Right Size Parking Project
King County Metro Transit

Research Methods
Phase II – Model Development

November 28, 2011

Prepared by:
Center for Neighborhood Technology

BACKGROUND

King County Metro has been awarded a grant in the Federal Highway Administration (FHWA) Value Pricing Program. The project will assemble local information on multifamily residential parking utilization to guide parking supply and management decisions in the future.

As part of the grant effort, research will be conducted in two phases:

- **Phase I** – Site selection and field data collection consisting of on-site parking inventory and utilization counts and assembly of physical building and pricing information.
- **Phase II** – Independent variable data collection, statistical analysis, and model development will be done to predict parking utilization.

A literature review prefaces the methodology discussion for these two phases of the research. Beginning with an overview of the present standards for estimating parking demand and utilization, the literature review highlights studies which show that parking is often oversupplied. Many studies focused on the relationships between parking demand and household socio-demographic characteristics, housing type, qualities of the built environment, and parking price and supply are reviewed in consideration of the imbalance between supply and demand for parking at multifamily residential properties. The literature review also considers efforts to model these relationships as well as auto ownership models and their relationship to parking demand. Additionally, data sources that assess auto ownership or vehicle availability can potentially serve as proxy measures for estimating parking demand and the applicability of these sources is briefly addressed at the conclusion of the review.

The Phase I research methodology addresses the collection of parking utilization data that will help King County assess parking utilization in existing multifamily residential buildings to learn more about resident's parking needs. Approximately 240 properties will be the subject of parking field counts to assess residential parking utilization. Due to time and budget limitations, a specific site selection methodology was developed to ensure a representative sample is collected within the constraints of the project. In addition, information about the physical building characteristics and parking/housing pricing will be collected to support the Phase II research.

This document covers the Phase II research methodology. The collection of parking field data, as detailed in the Phase I methodology, will enable statistical analyses and the development of a model predicting parking utilization at multifamily residential developments based on a set of independent variables tested in regression analysis. Independent variables selected will capture the effects of housing characteristics, neighborhood household characteristics, accessibility, built form, and parking pricing and supply on parking utilization. The resulting predicted utilization will be displayed on a website in which users will have the ability to incorporate development specific details to assess the resulting estimated parking utilization.

RESEARCH OBJECTIVES

As discussed in the literature review, a lack of consensus exists on the factors that drive demand for parking and utilization and account for the variation in auto ownership. While socio-demographic, housing, and built environment variables have all been shown to have an impact on parking and vehicle availability, their relative influence is a source of debate. There is more agreement on the fact that parking supply and pricing have a significant impact on parking demand and auto ownership, but these variables have been understudied. The appropriateness of using data on auto ownership as a proxy for parking demand is also an unresolved question. This research will attempt to address and provide clarity on these issues.

This Phase II research methodology is designed to identify and analyze independent or explanatory variables thought to affect multifamily residential parking utilization. A model will be developed to capture these relationships and estimate parking utilization as a function of the variables selected. The collection of field data will enable a unique analysis of variables often difficult to capture, including the supply and pricing of parking in conjunction with the resulting utilization. While the occupancy counts collected in Phase I field work will serve as the primary dependent variable representing parking utilization, the appropriateness of using data on registered vehicles as a proxy for parking utilization will also be analyzed. A website will be developed using the parking utilization model and will provide estimates to guide stakeholders' decisions about building new parking and managing existing parking. The objectives of the Phase II research design are to explain how the research will:

- Identify independent variables to be tested in regression analysis of occupancy counts collected in Phase I
- Conduct statistical analyses to test independent variables' significance in predicting parking utilization
- Develop model of parking utilization using regression analysis
- Test alternative methods of estimating parking utilization by evaluating correlation between Department of Licensing registration data and collected occupancy counts as to enable greater replicability without the need for field work
- Provide a set of standard and replicable methods to quantify parking utilization impacts that are conducive to an interactive website application

Modeling Parking Utilization

Using the parking occupancy counts collected in the field work described in Phase I as the dependent variable representing parking utilization, regression analysis will be used to estimate the relationship to several independent or explanatory variables. The resulting regression function will be used to model parking utilization.

The dependent variable of parking utilization will be defined as observed parking occupancy divided by the number of occupied dwelling units in the building. One potential difficulty or limitation of this definition may occur if the field data collection results in a case where 100 percent of the parking spaces are utilized. In these cases, there will be no way to determine if this is the result of supply perfectly meeting demand or parking being undersupplied or underpriced. In other words, is there a significant level of overflow that is not being accounted for in the occupancy counts on-site? Because the validity of these utilization rates cannot be determined, data from the Department of Licensing registrations will be considered to better understand the magnitude of the potential overflow. Determinations will be made on a case by case basis as to the appropriateness of the property's inclusion in the regression.

To construct this regression analysis, independent variables must be selected and tested. As highlighted in the literature review, little consensus exists on the predominant drivers of parking demand, although individual studies provide insight on variables to be tested and overall categories of variables to consider. The following theoretical framework has been constructed to guide the regression analysis and model development.

Parking Supply and Price:
As the two predominant indicators of demand, it is believed that parking supply and price will have a large impact on parking utilization. As basic theory suggests, low supply should correspond to low utilization and high prices should also indicate low utilization. Clearly, this can vary depending on context, but this basic trend should hold true in the data. Because price and demand (in this case, utilization) can both change in response to changes in each other, this can cause a loop of causality. To address these potential issues of endogeneity, a two-stage approach will be considered.

Property/Development Characteristics:
Characteristics specific to the properties studied have been shown to be explainers of parking utilization. Higher density developments (both by lot size as well as by floor space) and lower rents have both correlated with low utilization rates.

Neighborhood Household Characteristics:
To the extent that neighborhood demographics reflect the building/development demographics, high income, large household size, and many commuters per household in the neighborhood should all correlate with high auto ownership and thus high parking utilization rates.

ACCESSIBILITY:
High levels of access to public transit, jobs, and services are all expected to correspond with low parking utilization rates. Conversely, poor access requiring high dependence on auto travel should lead to high parking utilization.

BUILT FORM AND DEVELOPMENT PATTERNS:
Much research on auto ownership and auto use has highlighted the significance of built form and development patterns on auto dependence. High density, interconnected street networks, and a mix of land uses all correlate with low auto ownership and use, and therefore, should correspond with low parking utilization.

Because factors can be captured and represented in many different ways with many different variables, an extensive list of potential explanatory variables will be analyzed (see Attachment A). For example, while it is expected that transit access will correlate with parking utilization rates, the best measure of transit access to best explain utilization rates is unknown. Due to this, potential independent variables have been grouped into these five categories—parking supply and price, property/development characteristics, neighborhood household characteristics, accessibility, and built form and development patterns—enabling consideration of the greatest number of possible variables to capture these factors.

The dependent variable in the regression, parking spaces utilized per occupied dwelling unit, will be collected in the field work at the address/parcel level. This will therefore be the level of analysis. Independent variables will be related to this analysis level in one of two ways: (1) in cases of points of interest (e.g., job locations), gravity measures will be used to aggregate points weighting by their distance from the parcel centroid; and (2) in cases of average spatial characteristics (e.g., residential density), variables will be aggregated to a half mile buffer around the parcel centroid from the smallest geographic unit available (e.g., Census block group).

Within each group of independent variables, the correlation will be tested between each independent variable and the dependent variable. The dependent variable will be regressed against the independent variable most highly correlated first, with additional independent variables added to the fit, one at a time, based on the strength of the correlation. A t-value, indicating the significance of the independent variable in describing the relationship to the dependent variable, in conjunction with the R-squared value of the fit, will be used to assess the inclusion of the additional variables. As a general approach, an independent variable will be included in the fit if the R-squared value increases (even marginally) and the t-value is greater than one. This approach will be somewhat flexible, however, to allow for consideration of factors such as how previously added variables are impacted by the inclusion of additional variables. This process will serve to isolate the set of best independent variables providing the greatest fit within each category by eliminating highly correlated variables with redundant influence. For example, residential density and gross density will likely both correlate with the dependent variable. However, if residential density shows a stronger correlation and is therefore fit first, gross density, when added to the fit, may have a high t-value but not increase the R-squared value. This will indicate that while gross density has a strong relationship with the dependent variable, this impact is redundant with that of residential density.

Once the independent variables have been narrowed to the list that best captures five factors identified in the theoretical framework, this process of regressing the occupancy counts against each independent variable one at a time will be repeated, this time without the categorization. Again, only those independent variables that add to the overall explanatory power of the fit will be included in the final regression function. The best fit that is obtained will represent the extent to which the set of independent variables selected describe the variation in parking utilization as observed in the occupancy counts. Further research, if necessary, will be suggested to assess the variation unable to be described.

ALTERNATIVE METHODS OF ESTIMATING PARKING UTILIZATION

Because field data collection is difficult, time consuming, and therefore expensive, it is desirable to represent parking utilization with more readily available and frequently updated data sources, as this will increase the capacity of this research to be replicated in areas where field work is not feasible. As the literature suggests, various measures of auto ownership have been tested as proxies for parking demand or utilization with mixed results. The US Census previously, and now the American Community Survey (ACS), report Vehicles Available, but at the smallest geographic scale of a Census block group. Clearly parcel level data within a block group can vary significantly. Vehicle registrations (available in King County from the Department of Licensing) provide address level data, but questions persist on the potential mismatch between the address of vehicle registration and the home location parking utilization (e.g., vehicles registered in areas with lower insurance costs or taxes).

In the interest of estimating parking utilization without field work, this research will aim to assess the appropriateness of using the Department of Licensing (DOL) registration data or Vehicles Available as proxies for parking utilization. To do so, the utilization rates at the properties where counts were collected will be compared with the DOL registration data for those addresses and the ACS data for the surrounding Census block group. The correlation between the datasets will be tested using a significance level of alpha equal to 0.10 and the critical value of the correlation will be determined after the sample size is known. If the correlation is deemed statistically significant, regression analysis will be used to determine if the regression coefficient is consistent with 1.0 (with a zero intercept) with 90 percent confidence as indicated by the t-value. This will be the necessary threshold for concluding that the two measures of vehicles per residential unit are suitable proxies for each other.

If these criteria are met, the regression analysis of utilization will be repeated using the same methods, however using the DOL registration data or the ACS Vehicles Available data as the dependent variable representing parking utilization, thereby providing a significantly larger sample size with greater geographic coverage. In this regression analysis, however, some independent variables resulting from the field work (e.g., parking supply) will have to be excluded. The resulting goodness of fit, when compared to that from the regression of the utilization counts from the field data, will indicate if parking utilization can be estimated accurately without field work.

If neither the DOL registration data nor the ACS Vehicles Available data are deemed suitable proxies for the utilization counts data in and of themselves (as tested in the correlation and regression described above), the utilization counts from the field work data will then be regressed again, this time using the DOL registration data and the ACS Vehicles Available data as independent variables. In other words, if a strong correlation is found between the DOL registration data and the utilization data, but the regression coefficient is not consistent with one (making them unsuitable proxies), additional independent variables will be added to the regression to better estimate the dependent variable. While the occupancy counts from the field work will be used as the dependent variable here, this regression will assess if these data can be fit without the use of field work data as independent variables. Additional independent variables will be included along with the DOL registration data or the ACS Vehicles Available data, however excluding variables collected in the field work. This will result in a regression equation that will relate the DOL registration data, the ACS Vehicles Available data, and neighborhood data to collected utilization counts. If this regression provides a reliable fit, it will result in an equation and a model to estimate parking utilization without using the field work data as input variables, and thus be applicable to larger areas than just King County.

To assess the best possible fit for estimating parking utilization while concurrently assessing the need for field work in obtaining these results, this sequence of tests and regression analyses is structured to answer the following questions:

1. How well, as assessed through goodness of fit, can occupancy counts representing parking utilization be estimated through the use of independent variables in a regression analysis?
2. Can DOL registration data or ACS Vehicles Available data represent utilization counts data with 90% confidence?
3. If DOL registration data or the ACS Vehicles Available data are deemed an appropriate proxy for utilization counts, how well, as assessed through goodness of fit, can these data be estimated in a regression analysis *not* utilizing any data collected through field work?
4. If DOL registration data and the ACS Vehicles Available data are *not* deemed appropriate proxies for utilization counts, can the occupancy counts be accurately regressed, as assessed through goodness of fit, using DOL registration data and ACS Vehicles Available data in conjunction with other independent variables (excluding data as independent variables requiring field work)?

ESTIMATING PARKING UTILIZATION IMPACTS

Using best available research findings and accepted rule of thumb assumptions in the industry, impacts will be estimated on variables such as greenhouse gas emissions, vehicle miles traveled, mode split, development costs, and housing costs. For example, CNT's H+T Index provides values for average vehicle miles traveled (VMT) per auto. Using these data in conjunction with the estimated parking utilization, impacts on average VMT can be predicted. Also, if it is assumed that the average auto emits 0.0092 metric tons of carbon dioxide per gallon of gasoline consumed, and that the average auto gets 20.3 miles per gallon of gasoline, the impact of varying parking utilization levels on greenhouse gas emissions can be calculated.

These impacts will be based, in part, on the trends identified in the research and analysis. For example, because the price of parking will be used as an independent variable in estimating parking utilization, an assessment of the impacts of parking prices on these various factors will also be possible. Therefore, if the cost of parking is increased, the model will estimate the resulting impact on parking utilization, and these subsequent impacts can also be calculated.

DEVELOPING PARKING UTILIZATION WEBSITE

To leverage the investment in this research, a website will be developed to support and guide parking supply and management decisions. Using the function developed in the regression analysis of the collected occupancy counts data, a model will estimate parking utilization in all developable parcels[1] in King County. Because the independent variables resulting from the field work will not be available for every parcel, characteristics specific to the building and development will be held constant. In other words, if average unit size, average rent, and the floor to area ratio are determined to be the significant factors defining the housing characteristics, an average value will be calculated for each of these variables. These values will then be used as inputs defining these characteristics in the model. This prototypical development will serve two purposes: it will enable the model to be run throughout the county, even where field data were not collected; and it will enable parking utilization to be assessed separate from the influence of the building or development under consideration. Many of the factors influencing parking utilization cannot be altered by stakeholders or in the context of deciding how much parking is necessary. By modeling utilization for the prototypical building, as a function of these variables on the ground, decision makers will be able to better understand the impacts of the factors they can control.

These controlled model outputs will be displayed through a mapping tool on the website. In other words, visitors to the site will first see a map of King County presenting estimated parking utilization for a prototypical building on all developable parcels in the county. Users

[1] as defined in the Location Criteria of the Phase I methodology

will then have the ability to select a parcel, input details specific to their proposed development, and see the new estimated parking utilization. Users will also, therefore, be able to alter these characteristics and compare the impacts. For example, if the price for planned parking at a proposed building is reduced by half, how will that impact the utilization of parking? The website will enable users to assess these impacts, and therefore, help guide stakeholders' decisions regarding appropriate parking levels.

This approach presents early concepts for the website. As both the analysis and the website development process progress, the most appropriate applications and presentations of the data will be considered and determined.

ATTACHMENT A: INDEPENDENT VARIABLES TO BE TESTED

Characteristics and Variables	Data / Source(s)	Geographic Level / Aggregation Method
Parking Supply and Price		
average cost per space per month	field work	parcel
car sharing access	local sources	gravity measure from parcel centroid
average neighborhood parking cost per day	local sources	block group -> parcel buffer
supply (on-site)	field work	parcel
Property / Development Characteristics		
average rent	field work	parcel
percent affordable units	field work	parcel
dwelling units by land area (footprint density)	field work, assessors data	parcel
dwelling units per 1000 SF (building density)	field work, assessors data	parcel
floor area ratio	field work	parcel
total units (by type and/or size)	field work, assessors data	parcel
proportion developed land within development	PSRC or KC GIS	parcel
price per square foot	field work	parcel
Neighborhood Household Characteristics		
household income (median, average, and per capita)	ACS (2005-2009 5-year estimates)	block group -> parcel buffer
average household size	ACS (2005-2009 5-year estimates)	block group -> parcel buffer
average commuters per household	ACS (2005-2009 5-year estimates)	block group -> parcel buffer
presence of children (possibly % population under 18)	ACS (2005-2009 5-year estimates)	block group -> parcel buffer
average age (possibly % in ranges)	ACS (2005-2009 5-year estimates)	block group -> parcel buffer
average commute distance	LED data	block -> parcel buffer
average autos per household	ACS (2005-2009 5-year estimates)	block group -> parcel buffer
selected monthly owner costs and gross rent	ACS (2005-2009 5-year estimates)	block group -> parcel buffer
Accessibility		
access by land coverage scaled by frequency of service (CNT's TCI)	GTFS data collected by CNT	parcel buffer
accessible area in a given time by transit (total area, jobs, etc)	Network analysis of GTFS data	polygon from parcel centroid
distance to nearest transit stop	GTFS data	network distance from parcel centroid
distance to nearest freeway interchange	TIGER/Line	network distance from parcel centroid
transit stop density	GTFS data	parcel buffer
job gravity	LED data	gravity measure from parcel centroid
service job gravity (access to amenities)	LED data	gravity measure from parcel centroid
job density	LED data	parcel buffer
activity measure (residential + employment)	ACS and LED	parcel buffer
gravity activity measure (residential + employment)	ACS and LED	gravity measure from parcel centroid
job type mix measures (gravity entropy)	NAICS codes in LED data	gravity measure from parcel centroid
proximity to schools		
walkscore	walkscore	parcel buffer
Built Form / Development Patterns		
entropy / land use mix measures	land use data	parcel buffer
residential density / gross density	Census TIGER/Line files	block group -> parcel buffer
average block size / intersection density / block density	Census TIGER/Line files	parcel buffer
link/node ratio	local sources	parcel buffer