PREDICTING THE APPRAISED UNIT VALUE OF UNIMPROVED PARCELS IN SAN FRANCISCO COUNTY, CA USING LEED SUSTAINABLE SITE CREDIT CRITERIA, PARCEL AREA, ZONING, POPULATION DENSITY

A Thesis

by

HYUN JEONG CHO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2011

Major Subject: Construction Management

PREDICTING THE APPRAISED UNIT VALUE OF UNIMPROVED PARCELS IN SAN FRANCISCO COUNTY, CA USING LEED SUSTAINABLE SITE CREDIT CRITERIA, PARCEL AREA, ZONING, POPULATION DENSITY

A Thesis

by

HYUN JEONG CHO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Paul Woods
Committee Members,	Valerian Miranda
	Mike Speed
Head of Department,	Joe Horlen

August 2011

Major Subject: Construction Management

ABSTRACT

Predicting the Appraised Unit Value of Unimproved Parcels in San Francisco, CA Using LEED Sustainable Site Credit Criteria, Parcel area, Zoning, and Population Density.

(August 2011)

Hyun Jeong Cho, B.E., Kyung Hee University, Seoul, Korea

Chair of Advisory Committee: Dr. Paul K. Woods

Nowadays many people recognize the need for sustainable development more than ever because of improper urban sprawl, rapid exhaustion of natural resources, and serious environmental problems.

Emission of carbon dioxide from transportation sources causes severe air contamination, and this will increase due to the continued use of private vehicles. Thus, local governments are trying to keep public transit on a satisfactory level due to rising commuting time for cities. The U.S. Census shows that the majority of people would prefer to use their private automobile rather than utilize existing mass transit systems. Therefore, it is up to local governments to consider setting up more efficient alternative mass transit systems to deal with the increasing pollution caused by automobiles.

Organizations adopt certain environmental standards for many different reasons, such as commitment to environmental issues affecting their industry. Other organizations could also benefit, both economically and environmentally, by utilizing such standards. The LEED (Leadership in Energy and Environmental Design) rating system is one of the more commonly-used environmental standards which presents guidelines for making decisions regarding land development while preserving the environment. However, only a few studies have attempted to evaluate this voluntary rating system which makes it difficult to justify the motivation of organizations that have adopted such voluntary standards. In this respect, this research primarily aims to explore the economic influence on the market value of undeveloped land through an analysis of public transportation in San Francisco, CA. Population density and area of each parcel are the factors considered to make the predictive model more powerful.

Findings in this study show that LEED PTA (Public Transportation Accessibility) criteria, and population density significantly affect the appraised land unit value in specific purposed zones. Particularly, the economic impact of public transportation accessibility tended to be positive. With these findings, the statistical model for predicting land value was created. The result of this research can assist developers to make better site selections to accelerate the growth of sustainable construction.

DEDICATION

To my parents

ACKNOWLEDGEMENTS

I am grateful to my committee chair, Dr. Paul K. Woods for his guidance and enormous help, and to my committee members, Dr. Speed and Dr. Miranda.

Thanks to my friends, colleagues, and the department faculty and staff for making my time at Texas A&M University an unforgettable experience. Especially, I appreciate Kiyoung Son for his help in generating part of data for this thesis, and Jinsu Kim for his statistical advice.

NOMENCLATURE

BART	Bay Area Rapid Transit
CA	California
GCS	Geographic Coordinate System
GIS	Geographic Information System
LEED	Leadership in Energy and Environmental Design
LEED-NC	Leadership in Energy and Environmental Design for New Construction
PTA	Public Transportation Accessibility
USGBC	United States Green Building Council
VIF	Variance Inflation Factors

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

Page

1. INTRODUCTION

1.1 Research Objective

The main purpose of this study is to discover the likely relationship between public transportation accessibility and the market value of real estate by predicting the appraised unit value of unimproved parcels in San Francisco, California using population density and lot size.

1.2 Problem Statement

Various voluntary standards for buildings incorporate a range of environmental costs and benefits, but it is necessary to see if these standards are essential and successful. “Green buildings” have gained much popularity, but the price effect of green building ratings has scarcely been gauged (Greg Kats, 2003). Hence, this study was conducted to help develop a process that could be used to identify factors for estimating the value of real estate in other countries or cities. The population of interest of this research is parcels within San Francisco County.

This thesis follows the style of *International Journal of Construction Education and Research*.

1.3 Significance of Research

1.3.1 Sustainable Development Regarding Public Transportation

The construction industry has undergone a steady decline in productivity and efficiency from the 1960's to the present time (Huang, 2009). Therefore, as interest in sustainable construction has increased, the need for an assessment system to measure the sustainability performance of buildings has also escalated (Redeclift, 1993). The term "sustainable development" comes from the most widely and currently accepted definition by the United Nations, Brundtland Commission in 1987. The report of WCED (World Commission on Environment and Development), "*Our Common Future,*" defined the concept of sustainability development as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (Brundtland, 1987).

The U.S Census shows that the majority of people would prefer to use their private automobiles rather than existing mass transit systems; (Bureau, 2000). Therefore, it is up to local governments to consider setting up more efficient alternative mass transit systems to deal with the increasing pollution caused by automobile use (Puentes, 2004). The rising commuting time in cities puts constant pressure on local governments to keep infrastructure such as highways, roads, and public transit at a satisfactory level (Puentes, 2008).

1.4 Zoning Policy

In this study, two zoning code categories will be investigated- mixed and residential. Mixed use zone consists of a building, building complexes, or neighborhoods for more than one purpose. Residential would mean purely land use for housing. Tables 1-1 and 1-2 explain how the two zone categories are determined in this research.

Table 1-1. Zoning Policy – mixed zoning

Mixed Use Zone			
RM-1	Residential - Mixed Districts, Low Density (Apartments and Houses), Low Density	One dwelling unit per 800 sq. ft. of lot area	Width: 25ft, Area: 2,500sq
RM-2	Residential - Mixed Districts, Moderate Density (Apartments and Houses), Moderate Density	One dwelling unit per 600 sq. ft. of lot area	Width: 25ft, Area: 2,500sq
RM-3	Residential - Mixed Districts, Medium Density (Apartments and Houses), Medium Density	One dwelling unit per 400 sq. ft. of lot area	Width: 25ft, Area: 2,500sq
RM-4	Residential - Mixed Districts, High Density (Apartments and Houses), High Density	One dwelling unit per 200 sq. ft. of lot area	Width: 25ft, Area: 2,500sq
RC-3	Residential - Commercial Combined Districts, Medium Density	One dwelling unit per 400 sq. ft. of lot area	Width: 25ft, Area: 2,500sq
RC-4	Residential - Commercial Combined Districts, High Density	One dwelling unit per 200 sq. ft. of lot area	Width: 25ft, Area: 2,500sq

Table 1-2. Zoning Policy – residential zoning

Residential Zone			
RH-1	Residential - House Districts, One Family	One dwelling unit per lot	Width: 25ft, Area: 2,500sq
RH-1 (D)	Residential - House Districts, One Family-Detached	One dwelling unit per lot	Width: 33ft, Area: 4,000sq
RH-1(S)	Residential - House Districts, One Family-Secondary Unit	One dwelling unit per lot	Width: 25ft, Area: 2,500sq
RH-2	Residential - House Districts, Two Family	Two dwelling unit per lot	Width: 25ft, Area: 2,500sq
RH-3	Residential - House Districts, Three Family	Three dwelling unit per lot	Width: 25ft, Area: 2,500sq

1.5 Hypothesis

To find the economic impact of LEED PTA credits on unimproved land value, the following research hypotheses will be tested.

1. For mixed and residential zoning, the appraised unit value of a PTA qualified parcel is higher when it meets the requirements for the LEED Public Transportation Access credit due to qualifying bus stops.
2. For mixed and residential zoning, the appraised unit value of a PTA qualified parcel is higher when it meets the requirements for the LEED Public Transportation Access credit due to qualifying light rail stations.
3. For mixed and residential zoning, the appraised unit value of a PTA qualified parcel is higher when it meets the requirements for the LEED Public Transportation Access credit due to qualifying commuter rail stations.

4. For mixed and residential zoning, the appraised unit value of a parcel decreases as the population density of the census block increases.
5. For mixed and residential zoning, the appraised unit value of a parcel decreases as the area of the census block increases.

1.5.1 General Model for Mixed and Residential Zones

Unit value of parcel (UV) = $\beta 0$ + $\beta 1$ *B+ $\beta 2$ *L+ $\beta 3$ * C+$\beta 4$ *A+ $\beta 5$ *P+ $\beta 6$*BL + $\beta 7$ *BC+ $\beta 8$ *LC+ $\beta 9$ * BLC

B: 0: does not meet LEED criteria for bus stops 1: meets criteria

L: 0: does not meet LEED criteria for light rail stations, 1: meets criteria

C: 0: does not meet criteria for commuter rail stations, 1: meets criteria

A: lot size in acres

P: Population density of census block (persons per square mile /10,000)

BL: interaction term between B and L

BC: interaction term between B and C

LC: interaction term between L and C

BLC: interaction term among B, L and C

For the statistical models, the confidence level of ANOVA test is 95%, and α value is 0.05.

2. LITERATURE REVIEW

2.1 Introduction of LEED Rating System

Leadership Energy and Environmental Design (LEED) is the most widely used system in the world (Fowler, 2006). Each category has a particular sustainability goal which is referred to as the intent of the credit. Satisfying the intent of a credit means achieving the points assigned to each credit. During the project review process, prerequisites and credits are updated with status such as Anticipated/ Clarify/ Achieved/ Denied. Credits are not earned during the design phase - points are earned only after the construction phase. Points are given to each credit when the requirements are met by the way performance is achieved at construction completion. A project can achieve up to 110 points and one of four levels which are Certified (40+), Silver (50+), Gold (60+) and Platinum (80+) according to the number of points. the number of buildings certified has grown annually. The number of projects certified has increased rapidly, thus, over 40,000 commercial and industrial buildings are under construction or have constructed in 117 countries from 2000 (Ginger Christ, 2011). This figure shows how popular LEED is in the building industry for measuring sustainability of projects all around the world. The system has prospered because of its numerous advantages which have been revised and upgraded over time (USGBC, 2008).

There are several noticeable benefits from using LEED, yet several studies question the effectiveness of transportation systems and some developers do not consider the importance of the LEED rating system although it is becoming more popular (Weber, 2010). Therefore, this research focuses on Sustainable Sites (SS), especially on SS Credit 4.1. which weighs the importance of public transportation accessibility among the seven topics of LEED for New Construction (LEED-NC). The proportion for category SS Credit 4.1 is 6, which is the largest of the SS parts. (USGBC, 2008)

2.2 Public Transportation in San Francisco County

San Francisco has emerged as the fourth most populous city in California, and 13th in the United States with an estimated population of 805,235 in 2010. (U.S. Census Bureau Delivers California's 2010 Census Population Totals, U.S. Census Bureau. 2011-03-08. Retrieved 2011-03-30). Many different types of public transportation are available in San Francisco, and about 30% of residents in San Francisco commute by public transportation in 2005 (Christie, 2008).

The San Francisco Municipal Railway (Muni) predominantly provides public transit within the city, which operates a combined light rail and subway system (the Muni Metro) as well as a bus network (Travel Resources: Public Transportation, 2011). While the Metro streetcars run on the surface streets and underground in the downtown area, the Muni also runs the highly visible F Market historic streetcar line that runs from Castro Street to Fisherman's Wharf. Furthermore, San Francisco's cable car system has been designated as a national historic landmark (Report on San Francisco’s cable cars, 2007). The Commuter rail is provided by the Bay Area Rapid Transit (BART), a regional rapid transit system connecting the San Francisco peninsula with East Bay through the Transbay Tube that runs under Market Street to the Civic Center where it turns south to the Mission District to northern San Mateo County (Regional Transit, 2011). The Caltrain rail system runs from San Francisco to San Jose which was operated for many years by Southern Pacific (Stations, 2011). The Transbay Terminal serves as a long-range bus service and a hub for regional bus systems such as AC Transit (Alameda & Contra Costa counties), SamTrans (San Mateo County), and Golden Gate Transit (Marin and Sonoma Counties). Amtrak also runs a shuttle bus from San Francisco to Emeryville (Amtrak timetable, 2011).

In this study, the use of Muni buses, Muni trains, trolleybuses, cable cars, BART, and Caltrain in San Francisco county will be utilized because they are most commonly used by commuters.

2.3 Population Density

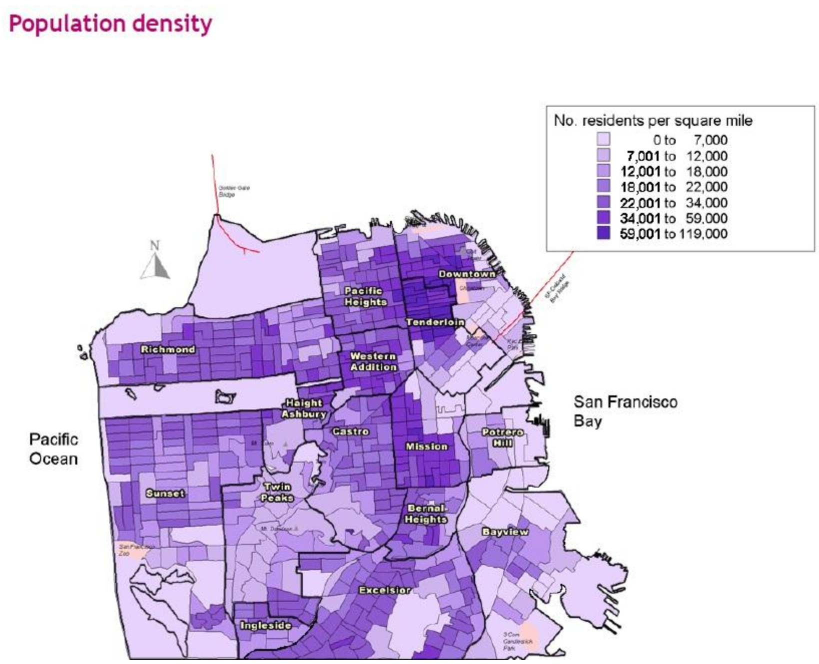


Figure 2-1. Population density (persons/mile^2) in 1999, San Francisco County (McFarland, 2007).

Figure 2-1 describes the distribution of population density in persons per square mile of San Francisco County in 1999. Population density measures how many people or living organisms are in a unit area or unit volume. Population density for people commonly means the number of people per unit area per square kilometer or mile. Usually this can be estimated for a world, county, state, country, city, or a smaller territory. (Rosenberg, 2011)

2.4 Young Jun Park's Study

The primary objectives of Young Jun Park's (2009) research were identifying the relation between LEED criteria and the appraised value of sites, specifically for Houston, Texas. The criteria utilized in the LEED metric for sustainable site selection are Sustainable Site Credit (SSC) #1: Site Selection, SSC #3: Brownfield, and SSC #4.1: Public Transportation Access. The independent variable for this model was LEED sustainable site criteria while the dependent variable was the unit appraised value of land. In order to evaluate the relationship, linear regression was used for quantitative analysis regarding economic profit as well as environmental preservation.

After evaluating the results of statistical analysis, SSC#4.1 was the most significant, including detail components. These results show that the environment was preserved while enhancing the development density near public transportation access (Young Jun Park, 2011).

2.5 Bhagyashri Joshi's Study

Joshi proposed research to identify the economic benefits of "LEED-NC Sustainable Sites" which predicted the appraised values of unimproved parcels in Houston, Texas based on the LEED sustainable rating for Public Transportation Access. Although Joshi established two models in her research, it focuses primarily on Model 1 which utilizes a dependent variable to measure the appraised value of a parcel while the independent variable measures the number of bus stops for a given parcel that meets LEED criteria, the number of light rails for a given parcel that meets LEED criteria, and

Area of a given parcel.

Joshi utilized multiple regressions to evaluate the relationship between the independent and dependent variable and analyze the predicted transformed unit value. This model presented a significant relationship between the transformed unit value of parcels and the measurements required to earn LEED criteria. According to the analysis, an increase in the number of light rail stations led to an increase in the transformed appraised unit value of a parcel; however, the number of bus stops which met LEED criteria for a given parcel had the opposite relationship. These different effects might be a potential link between socio-economic status and transportation mode (B. B. Joshi, 2011).

3. DATA COLLECTION

The target area for this research is limited using only unimproved parcels within San Francisco County shown as the blue colored area of the map in Figure 3-1. The data of each block group was used in this research because it is the smallest cartographical unit that the U.S. Census Bureau offers. Each parcel served as an observational unit for the data which was collected and analyzed; 122 mixed zone parcels and 308 residential zone parcels were selected randomly in San Francisco County, California. Data for this research is quantitative and analyzed by SPSS and SAS, widely used statistical tools.

Appraised unit value of each unimproved parcel is the dependent variable. The independent variable is whether or not a parcel meets the LEED PTA criteria for the number of bus stops, light rail stations, and commuter rail stations. LEED PTA requires a parcel to have bus stops located on at least 2 bus routes within ¼ mile while number of light rail stations and commuter rail stations within ½ mile. In summary, below is the process for collecting data.

3.1 Data Collection Process

1. Basic data of all population is collected from San Francisco County, California;
2. All unimproved parcels are included, but the improved parcels are excluded from the population.
3. The unimproved parcels are arranged by zoning code: mixed and residental zoning;
4. ArcGIS, a GIS tool, is used to create a map to mark public transportation routes and points on the map;

5. Centroid of each parcel is found, and calculated the distance between parcels and public transit points, and determine the transit points are meet the LEED PTA criteria or not.
6. Information of the selected parcels including parcel ID, appraised land values, parcel sizes, population density is gathered;

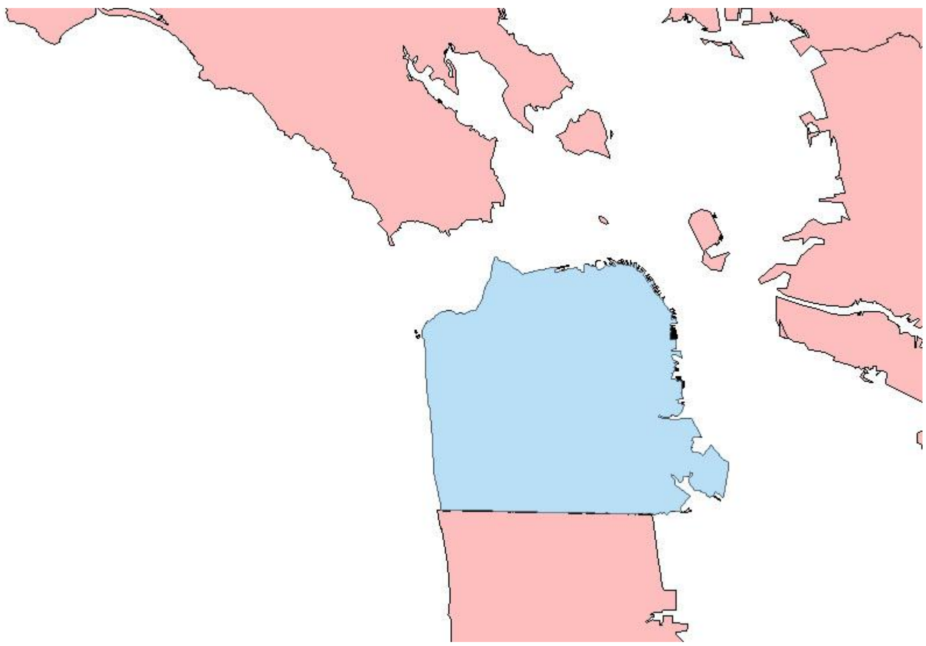


Figure 3-1. Target area of this research

3.2 Population of Interest

Population is limited as all unimproved parcels, which are within the city limits of San Francisco, San Francisco County, California.

The first step is to identify population and samples. The population must include every unimproved parcel in San Francisco, CA, and the samples are to be selected randomly from all the population.

3.3 Sample Selection

The observational unit for this research will be selected from all vacant parcels in San Francisco County. As of September 2010, 2,539 residential purposed parcels were vacant, and 12% of them were chosen as the sample group for this study. For residential and commercial combined purposed parcels, the number of vacant parcels were 273, and 33% of them were chosen for the sample in this study. In short, total 430 zoning area, 308 for residential, 122 for mixed use, were randomly selected among them for this study. In addition, the size of selected parcels is limited according to the zoning code.

In Table 3-1, the number of populations and the number of selected samples of both zones are described.

Table 3-1. Sample selection method

Zoning	Name of District	Number of Vacant Parcel	Total	Minimum Lot Size	Number of Sample Selected
RH-1	Residential - House Districts, One Family	1,050		625	
RH-1 (D)	Residential - House Districts, One Family- Detached	819		1000	
RH-1(S)	Residential - House Districts, One Family- Secondary Unit	11		625	
RH-2	Residential - House Districts, Two Family	504		625	
RH-3	Residential - House Districts, Three Family	155	2,539	625	308
RM-1	Residential - Mixed Districts, Low Density (Apartments and Houses), Low Density	154		625	
RM-2	Residential - Mixed Districts, Moderate Density (Apartments and Houses), Moderate Density	35		625	
RM-3	Residential - Mixed Districts, Medium Density (Apartments and Houses), Medium Density	22		625	
RM-4	Residential - Mixed Districts, High Density (Apartments and Houses), High Density	42		625	
RC-3	Residential - Commercial Combined Districts, Medium Density	2		625	
RC-4	Residential - Commercial Combined Districts, High Density	18	273	625	122

3.4 GIS Data Retrieval

GIS is computer-based information system which uses geo-referenced data to answer questions for managing geographic data and other attributed data, then using the data to solve and manipulate various layers of spatial problems. Geographic data, known as spatial data, is a geographic representation of a data-like map, photos and graphics.

Attribute data is limited to characteristics or descriptions of data such as length, area, population, and address (Feagin, 2010). ArcGIS is one of the credible GIS tools that the ESRI (Environmental Systems Research Institute) recommends, it is used to collect and manage spatial data for this study (Shamsi, 2002).

To collect data, GIS formatted files were retrieved from San Francisco County, and the maps through ArcGIS 9.3 program to create for this study. All shape files for this research were produced with the 'GCS North America 1983' coordinate system to gather qualified and unqualified parcels for PTA components and to produce new shape files to further conduct rigorous statistical analyses. In addition to these spatial data, the attribute data of shape files contain all necessary information for this study including ID, appraised land value, size and addresses. After retrieving the GIS data, tract number and block ID were collected, and population density for each census block was searched referring the block ID.

3.5 Collected Data

Table 3-2. Data description and sources

Data	Contents	Source
Zoning information	bayareablock 00.shp	Metropolitan Transportation Commission (http://www.mtc.ca.gov/maps_and_data/GIS/data.htm)
Land use	planning_lan duse.shp	Cal-Atlas (http://www.atlas.ca.gov/) Geospatial data from California government
Residential information according to parcel identification number.	planning_lan duse.shp	Cal-Atlas (http://www.atlas.ca.gov/)
Parks	Parks_shp	Cal-Atlas (http://www.atlas.ca.gov/)
Hydrology, river, stream, lakes, water bodies, and inland waters	Hydro24ca_t 83.shp	Cal-Atlas (http://www.atlas.ca.gov/)
State highways	ca_major_roa ds.shp	Cal-Atlas (http://www.atlas.ca.gov/)
Latitude, longitude, route ID, and name of bus stops	Bus_Stops.sh p	Metropolitan Transportation Commission (http://www.mtc.ca.gov/maps_and_data/GIS/data.htm)
Latitude, longitude, route ID, and name of light rail and commuter rail stations	Transit_Rout es.shp	Metropolitan Transportation Commission (http://www.mtc.ca.gov/maps_and_data/GIS/data.htm)
Block and lot information, account number and land value of each block	SFViewer	San Francisco GIS Application (http://gispub02.sfgov.org/website/sfviewer/INDEX.htm)
Population density of each parcel	Census 2000	American Fact Finder (http://factfinder.census.gov/home/saff/main.html?_lang=en)

Table 3-2 indicates how the geospatial data set, maps and other information were gathered for this study. Data about land use, parcel identification number, parks, rivers, lakes, and roads such as highways was collected from Cal-Atlas website offered by California government agencies, partners and stakeholders (Cal-Atlas, 2010). Table 3-2 also shows that the spatial information regarding bus stops, light rail stations, commuter rail stations, zoning, block, and land value was accessed from (MTC, 2010). Block ID, lot information, account number and land value of each block was gathered from the SFViewer website of the San Francisco government (SFViewer, 2010). Population density data was collected from the American Fact Finder website of the US Census Bureau (Bureau, 2000).

Figure 3-2. Parcels in San Francisco County

3.5.1 Parcels

All parcels in San Francisco County are shown in Figure 3-2. This file was captured from ArcGIS, and contains account number, and addresses in attribute table of each parcel.

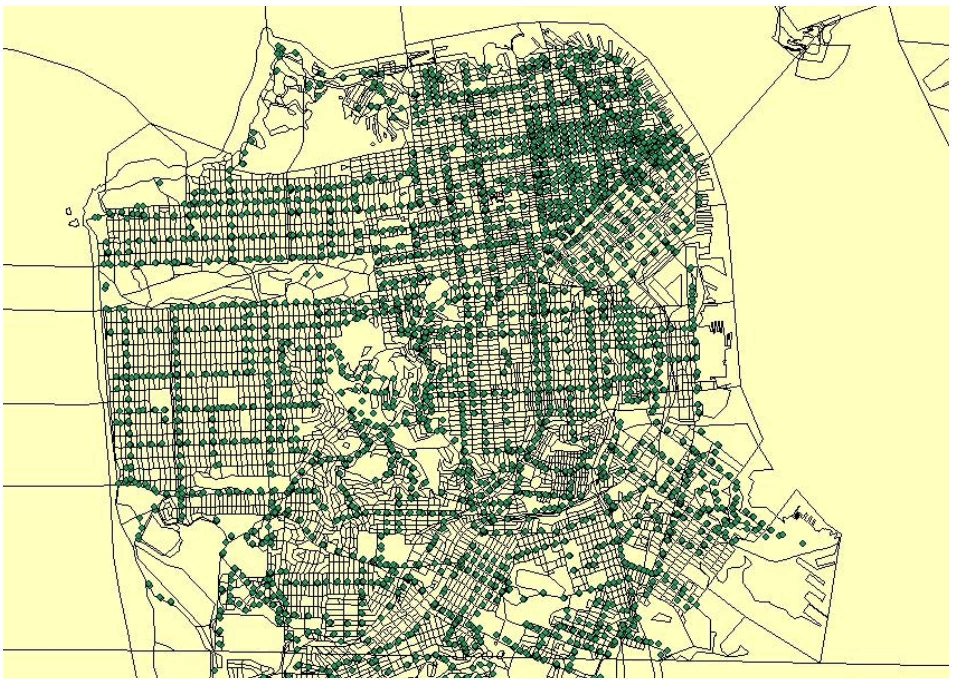


Figure 3-3. Bus stops in San Francisco County

3.5.2 Bus Stops

In Figure 3-3, all bus stops in San Francisco County are painted green. Using this GIS file, the distance between each parcel and bus stops was calculated.

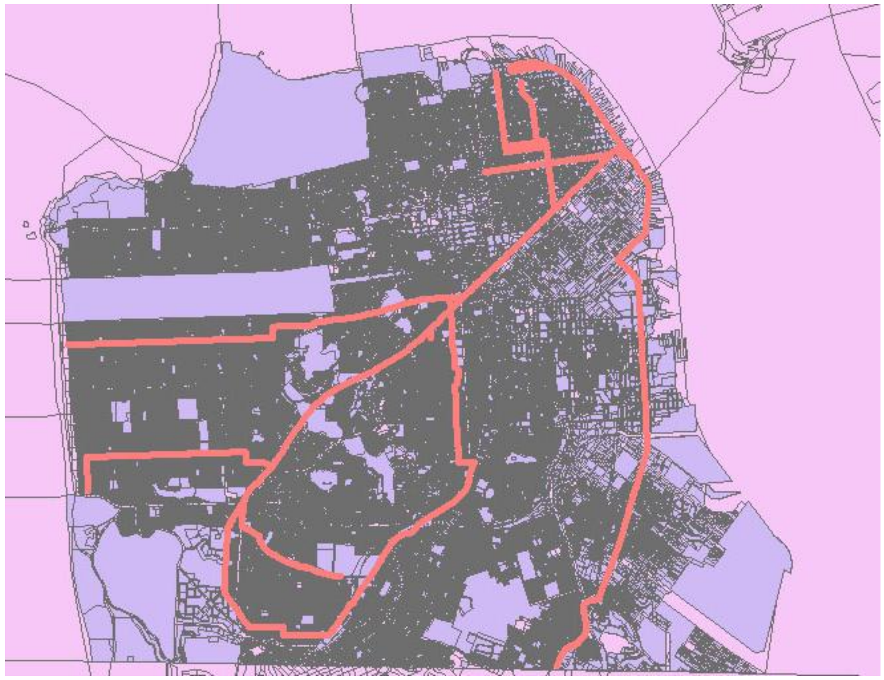


Figure 3-4. Light rail routes in San Francisco County

3.5.3 Light Rail Routes

In Figure 3-4, the light rail routes which include trolley, MUNI, and cable cars in San Francisco County are drawn in red line. Using this GIS file, the distance between each parcel and light rail station was calculated.

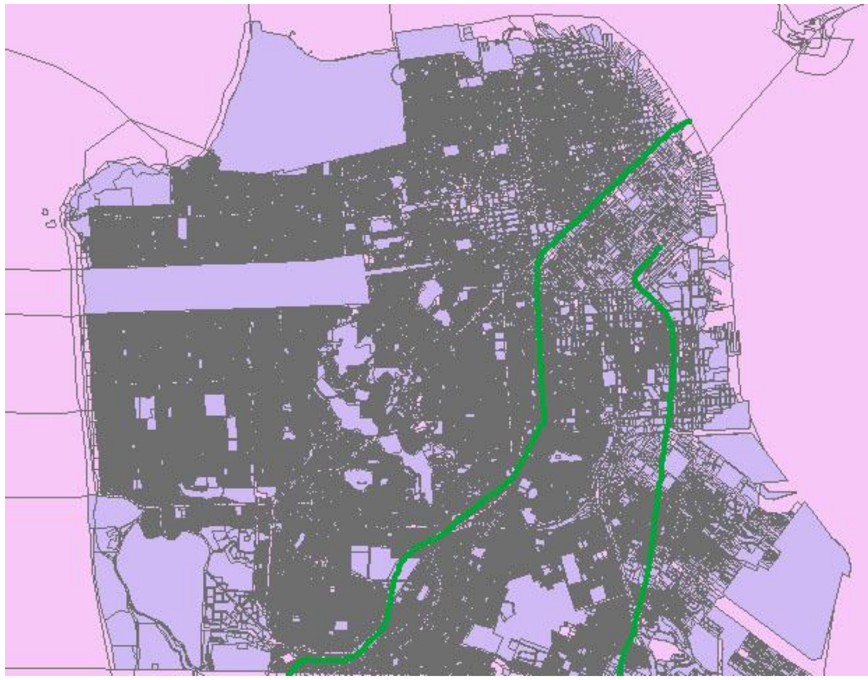


Figure 3-5. Commuter rail routes in San Francisco County

3.5.4 Commuter Rail Routes

In Figure 3-5, commuter rail routes including BART and Caltrain in San Francisco County are drawn in green line. Using this GIS file, the distance between each parcel and commuter rail stations was calculated.

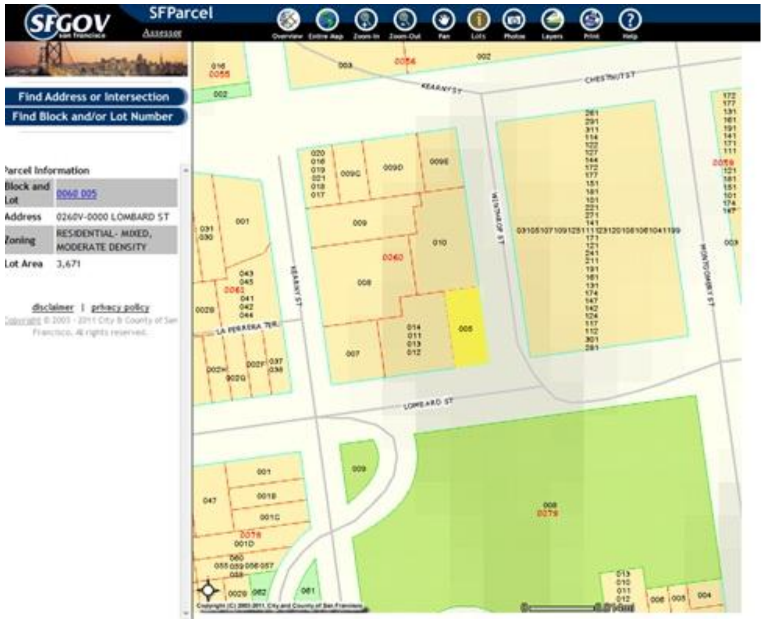
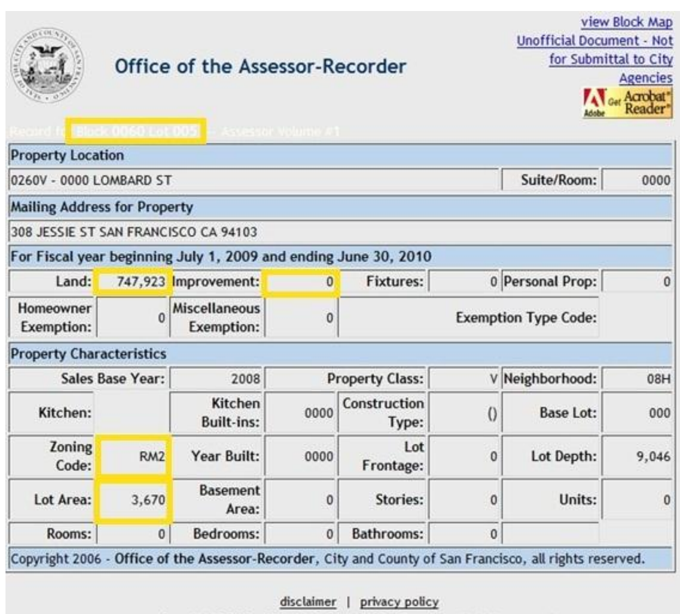


Figure 3-6. Land value data retrieval - phase 1



Office of the Assessor-Recorder

view Block Map
Unofficial Document - Not for Submittal to City Agencies

Record for Block 0060 Lot 005 - Assessor Volume #1

Property Location							
0260V - 0000 LOMBARD ST						Suite/Room:	0000
Mailing Address for Property							
308 JESSIE ST SAN FRANCISCO CA 94103							
For Fiscal year beginning July 1, 2009 and ending June 30, 2010							
Land:	747,923	Improvement:	0	Fixtures:	0	Personal Prop:	0
Homeowner Exemption:	0	Miscellaneous Exemption:	0			Exemption Type Code:	
Property Characteristics							
Sales Base Year:			2008	Property Class:	V	Neighborhood:	08H
Kitchen:		Kitchen Built-ins:	0000	Construction Type:	()	Base Lot:	000
Zoning Code:	RM2	Year Built:	0000	Lot Frontage:	0	Lot Depth:	9,046
Lot Area:	3,670	Basement Area:	0	Stories:	0	Units:	0
Rooms:	0	Bedrooms:	0	Bathrooms:	0		

Copyright 2006 - Office of the Assessor-Recorder, City and County of San Francisco, all rights reserved.

disclaimer | privacy policy

Copyright © 2003 - 2011 City & County of San Francisco. All rights reserved.

Figure 3-7. Land value data retrieval - phase 2

3.5.5 Land Value

Through the office of the Assessor-Recorder website, parcel information, land value, zoning code, and the area of lots were collected. Figures 3-6 and 3-7 describe the phases of data collecting (SFViewer, 2010).

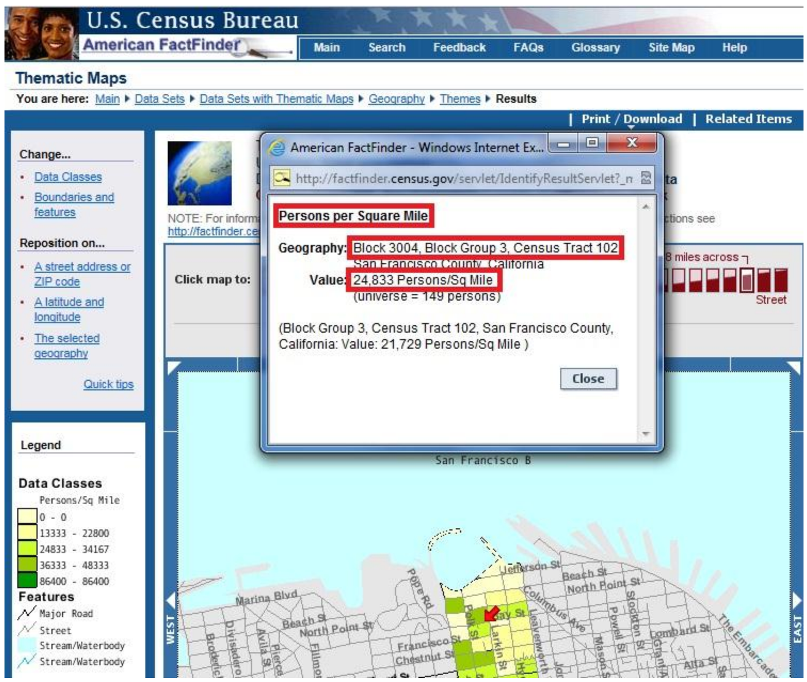


Figure 3-8. Data retrieval – population density in persons per square mile

3.5.6 Population Density

Figure 3-8 shows how the population density information for each parcel was collected through The American Fact Finder website offered by the U.S. Census Bureau.

Table 3-3 shows an example of the whole collected data.

Table 3-3. Data collection example

Zoning	ID	Land value($)	$/Acre/ 1,000,000	Population Density /10000	Lot size (Acre)	Number of bus stops meet LEED	Closest distance from bus stops (ft)	bus	Number of lightrail stations meet LEED	Closest distance from lightrail stations (ft)	light rail	Number of commuter rail stations meet LEED	Closest distance from commuter rail stations (ft)	commuter rail
RH-3	26030	250,179	3.16	1.2	0.079051	7	122.99	1	2	161.31	1	0	8,187.21	0
RH-3	44021	177,024	2.75	2.94	0.0644	7	297.96	1	2	1,212.92	1	0	7,199.37	0
RH-3	44022	177,024	2.75	2.94	0.0644	4	287.93	1	2	1,224.34	1	0	7,224.78	0
RH-3	87019	86,793	1.1	5.27	0.079028	5	786.98	1	2	1,517.46	1	0	4,615.48	0
RH-3	106005	12,864	0.19	5.27	0.068908	2	1,125.54	1	1	793.44	1	0	9,714.19	0
RH-3	106041	151,178	1.15	0.44	0.1311	3	334.00	1	2	1,176.63	1	0	3,682.10	0
RH-3	113011	12,637	0.17	0.44	0.07245	2	336.46	1	2	1,345.51	1	0	3,295.58	0
RH-3	134029	508,791	5.85	5.27	0.086963	7	325.47	1	3	1,273.91	1	0	3,152.92	0
RH-3	134031	549,494	5.8	0.44	0.09476	9	266.36	1	3	1,197.45	1	0	3,144.50	0
RH-2	501017	531,861	6.73	1.2	0.079028	5	187.60	1	3	768.13	1	0	7,384.46	0
RH-3	571014	1,167,355	14.77	5.01	0.079028	6	158.87	1	4	1,214.46	1	0	6,169.82	0
RH-2	641006	240,758	2.03	5.63	0.118519	3	248.72	1	1	805.15	1	0	5,162.12	0
RH-2	663002	123,478	0.69	3.27	0.17963	0	1,452.94	0	1	1,644.95	1	0	5,065.91	0
RH-3	679037	79,811	0.84	2.54	0.09476	3	369.37	1	0	4,229.97	0	0	6,801.04	0
RH-3	869007	208,086	2.74	4.13	0.075831	10	496.76	1	4	488.64	1	0	3,018.90	0
RH-1(D)	957016	1,369,959	11.92	1.26	0.114908	9	853.39	1	0	6,868.91	0	0	10,438.86	0
RH-1(D)	957020	532,087	5.77	1.26	0.092161	6	717.05	1	0	6,774.99	0	0	10,481.80	0
RH-1(D)	961021	523,888	2.15	1.26	0.244099	0	1,360.45	0	0	5,201.28	0	0	9,152.95	0
RH-1(D)	961023	327,427	2.15	1.26	0.151961	5	342.68	1	0	5,330.08	0	0	9,055.22	0
RH-2	1067034	228,891	3.32	1.86	0.068908	6	468.84	1	0	6,321.99	0	0	10,743.36	0
RH-2	1070002	51,353	0.4	1.86	0.129927	0	1,493.69	0	0	6,398.65	0	0	9,845.75	0
RH-3	1081030	36,139	0.42	2.66	0.086963	2	1,223.36	1	0	5,905.90	0	0	8,457.94	0
RH-3	1213011	193,955	2.45	3.19	0.079051	6	72.10	1	0	2,642.26	0	0	10,178.67	0

4. DATA ANALYSIS

4.1 Explanation of Variables

Seven variables are considered in this research, unit value, population density, lot size, and indicators of LEED criteria for bus stops, light rail stations, and commuter rail stations. Unit value and population density are continuous variables, and the two indicators are categorical variables. Table 4-1 and Table 4-2 show summary statistics for all variables.

4.1.1 Continuous Variables

Unit value, population density, and lot size are the continuous variables. Unit value of each parcel is the only dependent variable in this study. Two previous studies used dollars per square foot as the unit of land unit value (B. Joshi, 2009; Y. J. Park, 2009); however, millions of dollars per acre was used rather than dollars per square foot.

Independent variables, population density, and lot size of each parcel have been considered.

4.1.2 Categorical Variables

There are three categorical variables used in this research. In order to represent the relationship between public transportation accessibility and land value and determine whether or not a parcel meets the LEED criteria for the number of bus stops, light rail stations, and commuter rail stations are used as categorical values. One represents that the parcel meets LEED PTA criteria for each transit, and zero is assigned if it does not.

4.2 Overall Data Description

Table 4-1. Statistical description of mixed zones

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
unitvalue ($1,000,000/acre)	122	.20	28.66	4.7998	6.08788	37.062
bus	122	0	1	.9180	.27545	.076
lightrail	122	0	1	.6639	.47431	.225
commuter rail	122	0	1	.3607	.48217	.232
pdensity (10,000 persons/mile^2)	122	.36	15.72	4.3509	3.5424	12.549
Lotsize(acre)	122	.06	1.15	.1336	.13052	.017
Valid N (listwise)	122					

The mean of unit value is $4,799,800 per acre. It is assumed that 1 meets the LEED criteria for public transportation accessibility while 0 does not meet the criteria. It can be interpreted as the higher the mean value of each transportation system, the easier the access to public transportation of each parcel. For bus, it is 0.918, 0.6639 for light rail stations and 0.3607 for commuter rail stations. The mean value of population density in mixed purposed zones is approximately 43,509 persons per square mile.

Table 4-2. Statistical description of residential zones

	N	Minimum	Maximum	Mean	Standard Deviation	Variance
unitvalue ($1,000,000/acre)	308	.01	14.77	1.9222	2.47389	6.120
bus	308	0	1	.8994	.30135	.091
lightrail	308	0	1	.7208	.44935	.202
commuterrail	308	0	1	.3182	.46653	.218
lotsize	308	.06	6.88	.1621	.56827	.323
pdensity (10,000 persons/mile^2)	308	0.05	5.63	2.1228	1.04364	1.089
Valid N (listwise)	308					

In Table 4-2, the mean of unit value (in US dollars per acre) is $1,922,200 per acre which is less expensive than the mean value of mixed zoning in San Francisco which is $4,799,800 per acre. The highest unit value is $14,770,000 while the minimum value is only $100,000 per acre.

It is assumed that 1 meets the LEED criteria for public transportation accessibility while 0 does not meet the criteria. Thus, it can be interpreted that the higher the mean value of each transportation system, the easier the access to public transportation for each parcel. For bus, it is 0.8994, 0.7208 for light rail stations and 0.3182 for commuter rail stations. The mean of population density in the residential purposed zoning is about 21,226 persons per square mile.

4.2.1 Continuous Variables

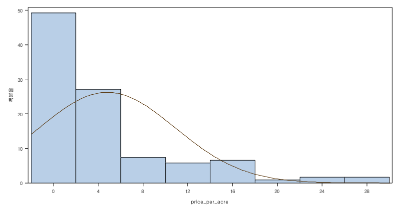
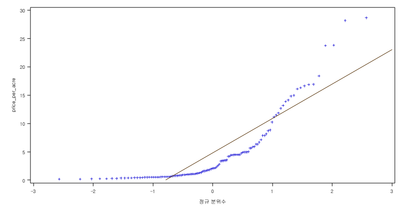
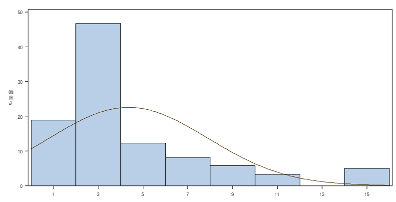
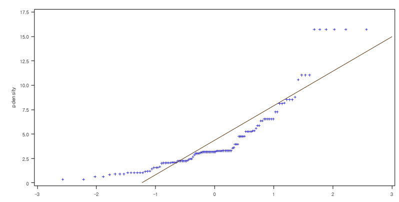
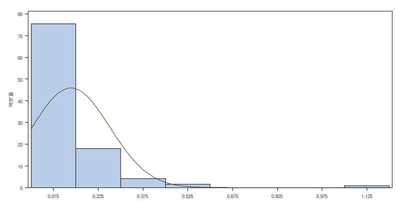
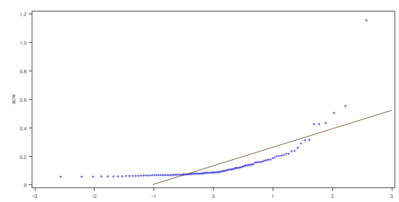
Variables	Histograms	Q-Q Plots
unitvalue ($1,000,000/acre)		
pdensity (10,000 persons/mile^2)		
Lot size (acre)		

Figure 4-1. Histograms and Q-Q plots of continuous variables in mixed zones

In Figure 4-1, all continuous variables are right skewed, which means they are not normally distributed. Since unit value is used as dependent variable in the regression model that should satisfy the normality assumption (Fan, 2010). Hence it seems that unit value needs to be transformed.

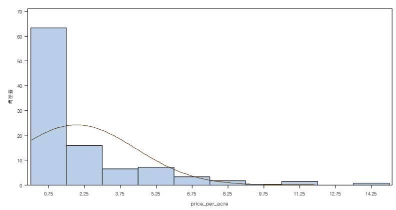
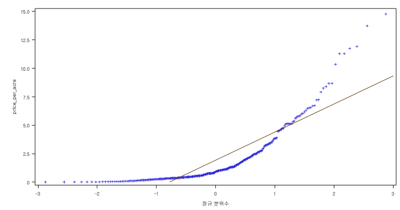
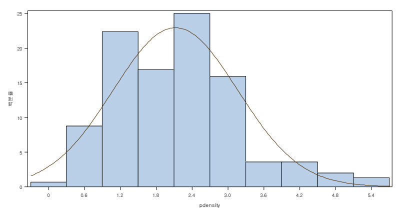
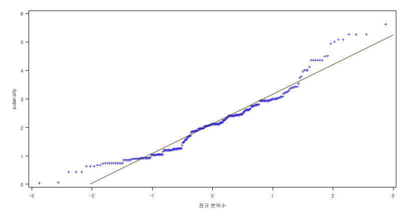
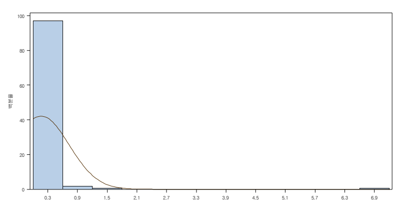
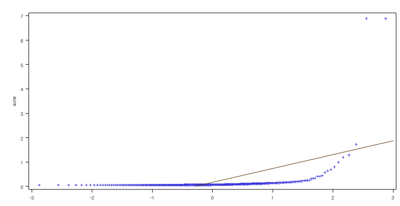
Variables	Histograms	Q-Q Plots
unitvalue ($1,000,000/acre)		
pdensity (10,000 persons/ mile^2)		
Lot size (acre)		

Figure 4-2. Histograms and Q-Q plots of continuous variables in residential zone

Same as in the mixed zone, all continuous variables in the residential zone are severely right skewed in Figure 4-2, but they do not seem to have severe outliers. Unit value of residential zone also needs to be transformed because it is not normally distributed (Osborne, 2010).

4.2.2 Categorical Variables

Three categorical variables are used in this research. To indicate satisfaction of LEED criteria for bus stops, light rail stations, and commuter rail stations. 1 indicates that the area satisfies LEED criteria, and otherwise zero is assigned.

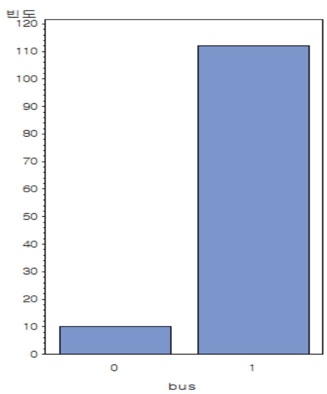
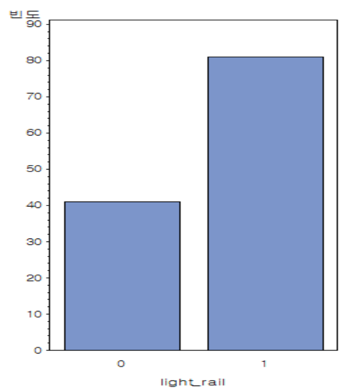
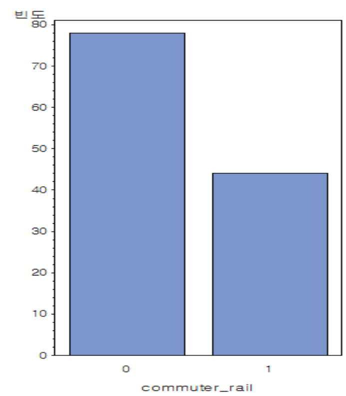
Bus	Light rail	Commuter rail
		

Figure 4-3. Bar chart of bus, light rail, and commuter rail in mixed zone

Shown in Figures 4-3 and 4-4, there are a few parcels which did not meet the LEED criteria for bus stops, and about 36% of parcels in the samples meet the LEED PTA criteria for commuter rail stations.

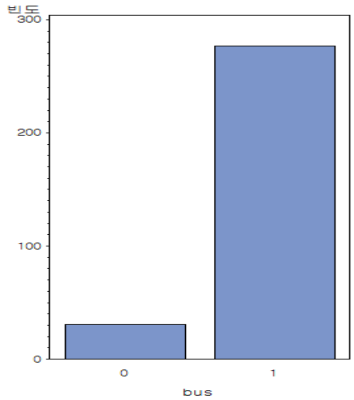
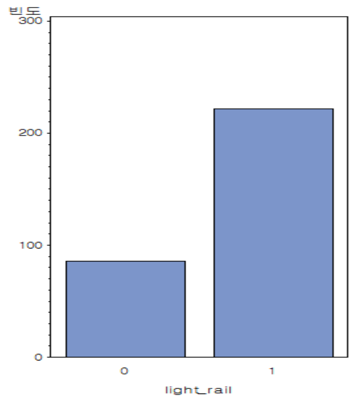
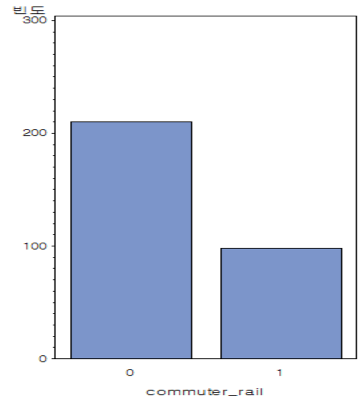
Bus	Light rail	Commuter rail
		

Figure 4-4. Bar chart of bus, light rail, and commuter rail in residential zone

4.3 Relationship Between Variables

4.3.1 Continuous Variables

There are several methods to check the relationship between continuous variables; Pearson correlation coefficient is the one most widely used. It has a value from -1.0 to 1.0. If the absolute value is close to 1, it shows a strong correlation, or if it is close to zero, it shows no relationship between two continuous variables. The sign indicates negative or positive correlation. We can also test whether the estimated correlation is significant or not by using a T-test. However, to use the Pearson correlation, variables should be normally distributed, and as noted above, they are not normally distributed. Therefore, we cannot fully trust these results, but can see positive or negative signs and relative size of the measurement for the relationship.

Table 4-3. Pearson correlation coefficients and significance test p-values between continuous variables in mixed zone

Pearson correlation coefficient (p-value)	Unit value	Population density	Lot size
Unit value	1.00000	0.30731 (0.0006)	-0.09184 (0.3144)
Population density	0.30731 (0.0006)	1.00000	-0.15815 (0.0819)
Lot size	-0.09184 (0.3144)	-0.15815 (0.0819)	1.00000

Table 4-4. Pearson correlation coefficients and significance test p-values between continuous variables in residential zone

Pearson correlation coefficient (p-value)	Unit value	Population density	Lot size
Unit value	1.00000	-0.02242 (0.6951)	-0.04226 (0.4599)
Population density	-0.02242 (0.6951)	1.00000	-0.08337 (0.1443)
Lot size	-0.04226 (0.4599)	-0.08337 (0.1443)	1.00000

In Table 4-3, population density is correlated with unit value because the P-value is less than 0.05. Otherwise, in Table 4-4, none of the continuous variables are correlated with unit value.

4.3.2. Categorical Variables

To find the relationships between categorical variables, a chi-square independent test was used. The test only shows whether or not two categorical variables are independent. Because category 1 does not mean 1 unit is greater than the category 0, it cannot be said that there are positive or negative correlations or strong or weak correlations between two categorical variables.

Table 4-5. Independent test for categorical variables in mixed zone

	Chi-square test statistics	p-value
Bus vs. Light rail	0.1996	0.6551
Bus vs. Commuter rail	0.1738	0.6768
Light rail vs. Commuter rail	18.5377	<0.0001

Table 4-6. Independent test for categorical variables in residential zone

	Chi-square test statistics	p-value
Bus vs. Light rail	3.3633	0.0667
Bus vs. Commuter rail	0.1233	0.7255
Light rail vs. Commuter rail	5.2017	0.0226

As shown in Table 4-5 and Table 4-6, it seems that light rail and commuter rail are not independent in both mixed and residential zone while bus is independent with both light rail and commuter rail. This means a parcel that meets LEED criteria due to qualifying light rail stations tends to have qualifying commuter rail stations.

4.4 Regression Model

Unit value is considered as the dependent variable, and others are considered as independent variables for the regression model. Objective of this model is to find the effects on unit value caused by changes in the value of independent variables. Therefore, the relationship between unit value and others needs to be checked. Also, independent variables should be independent of each other, but some of them are shown not to be independent in the previous section. Hence, the interaction effect needs to be considered (William Mendenhall, 1996).

4.4.1 Dependent Variable – Unit Value

4.4.1.1 Transformation of Dependent Variable

As shown in Section 4.2.1, unit values in both zones are not normally distributed, and it should be transformed for applying to the regression model.

Box-cox transformation is one method to make the model residuals normally distributed. Figure 4-5 through Figure 4-8 show the relationship between RMSE and lambda values. RMSE stands for “root mean square error.” It has a minimum error value when the lambda value has also a minimum value. The lambda value is 0 when the RMSE has the smallest value (Minbo Kim, 1993). There are two types of Box-Cox transformations; power transformation and log transformation. Log transformation is commonly used where lambda is zero, but the lambda values in this result are very close to zero. Thus, the natural log transformation was determined for the dependent variable, unit value for both zones in this study.

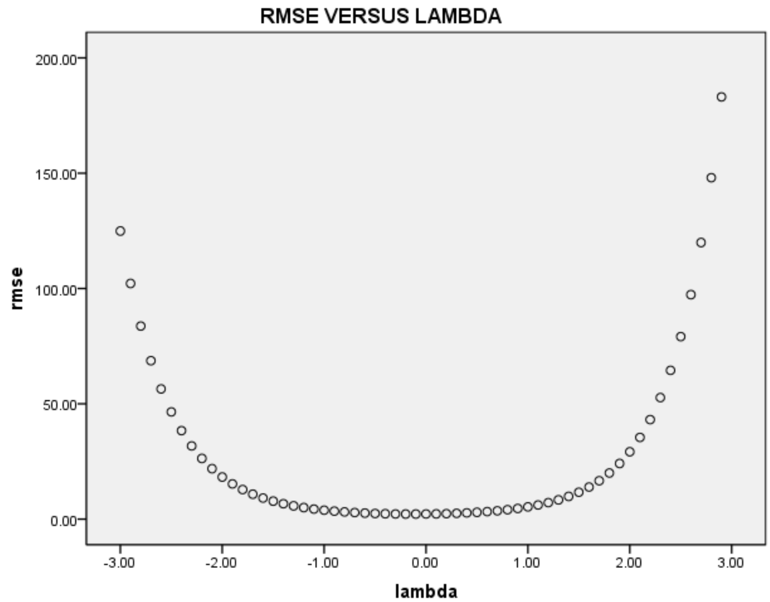


Figure 4-5. RMSE VS lambda graph for data transformation of mixed zoning

	lambda	rmse
25	-.60	2.62
26	-.50	2.45
27	-.40	2.33
28	-.30	2.25
29	-.20	2.21
30	-.10	2.20
31	1.53E-015	2.22
32	.10	2.28
33	.20	2.38
34	.30	2.52
35	.40	2.71
36	.50	2.94

Figure 4-6. RMSE VS lambda for data transformation of mixed zoning

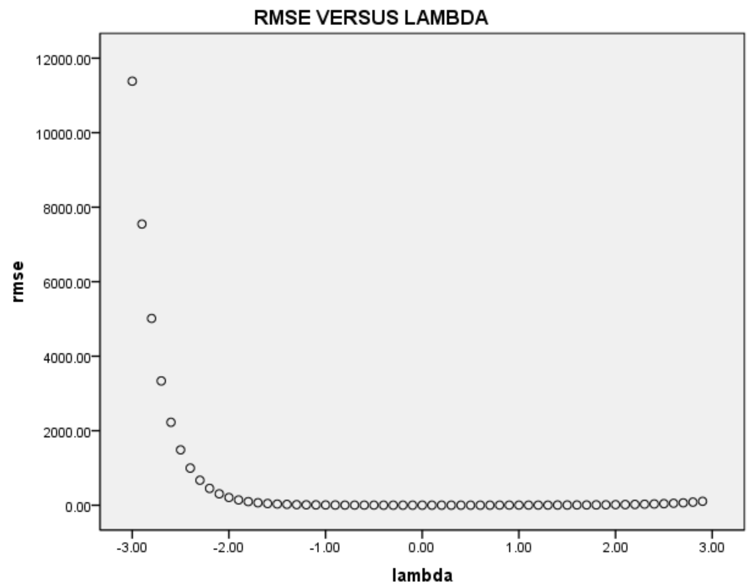


Figure 4-7. RMSE VS lambda graph for data transformation of residential zoning

	lambda	rmse
25	-.60	2.34
26	-.50	1.93
27	-.40	1.64
28	-.30	1.44
29	-.20	1.30
30	-.10	1.21
31	1.53E-015	1.16
32	.10	1.15
33	.20	1.16
34	.30	1.20
35	.40	1.27
36	.50	1.37
37	.60	1.49

Figure 4-8. RMSE VS lambda for data transformation of residential zoning

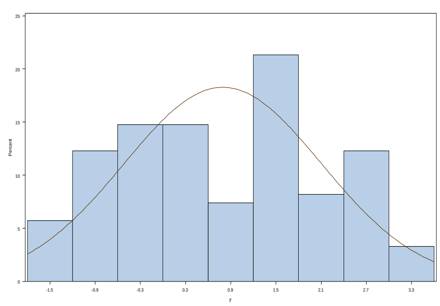
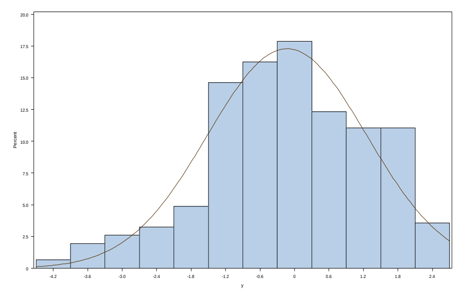
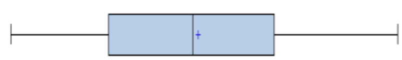
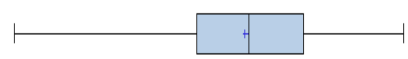
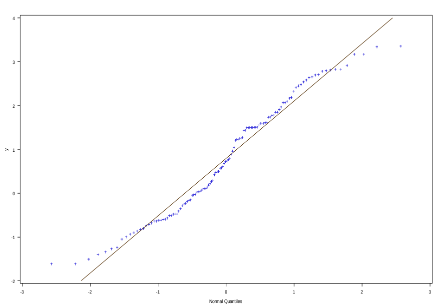
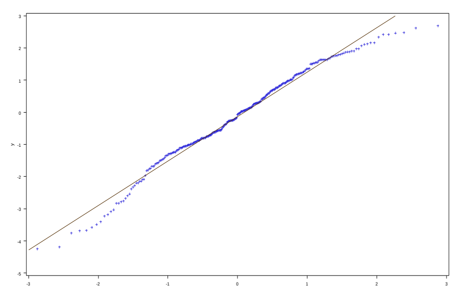
Graph	Mixed zone	Residential zone
Histograms		
Box-plots		
Q-Q plots		

Figure 4-9. Graphical descriptions of transformed unit value

Transformed unit values apparently are not exactly normal, but are also not seriously skewed. This can be ignored unless residual of the model does not satisfy normal assumption.

4.4.2 Unit Value and Continuous Independent Variables - Scatter Plots

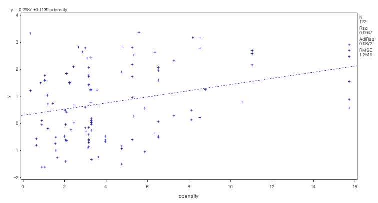
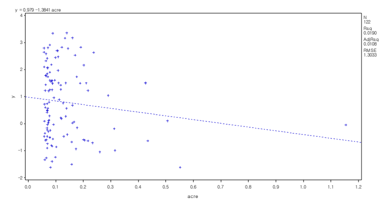
Unit value vs. population	Unit value vs. Lot size
	

Figure 4-10. Scatter plots of unit value vs. population density and lot size in mixed zone

In Mixed Zones, population density positively affects unit value and lot size negatively affects Unit Value . Pearson correlation in Section 4.3.1 supports this. Moreover, in residential zone, continuous variables are not correlated with unit variable, and Figures shown below show that slopes are almost zero for both cases.

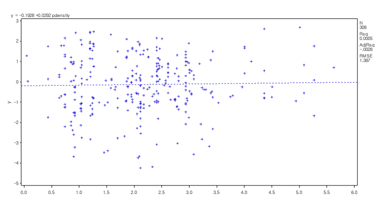
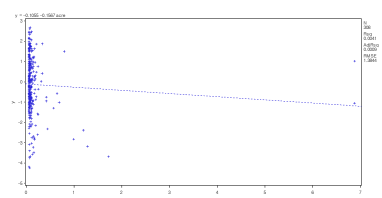
Unit value vs. population	Unit value vs. Lot size
	

Figure 4-11. Scatter plots of unit value vs. population density and lot size in residential zone

4.5 Unit Value and Categorical Independent Variables - Box Plots

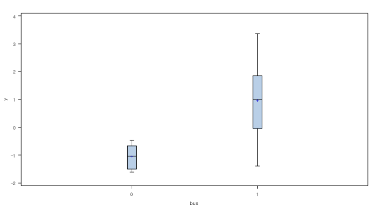
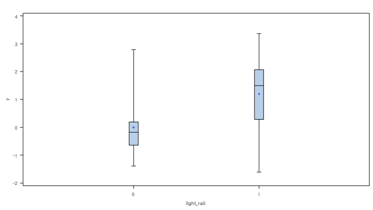
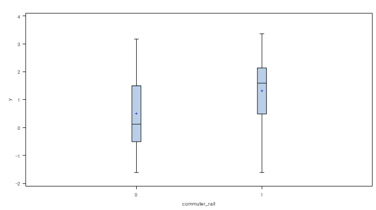
Unit value vs. Bus	Unit value vs. Light rail	Unit value vs. Commuter rail
		

Figure 4-12. Boxplots of unit value for bus, light rail, commuter rail in mixed zone

For each category of bus and light rail, a distribution of unit value appears to be different with different means in mixed zone. However, other cases seem to have the same distribution in each category. To decide whether these variables affect unit value or not, a statistical model should be used.

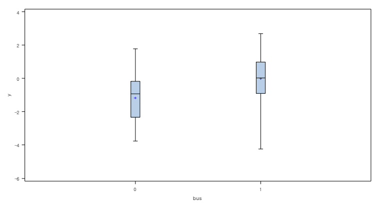
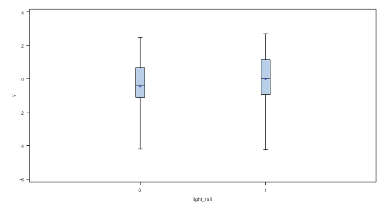
Unit value vs. population	Unit value vs. Lot size
	

Figure 4-13. Boxplots of unit value for bus, light rail, commuter rail in residential zone

4.6 Model Selection

Different from the hypotheses in Sections 1.4.1 and 1.4.2, the interaction terms between the variables should be considered in the predicting model. To develop appropriate models with significant independent variables, variable selection is necessary. There are three major variable selection methods, forward selection, backward elimination, and stepwise selection (Fan, 2010). In this research, stepwise selection is used to select significant variables. The full model before variable selection is shown below:

$$\text{Log(Unit value)} = \beta 0 + \beta 1 * B + \beta 2 * L + \beta 3 * C + \beta 4 * A + \beta 5 * P + \beta 6 * BL + \beta 7 * BC + \beta 8 * LC + \beta 9 * BLC + \varepsilon$$

B: 0: does not meet LEED criteria for bus stops 1: meets criteria

L: 0: does not meet LEED criteria for light rail stations, 1: meets criteria

C: 0: does not meet criteria for commuter rail stations, 1: meets criteria

A: lot size in acre

P: Population density of census block (persons per square mile /10,000)

BL: interaction term between B and L

BC: interaction term between B and C

LC: interaction term between L and C

BLC: interaction term among B, L and C

In each selection step, significance level of entry,0.05, and that of stay,0.1, were decided. Following steps are the explanation of stepwise selection. The mixed zone data is used for the steps. The residential zone model also followed same steps. Tables 4-7 through 4-10 explain the stepwise selection process for selecting significant variables.

Table 4-7. Stepwise selection phase 1 - first entry variables

Variables	Estimates	Standard error	T value	P-value
Intercept	-0.51938	0.74057	-0.70	0.4846
B	0.67646	0.74722	0.91	0.3672
L	-0.55295	0.86082	-0.64	0.5220
C	-0.35059	1.02848	-0.34	0.7338
B*L	1.60282	0.89379	1.79	0.0756
B*C	-0.52190	1.27037	-0.41	0.6820
L*C	-0.14216	1.52577	-0.09	0.9259
B*L*C	1.32612	1.71786	0.77	0.4418
P	0.03037	0.02908	1.04	0.2986
A	-0.99855	0.74454	-1.34	0.1826

When the full model is fitted, the most significant variable, B*L, is selected as the first entry variable.

Table 4-8. Select the next entry variable and decide to stay or remove for the first variables.

Variables	Estimates	Standard error	T value	P-value
Intercept	-0.30261	0.17583	-1.72	0.0878
B*L	1.44573	0.20948	6.90	<.0001
P	0.04779	0.02890	1.65	0.1008

When the second significant variable, P, was entered in the model, B*L is still significant, but P was not significant. Therefore, P will not be selected.

Table 4-9. Repeat step 2 for every variable

Variables	Estimates	Standard error	T value	P-value
Intercept	-1.04916	0.32731	-3.21	0.0017
B*L	1.32159	0.20794	6.36	<.0001
B	1.12282	0.36889	3.04	0.0029

When B was entered in the model, B*L is still significant, and B is also significant. Thus, B is selected as the second variable, but, in other cases, no one was significant.

Table 4-10. Select the next entry variable and decide to stay or remove for the first and second variables

Variables	Estimates	Standard error	T value	P-value
Intercept	-1.17090	0.33333	-3.51	0.0006
B*L	1.21405	0.21659	5.61	<.0001
B	1.10970	0.36637	3.03	0.0030
P	0.04594	0.02796	1.64	0.1030

When P was entered in the model, B*L and B were still significant, but P was not significant. So, P is not selected as the third variable. And also no other variable was selected as the third variables. Therefore, we stop variable selection in this step. Because the interaction term, B*L is selected, Light rail should be in the model to make model interpretable even though it was not significant.

4.7 Parameter Estimates

With selected variables, models that would be estimated are following, and Tables 4-11 and 4-13 show the estimates and significance test.

Mixed zone: $\text{Log(unit price)}=\beta_0+\beta_1 B+\beta_2 L+\beta_7 B*L+\varepsilon$ (1)

Residential zone: $\text{Log(unit price)}=\beta_0+\beta_1 B+\beta_2 L+\varepsilon$ (2)

Table 4-11. Parameter estimates and significance test for regression coefficients in mixed zone

Parameters	Estimates	Standard error	t	p-value
β_0 (Intercept)	-0.74577	0.51845	-1.44	0.1530
β_1(B)	0.81943	0.54576	1.50	0.1359
β_2(L)	-0.50565	0.66932	-0.76	0.4515
β_7(B, L)	1.82724	0.70099	2.61	0.0103

In this regression model each coefficient does not mean the effect of the factor with respect to the coefficient. Because all chosen independent variables are binary data, not continuous, the model became a cell mean model of a two-way ANOVA. Therefore, estimated parameters represent mean differences of each cell with two factorized independent variables.

Table 4-12. Means of appraised unit value of factorized cells

Bus stops \ Light rail stations	0	1
0	β_0	$\beta_0+\beta_2$
1	$\beta_0+\beta_1$	$\beta_0+\beta_1+\beta_2+\beta_7$

In Table 4-12, since only β_7 is significant, certain groupings can be made. Group1= {(0,0), (0,1), (1,0)}, Group2= {(1,1)}

Therefore, only when both bus stops and light rail stations meet LEED criteria does appraised unit value increase, otherwise, it does not change.

Table 4-13. Parameter estimates and significance test for regression coefficients in residential zone

Parameters	Estimates	Standard error	t	p-value
β_0(Intercept)	-1.37837	0.25947	-5.31	<.0001
β_1(B)	1.10097	0.25429	4.33	<.0001
β_2(L)	0.35700	0.17054	2.09	0.0371

For the residential zone model, when a parcel meets LEED criteria due to qualifying bus stops, appraised unit value increases. Independently with bus stops, when a parcel meets LEED criteria due to qualifying light rail stations, appraised unit value increases, but its increased rate is lower than that due to bus stops.

Goodness of fit test in ANOVA tables for the regression models in Tables 4-14 and 4-15 show that these models are very significant. Therefore, these models are appropriate and can be used for predicting unit value.

Table 4-14. ANOVA table of the model in mixed zone

Source	df	Sum of squares	Mean square	F	p-value
Model	3	80.89790	26.96597	25.08	<.0001
Error	118	126.87035	1.07517		
Total	121	207.76824			

Table 4-15. ANOVA table of the model in residential zone

Source	df	Sum of squares	Mean square	F	p-value
Model	2	45.10895	22.55448	12.65	<.0001
Error	305	543.83585	1.78307		
Total	307	588.94481			

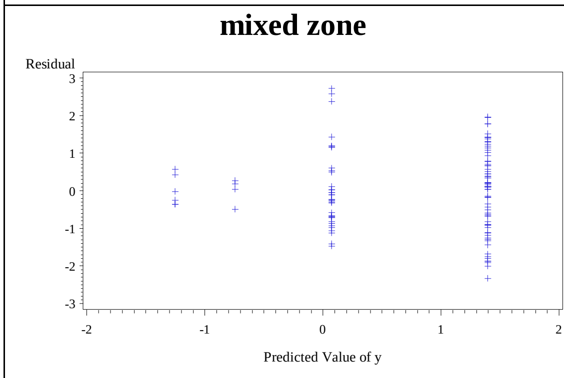
For mixed zone model, the adjusted R^2 is 0.3738, and for residential zone model the adjusted R^2 is 0.0705.

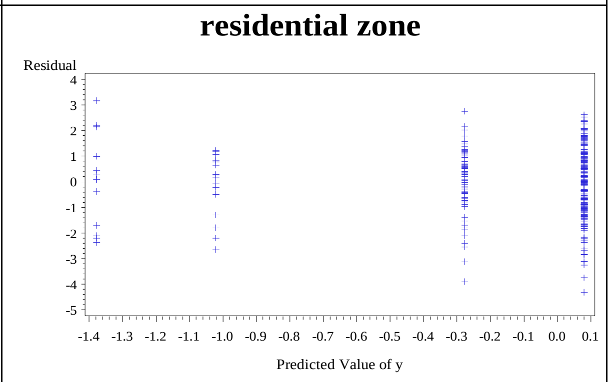
4.8 Residual Assumptions Check

In the linear regression model, there are two conditions that should be satisfied for error term, which is ε in equation (1) and (2) above. The first condition is that errors should be independent of each other. The second condition is that variance should be constant. The last one is that errors should be normally distributed. The first condition can be said to be already satisfied because all samples are randomly selected, which means, all errors are independently generated.

4.8.1 Residual Plot

A residual plot can be used to see whether the variance is constant or not.

<table>
<tr><td>Mixed zone model</td><td>Residential zone model</td></tr>
<tr><td>


</td><td>


</td></tr>
</table>

Figure 4-14. Residuals vs. fitted values

In Figure 4-14, residuals have no certain pattern as a function of fitted values. Therefore, the constant variance assumption is satisfied in both cases.

4.8.2 Normality Test

To check the normality assumption, box-plots and Q-Q plots were used. There are some tests for numerically testing normality such as the Kolmogorov-Smirnov test or Shapiro-Wilks test, but these tests have high power when there are many observations, and so they are too sensitive to accept the null hypothesis, which is normally distributed. In this research, only graphical approaches are used.

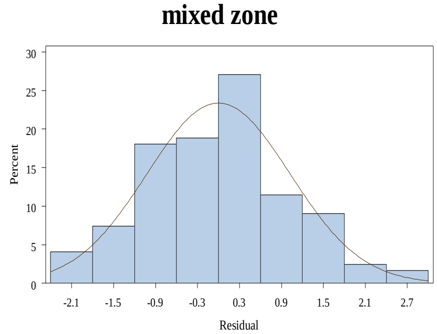
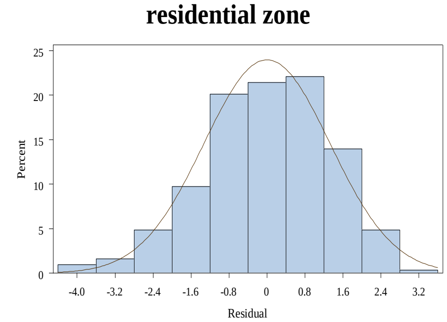
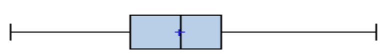
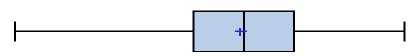
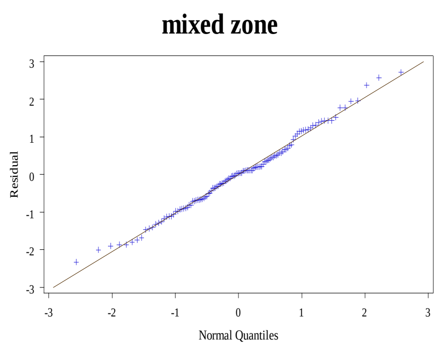
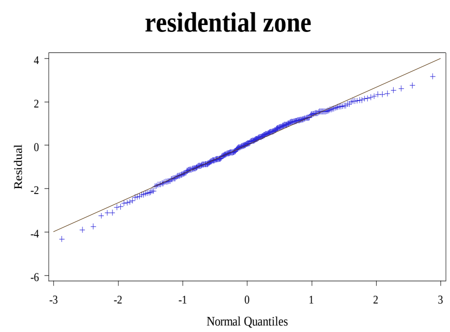
<table>
<tr><th>Graph</th><th>Mixed zone</th><th>Residential zone</th></tr>
<tr><td>Histograms</td><td>
</td><td>
</td></tr>
<tr><td>Box-plots</td><td></td><td></td></tr>
<tr><td>Q-Q plots</td><td>
</td><td>
</td></tr>
</table>

Figure 4-15. Normality tests - histograms, box-plots, and Q-Q plots for residuals

Figure 4-15 shows that residuals satisfy the normality assumption. Histograms and box-plots are symmetric with zero means, and Q-Q plots have almost a straight line.

4.9 Findings

4.9.1 Modeling Result

To predict unit value, a regression model was conducted with independent variables that mostly affect the unit value. Following are the estimated model equations.

Mixed zone: Log(unitvalue)= -0.74 577+0.81943*B-0.50565*L+1.82724*BL

Residential zone: Log(unitvalue)=-1.37837+1.10097*B+0.357*L

Where B is an indicator of LEED criteria for bus, L is an indicator of LEED criteria for light rail stations. BL is an interaction of two indicators of LEED criteria for bus and light rail.

The adjusted R^2 for mixed zone is 0.3738, and for residential zone model the adjusted R^2 is 0.0705. Hence, 37.4 % of the variability in the transformed unit value of the parcels can be explained by the significant independent variables, whereas 62.6% of the variability was explained by some other factors which are not considered for mixed zone in this study. In addition, for residential zone, only 7% of the variability in the transformed unit value of the parcels is explained by the significant independent variables.

4.9.2 Results and Interpretation

For the mixed zone model, interaction of bus stops and light rail stations was used. For residential zone model, only two main effects, bus stops and light rail stations, were significant. Hence, the first and second hypotheses are accepted, and other hypotheses are rejected. When a land parcel that meets the requirements for LEED Public Transportation Access credit due to qualifying bus stops, the appraised unit value is higher than when it does not meet the requirements. And when a land parcel that meets the requirements for LEED Public Transportation Access credit due to qualifying light rail stations, the appraised unit value is higher than when it does not meet the requirements.

5. CONCLUSION

Sustainable buildings have now become more popular than in the past, but the price effect of green building ratings has scarcely been calculated (Greg Kats, 2003). Hence, this study aimed to formulate a predictive model for estimating the value of real estate in San Francisco County, CA.

As a result of this study, the two statistical models above are suggested to predict the unit value of unimproved parcels in each zone in San Francisco County, CA. Findings show that the LEED PTA criteria and population density effect land unit value. Even though population density was not included in the final prediction model at this time, the correlation tells that the effectiveness of this factor is still significant. Findings indicate that if a parcel meets the PTA criteria, the land value of it tend to be higher. It can be interpreted as there is a relationship between a land value and PTA criteria.

In Joshi's study (B. B. Joshi, 2011), the number of public transit stations was used as the independent variable for predicting models. The statistical models created by the multiple regression method, suggested that the LEED criteria influenced the appraised value of properties in Houston, Texas as in San Francisco. However, in this study, number of qualifying bus and rail stops as well as the distance from the parcel to bus stops and light rail stations from each parcel were used as independent variables. To create a more accurate predicting model, for San Francisco the interaction term was considered, and the predicting models showed that the LEED PTA criteria positively influenced the appraised unit value in San Francisco County, California. Moreover, population density of each census block was considered as another main factor in this study, and was found to positively influence unit value, but was not included in the prediction model.

The result of this research is expected to encourage developers to make better site decisions for new buildings to accelerate the use of sustainable construction.

5.1 Limitations

Only unimproved parcels within San Francisco County, California are targeted in this research.

San Francisco County updates the GIS Parcel data once a year. Therefore all data, appraised unit land value and parcel information, gained from the website is for fiscal year beginning July 1, 2009 and ending June 30, 2010.

Population density in this study comes from Census 2000. Population density of the block group where each parcel is located did not exactly reflect the exact population density of the parcels, but data of each block group was used in the research because it is the most detailed spatial unit provided by the U.S. Census Bureau.

This research focused only on Sustainable Sites Credit 4.1-Alternative Transportation: Public Transportation Access criteria of LEED-NC Version 3.0.

The actual walking distance from the door of a building to the entrance of the bus stops or rail stations should be considered in the study. However, since unimproved parcels do not have buildings on them, the distance from the centroids of given parcels to bus stops or rail stations was calculated.

5.2 Delimitation

Statistical models of this research reflect only unimproved parcels in San Francisco, CA, and may be less effective if applied to other areas. This research focused only on the criteria of LEED-NC for Sustainable Sites Credit 4.1 section, Alternative Transportation: Public Transportation Access. Findings and corresponding interpretations can only be applied to this limited LEED credit.

Data gathered in this study reflects the conditions over a decade ago. The U.S Census Bureau offers the census information only every 10 years in years ending with 0. In this research, population density data is gathered from Census 2000 which is based on the data of the year 1999 because Census 2010 is not published yet.

Since sale price information is difficult to obtain, appraised value of parcels is the best alternative for making meaningful comparisons. However, only unimproved land with zero improvement value was considered in the study.

The correlations of five independent variables were analyzed at this time:

1. Whether or not a parcel meets the LEED criteria for the number of bus stops
2. Whether or not a parcel meets the LEED criteria for the number of light rail stations
3. Whether or not a parcel meets the LEED criteria for the number of commuter rail stations
4. Area of a given parcels
5. Population density of census block

Many other independent variables were not counted in the current research, but could possibly be studied in future research.

5.3 Recommendations for Future Research

The opportunities for additional areas of research are numerous, mostly as a result of lack of widely spread implementation of similar research. This lack of use creates a void of knowledge because of the near absence of any comparable experience. Future research can be implemented in numerous cities in the U.S or other countries if needed.

The adjusted R^2 value of the statistical model was not higher than expected. Even though the result is still credible, it will be more useful if other possible variables are incorporated to create a more credible regression model with more reliable factors in future research. Other LEED criteria for sustainable sites or water efficiency are suggested in future research.

REFERENCES

Amtrak timetable. (2011). from http://www.amtrak.com/servlet/ContentServer?c=am2Station&pagename=am%2Fam2Station%2FStation_Page&cid=1229726269959

Breheny, M. J. (1992). Sustainable development and urban form. *European Research in Regional Science, 2*, pp. 31-33

Brundtland, G. H. (1987). *World commission on environment and development: Our common future*, United Nations, Retrieved from http://www.un-documents.net/a42r187.htm

Cal-Atlas. (2010). GIS data. Retrived from http://atlas.ca.gov/

Christie, L. (2008). New Yorkers are top transit users. *CNN Money.com.* Retrieved from http://money.cnn.com/2007/06/13/real_estate/public_transit_commutes/index.html

Corbett, C and Muthulingam, S (2007). *Adoption of voluntary environmental standards: The role of signaling and intrinsic benefits in the diffusion of the LEED green building standards*. Retrived rom http://www.environment.ucla.edu/media/files/pdf_CC34.pdf

Fan, R. (2010). STAT 652 class notes, Texas A&M University, College Station, TX.

Feagin, R. (2010). AGSM 461 class notes, Texas A&M University, College Station, TX.

Ginger Christ, A. F. (2011). Number of LEED-certified buildings growing. *Austin Business Journal*. Retrieved from http://www.bizjournals.com/austin/news/2011/04/14/number-of-leed-certified-buildings.html

Greg Kats, L. A., Berman. A, Mills. E, and Perlman. J. (2003). The costs and financial benefits of green buildings. From http://evanmills.lbl.gov/pubs/pdf/green_buildings.pdf

Harding, V. R. (2009). Green buildings: Understanding what LEED is all about, Retrived from http://www.pepperlaw.com/publications_update.aspx?ArticleKey=1624

Huang, A. L., Chapman, R. E., and Butry, D. T. (2009). Metrics and tools for measuring construction productivity: *Technical and Empirical Considerations, NIST Special Publication 1101*, pp. 30-33.

Joshi, B. B. (2009). *Prediction of unit value of un-improved parcels of Harris County, Texas using LEED sustainable sites criteria of public transportation access.* M.S. Thesis, Texas A&M University, College Station,TX.

McFarland, W. (2007). Atlas of HIV/AIDS in San Francisco, CA 1981-2000, Retrived from http://www.sfdph.org/dph/files/reports/RptsHIVAIDS/SFAtlasHIVAIDS19912000/web00.pdf

Minbo Kim, R. C. H. (1993). The Box-Cox transformation of variables in regression. *Empirical Economics, 18*, 307-319.

MTC. (2010). Transportaion data. from http://www.mtc.ca.gov/about_mtc/

Osborne, J. W. (2010). Improving your data transformations: Applying Box-Cox transformations as a best practice. *Practical Assessment, Research, and Evaluation, 15(12)*, 1-9.

Park, Y. J. (2009). *Predicting the unit appraisal value of the unimproved and private land in the City of Houston by LEED sustainable site credits.* Ph.D. Dissertation, Texas A&M University, College Station, TX.

Puentes, R. (2004). Transportation finance: Current issues, challenges, and impacts on local governments. Retrived from http://www.brookings.edu/~/media/Files/rc/speeches/2004/1203metropolitanpolicy_puentes/20041203_transfinance.pdf

Puentes, R. (2008). *Beginning again: A metropolitan transportation vision for the 21st century*. Retrived from http://www.brookings.edu/~/media/Files/rc/testimonies/2008/0409_transportation_puentes/0409_transportation_puentes.pdf

Redeclift, M. (1993). Sustainable development: Needs, values, rights. *Environmental values 2*, 3-20

Report of the world commission on environment and development. (1987). Retrieved from http://www.un.org/documents/ga/res/42/ares42-187.htm.

Rosenberg, M. (2011). Population density. *About.com.* Retrived from http://geography.about.com/od/populationgeography/a/popdensity.htm.

SFViewer. (2010). San Francisco GIS application. Retrived from http://gispub02.sfgov.org/website/sfviewer/INDEX.htm

Shamsi, U. M. (2002). *GIS tools for water, wastewater, and stormwater systems*., New York: American Society of Civil Engineers Publications.

Travel resources: Public transportation. (2011). from http://www.sfgate.com/travel/resources/transit/

U.S Census Bureau. (2000). *Census 2000*, Retrived from http://factfinder.census.gov/home/saff/main.html?_lang=en

U.S Green Building Council. (USGBC) (2008). LEED for new construction and major renovations version 3.0. , Washington, DC: US Green Building Council.

Weber, A. (2010). *Touchstone for sustainable development? The promises and pitfalls of LEED-ND.* Ithaca, NY: Cornell University.

William Mendenhall, T. S. (1996). *A second course in statistics - regression analysis* (5 ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

VITA

Name: Hyun Jeong Cho

Address: Department of Construction Science,
Langford Building A, College of Architecture,
Texas A&M University, College Station, TX 77843-3137

Email Address: leoninee@gmail.com

Education: B.E., Architectural Engineering, Kyung Hee University, 2006
M.S., Construction Management, Texas A&M University, 2011