PTR ~ PD)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.54462 -0.03735  0.01371  0.08079  0.18594 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.73490    0.02470  29.753  < 2e-16 ***
PD           0.09723    0.01705   5.703 2.07e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.119 on 77 degrees of freedom
Multiple R-squared:  0.297,	Adjusted R-squared:  0.2879 
F-statistic: 32.53 on 1 and 77 DF,  p-value: 2.074e-07
```

Figure C. Linear regression results between parking time ratio and pay duration

Figure C is the results of the model used the variable pay duration(PD) only to predict parking time ratio(PTR).

```
Call:
lm(formula = PTR ~ PD, weights = 1/exp(fitted(model1)))

Weighted Residuals:
    Min      1Q  Median      3Q     Max 
-6.0213 -0.8247  0.2950  1.2607  2.7533 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.825619   0.014765  55.916  < 2e-16 ***
PD          0.042493   0.006938   6.124 4.01e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.829 on 74 degrees of freedom
Multiple R-squared:  0.3364,	Adjusted R-squared:  0.3274 
F-statistic: 37.51 on 1 and 74 DF,  p-value: 4.008e-08
```

Figure D. Linear regression results after adjustment

Figure D is the results of the model after avoiding the influence of abnormal data and Heteroscedasticity

```
Call:
lm(formula = PTR ~ PD, weights = 1/exp(fitted(model1)))

Weighted Residuals:
    Min      1Q  Median      3Q     Max 
-11.527  -1.072  -0.073   1.202   9.198 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.9737     0.0336    29.0   <2e-16 ***
PD           -0.0369     0.0119    -3.1   0.0022 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.32 on 189 degrees of freedom
Multiple R-squared:  0.0485,	Adjusted R-squared:  0.0434 
F-statistic: 9.63 on 1 and 189 DF,  p-value: 0.00221
```

Figure E. Linear regression result after adding under-paid parking data.

Figure E is the linear regression model results after adding under-paid parking data.

```
> model4<-cv.lm(data = parkings,form.lm = formula(PTR~PD,weights = 1/exp(fitted(model1))),m=5)
Analysis of Variance Table

Response: PTR
           Df Sum Sq Mean Sq F value Pr(>F)
PD          1   0.94   0.941    9.89 0.0019 **
Residuals 189  17.98   0.095
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


fold 1
Observations in test set: 38
                 1      5      8     19     23     34      47    52     54     57      59     61     64     67
PD           4.000  1.000  2.300  0.700  1.000  2.000  3.1000  0.30  1.000  1.000  1.8000  0.700  1.000  1.000
cvpred       0.817  0.949  0.892  0.962  0.949  0.905  0.8565  0.98  0.949  0.949  0.9138  0.962  0.949  0.949
PTR          0.063  0.367  0.413  0.525  0.583  0.708  0.7750  0.80  0.800  0.817  0.8180  0.825  0.833  0.833
CV residual -0.754 -0.582 -0.479 -0.437 -0.366 -0.197 -0.0815 -0.18 -0.149 -0.132 -0.0958 -0.137 -0.116 -0.116
                68      71     73     74      80     86      91      94      99     100     111      116
PD          4.0000  1.3000  1.500  2.000  1.0000  2.000  0.8000  1.3000 2.00000  1.0000 1.30000 1.000000
cvpred      0.8168  0.9358  0.927  0.905  0.9491  0.905  0.9579  0.9358 0.90497  0.9491 0.93584 0.949066
PTR         0.8330  0.8500  0.856  0.858  0.8830  0.892  0.9000  0.9000 0.90800  0.9170 0.93800 0.950000
CV residual 0.0162 -0.0858 -0.071 -0.047 -0.0661 -0.013 -0.0579 -0.0358 0.00303 -0.0321 0.00216 0.000934
               124   129    130   141   147   149   150   161   186   187  188   191
PD          1.0000 2.500 1.3000 0.800 1.000 0.200 0.500 0.200 0.500 0.700 0.30 0.800
cvpred      0.9491 0.883 0.9358 0.958 0.949 0.984 0.971 0.984 0.971 0.962 0.98 0.958
PTR         0.9670 0.987 0.9870 1.080 1.100 1.100 1.107 1.200 1.667 1.700 1.80 2.320
CV residual 0.0179 0.104 0.0512 0.122 0.151 0.116 0.136 0.216 0.696 0.738 0.82 1.362

Sum of squares = 5.35    Mean square = 0.14    n = 38

fold 2
Observations in test set: 39
                 2      3     12    16     17    22     24     27     36     39     40     42     49     50
PD           0.100  0.700  0.700  1.00  0.800  1.00  2.300  0.500  1.000  1.000  1.000  1.700  1.000  2.000
cvpred       1.065  1.015  1.015  0.99  1.006  0.99  0.880  1.032  0.990  0.990  0.990  0.931  0.990  0.905
PTR          0.200  0.250  0.475  0.50  0.511  0.55  0.593  0.667  0.717  0.733  0.733  0.740  0.783  0.783
CV residual -0.865 -0.765 -0.540 -0.49 -0.495 -0.44 -0.287 -0.365 -0.273 -0.257 -0.257 -0.191 -0.207 -0.122
                51      58     62     65     70      83      87      95      96     103    108     112     114
PD           0.800  2.0000  1.500 3.0000  0.300  1.2000  2.0000  1.0000  1.5000  0.7000 2.0000  0.8000  0.3000
cvpred       1.006  0.9054  0.947 0.8211  1.049  0.9727  0.9054  0.9896  0.9475  1.0149 0.9054  1.0064  1.0485
PTR          0.796  0.8170  0.830 0.8330  0.850  0.8860  0.8920  0.9000  0.9000  0.9250 0.9330  0.9400  0.9500
CV residual -0.210 -0.0884 -0.117 0.0119 -0.199 -0.0867 -0.0134 -0.0896 -0.0475 -0.0899 0.0276 -0.0664 -0.0985
                123   125    126   131    142   143   151   154   155   159   178   183
PD           1.0000 4.000 2.0000 2.000 1.0000 2.000 1.300 1.000 1.300 1.000 1.000 0.300
cvpred       0.9896 0.737 0.9054 0.905 0.9896 0.905 0.964 0.990 0.964 0.990 0.990 1.049
PTR          0.9670 0.971 0.9750 1.008 1.0830 1.083 1.111 1.133 1.143 1.167 1.367 1.500
CV residual -0.0226 0.234 0.0696 0.103 0.0934 0.178 0.147 0.143 0.179 0.177 0.377 0.451

Sum of squares = 3.53    Mean square = 0.09    n = 39

fold 3
Observations in test set: 38
                 4      9     15     29     30     33    43      44     45     46     48      56     66     69
PD           1.500  1.500  0.200  2.000  2.000  2.600  1.30  3.5000  0.500  1.000  1.000  2.0000  1.000 3.0000
cvpred       0.916  0.916  1.004  0.882  0.882  0.842  0.93  0.7807  0.984  0.950  0.950  0.8823  0.950 0.8145
PTR          0.344  0.432  0.500  0.667  0.683  0.705  0.75  0.7520  0.767  0.767  0.783  0.8080  0.833 0.8500
CV residual -0.572 -0.484 -0.504 -0.215 -0.199 -0.137 -0.18 -0.0287 -0.217 -0.183 -0.167 -0.0743 -0.117 0.0355
                 78      85      92     97    105   118   120    122    127   128   135   136  137  145   158
PD           1.3000 2.00000  0.8000  0.200 2.0000 6.000 4.000 1.3000 2.0000 2.500 2.000 1.200 1.00 0.70 2.000
cvpred       0.9297 0.88228  0.9636  1.004 0.8823 0.611 0.747 0.9297 0.8823 0.848 0.882 0.936 0.95 0.97 0.882
PTR          0.8750 0.89200  0.9000  0.900 0.9330 0.953 0.958 0.9630 0.9750 0.980 1.042 1.043 1.05 1.10 1.150
CV residual -0.0547 0.00972 -0.0636 -0.104 0.0507 0.342 0.211 0.0333 0.0927 0.132 0.160 0.107 0.10 0.13 0.268
              162  163   165   169   174   175   179   182   189
PD          0.200 1.00 1.500 0.500 0.200 0.800 0.100 0.300 0.500
cvpred      1.004 0.95 0.916 0.984 1.004 0.964 1.011 0.997 0.984
PTR         1.200 1.20 1.222 1.267 1.300 1.320 1.400 1.500 1.900
CV residual 0.196 0.25 0.306 0.283 0.296 0.356 0.389 0.503 0.916

Sum of squares = 3.16    Mean square = 0.08    n = 38
```

```
fold 4
Observations in test set: 38
                6     10     11     14     20     28     31    32     35     37     41    63     72    79
PD          0.200  2.000  0.300  5.000  1.000  0.500  0.200  2.00  2.500  1.000  2.000 6.000 3.3000 4.000
cvpred      0.995  0.890  0.989  0.715  0.948  0.977  0.995  0.89  0.860  0.948  0.890 0.656 0.8138 0.773
PTR         0.400  0.442  0.450  0.483  0.533  0.667  0.700  0.70  0.713  0.733  0.733 0.831 0.8520 0.879
CV residual -0.595 -0.448 -0.539 -0.232 -0.415 -0.310 -0.295 -0.19 -0.147 -0.215 -0.157 0.175 0.0382 0.106
               82      89      90      93     98    106    107    109    110    115   133   134    140   153
PD          1.000  1.5000  1.3000  0.2000 2.3000 2.0000  1.000 2.0000  1.000 1.3000 2.000 2.000 0.5000 0.800
cvpred      0.948  0.9188  0.9305  0.9947 0.8721 0.8896  0.948 0.8896  0.948 0.9305 0.890 0.890 0.9772 0.960
PTR         0.883  0.9000  0.9000  0.9000 0.9040 0.9330  0.933 0.9330  0.933 0.9500 1.025 1.033 1.0670 1.122
CV residual -0.065 -0.0188 -0.0305 -0.0947 0.0319 0.0434 -0.015 0.0434 -0.015 0.0195 0.135 0.143 0.0898 0.162
              157   168   170   171   172   173   180   181   184   190
PD          0.300 0.500 0.700 1.000 0.700 0.200 1.500 0.300 0.100 0.100
cvpred      0.989 0.977 0.966 0.948 0.966 0.995 0.919 0.989 1.001 1.001
PTR         1.150 1.267 1.275 1.283 1.300 1.300 1.432 1.500 1.600 2.000
CV residual 0.161 0.290 0.309 0.335 0.334 0.305 0.513 0.511 0.599 0.999

Sum of squares = 3.92    Mean square = 0.1    n = 38

fold 5
Observations in test set: 38
               7     13     18     21     25     26     38     53     55      60     75      76      77
PD          2.00  0.700  1.000  0.300  0.200  1.000  0.800  0.300  1.400  2.0000 2.5000  0.9000  1.3000
cvpred      0.88  0.985  0.961  1.018  1.026  0.961  0.977  1.018  0.929  0.8805 0.8402  0.9692  0.9369
PTR         0.40  0.475  0.517  0.550  0.600  0.600  0.733  0.800  0.802  0.8250 0.8730  0.8750  0.8750
CV residual -0.48 -0.510 -0.444 -0.468 -0.426 -0.361 -0.244 -0.218 -0.127 -0.0555 0.0328 -0.0942 -0.0619
                81      84    88     101     102    104     113   117    119    121   132  138   139   144
PD          1.0000  1.5000 4.000  1.0000  1.0000 2.0000  0.7000 3.200 2.0000 1.7000 1.500 3.00 2.000 1.300
cvpred      0.9611  0.9208 0.719  0.9611  0.9611 0.8805  0.9853 0.784 0.8805 0.9047 0.921 0.80 0.880 0.937
PTR         0.8830  0.8860 0.892  0.9170  0.9170 0.9330  0.9500 0.953 0.9580 0.9600 1.022 1.05 1.058 1.093
CV residual -0.0781 -0.0348 0.173 -0.0441 -0.0441 0.0525 -0.0353 0.169 0.0775 0.0553 0.101 0.25 0.178 0.156
               146   148   152   156   160   164   166   167   176   177   185
PD          0.3000 1.000 3.000 0.500 3.000 1.700 0.500 0.700 0.500 1.000 0.300
cvpred      1.0176 0.961 0.800 1.001 0.800 0.905 1.001 0.985 1.001 0.961 1.018
PTR         1.1000 1.100 1.111 1.143 1.192 1.210 1.229 1.250 1.333 1.350 1.600
CV residual 0.0824 0.139 0.311 0.142 0.392 0.305 0.228 0.265 0.332 0.389 0.582

Sum of squares = 2.68    Mean square = 0.07    n = 38

Overall (Sum over all 38 folds)
    ms
0.0976
```

Figure F. 5-fold cross validation results

Figure F shows the 5-fold cross validation process of the original model.

```
Call:
lm(formula = PTR ~ PD + 0)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.6060  0.0947  0.3768  0.6775  1.9951 

Coefficients:
   Estimate Std. Error t value Pr(>|t|)    
PD   0.4062     0.0289      14   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.692 on 190 degrees of freedom
Multiple R-squared:  0.509,	Adjusted R-squared:  0.507 
F-statistic:  197 on 1 and 190 DF,  p-value: <2e-16
```

Figure G. Linear regression results after adjustment

Figure G shows the results of the model without intercept.

```
Call:
lm(formula = PTR ~ factor(TP) + factor(PLT) + PD + factor(PLH) +
    factor(PR) + factor(D) + 0)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8160 -0.1755  0.0021  0.1521  1.3181

Coefficients: (2 not defined because of singularities)
               Estimate Std. Error t value Pr(>|t|)
factor(TP)1      1.0085     0.1215    8.30  2.3e-14 ***
factor(TP)2      1.0072     0.0998   10.09  < 2e-16 ***
factor(TP)3      1.1595     0.0973   11.91  < 2e-16 ***
factor(TP)4      0.9636     0.0634   15.20  < 2e-16 ***
factor(PLT)1     0.0281     0.0756    0.37   0.7099
PD              -0.0820     0.0269   -3.05   0.0026 **
factor(PLH)4     0.0606     0.0627    0.97   0.3351
factor(PLH)10    0.0758     0.0936    0.81   0.4189
factor(PR)2.25       NA         NA      NA       NA
factor(PR)3          NA         NA      NA       NA
factor(D)1      -0.0013     0.0621   -0.02   0.9833
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.309 on 182 degrees of freedom
Multiple R-squared:  0.906,     Adjusted R-squared:  0.902
F-statistic:  196 on 9 and 182 DF,  p-value: <2e-16
```

Figure H. The results of the model with all variables

Figure H shows the results of using all the variables in the model without intercept.

```
Start:  AIC=-440
PTR ~ factor(TP) + factor(PLT) + PD + factor(PLH) + factor(PR) +
    factor(D) + 0


Step:  AIC=-440
PTR ~ factor(TP) + factor(PLT) + PD + factor(PLH) + factor(D) -
    1

              Df Sum of Sq  RSS  AIC
- factor(PLH)  2      0.09 17.5 -443
- factor(D)    1      0.00 17.4 -442
- factor(PLT)  1      0.01 17.4 -442
<none>                     17.4 -440
- PD           1      0.89 18.3 -432
- factor(TP)   4     27.28 44.6 -268

Step:  AIC=-443
PTR ~ factor(TP) + factor(PLT) + PD + factor(D) - 1

              Df Sum of Sq  RSS  AIC
- factor(D)    1       0.0 17.5 -445
- factor(PLT)  1       0.1 17.6 -444
<none>                     17.5 -443
- PD           1       0.9 18.4 -435
- factor(TP)   4      48.6 66.0 -197

Step:  AIC=-445
PTR ~ factor(TP) + factor(PLT) + PD - 1

              Df Sum of Sq  RSS  AIC
- factor(PLT)  1       0.1 17.6 -445
<none>                     17.5 -445
- PD           1       0.9 18.4 -437
- factor(TP)   4      68.4 85.9 -149

Step:  AIC=-445
PTR ~ factor(TP) + PD - 1

              Df Sum of Sq  RSS  AIC
<none>                     17.6 -445
- PD           1       0.8 18.4 -439
- factor(TP)   4      73.4 91.0 -140
```

Figure I. The process of preventing unrelated variables

Figure I shows a process of preventing unrelated variables to get the final model.

```
Call:
lm(formula = PTR ~ factor(TP) + PD - 1)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.8056 -0.1616 -0.0095  0.1565  1.3683 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
factor(TP)1   1.0735     0.1082    9.92   <2e-16 ***
factor(TP)2   1.0908     0.0776   14.05   <2e-16 ***
factor(TP)3   1.1700     0.0913   12.81   <2e-16 ***
factor(TP)4   1.0133     0.0389   26.01   <2e-16 ***
PD           -0.0770     0.0260   -2.97   0.0034 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.308 on 186 degrees of freedom
Multiple R-squared:  0.905,	Adjusted R-squared:  0.903 
F-statistic:  355 on 5 and 186 DF,  p-value: <2e-16
```

Figure J. Results of final model

Figure J shows the results of the final model.

BIBLIOGRAPHY

[1] Chen Z, Yin Y, He F, et al. Parking Reservation for Managing Downtown Curbside Parking[C]//Transportation Research Board 94th Annual Meeting. 2015 (15-5008).

[2] Carvalho e Ferreira D, de Abreu e Silva J. Simulation of a Parking Reservations System to Mitigate Cruising for Parking[C]//Transportation Research Board 92nd Annual Meeting. 2013 (13-3718).

[3]Diallo A, Bourdeau J S, Morency C, et al. Methodology of parking analysis[J]. Canadian Journal of Civil Engineering, 2015, 42(4): 281-285.

[4] Vlahogianni E I, Kepaptsoglou K, Tsetsos V, et al. Exploiting new sensor technologies for real-time parking prediction in urban areas[C]//Transportation Research Board 93rd Annual Meeting Compendium of Papers. 2014: 14-1673.

[5] Maleck B, Sarasua W A, Ogle J H, et al. A Methodology Using GPS to Inventory University Campus Parking[J]. Journal of Transportation of the Institute of Transportation Engineers, 2014, 6(1).

[6] Bulan O, Loce R P, Wu W, et al. Video-based real-time on-street parking occupancy detection system[J]. Journal of Electronic Imaging, 2013, 22(4): 041109-041109.

[7] Lattunen A, Arvonen K, Katasonov A, et al. Real-Time Event-Based Information Collection from Street Parking[C]//20th ITS World Congress. 2013.

[8] Draper N R, Smith H, Pownell E. Applied regression analysis[M]. New York: Wiley, 1966.

[9]Hubbard, R. (2004). Blurring the Distinctions Between p’s and a’s in Psychological Research, Theory Psychology June 2004 vol. 14 no. 3 295-327

[10] Wilk, M.B.; Gnanadesikan, R. (1968), "Probability plotting methods for the analysis of data", *Biometrika* (Biometrika Trust) **55** (1): 1–17,doi:10.1093/biomet/55.1.1.

[11] Long, J. Scott; Trivedi, Pravin K. (1993). "Some Specification Tests for the Linear Regression Model". In Bollen, Kenneth A.; Long, J. Scott.*Testing Structural Equation Models*. London: Sage. pp. 66–110. ISBN 0-8039-4506-X.