Appendix F

Geocoding of Respondents' Origin-Destination Trips

Appendix F
Geocoding of Respondents' Origin-DestinationTrips

Each of the 760 "with meters" surveys and 760 "without meters" surveys contain geographic data describing the origins and destinations of the respondents' trips. The surveys also contain the respondents' home zip codes. Cambridge Systematics geocoded these geographic data to examine the similarity of the "with meters" and "without meters" samples, and to analyze whether geography plays a role in determining opinions about ramp meters. The process of geocoding entails converting textual data describing locations to pairs of geographic coordinates that can be represented as points on a map. The objective of geocoding information about the ramp meter surveys is to assign a latitude, longitude coordinate pair to each respondent's home, trip origin, and trip destination, so that each of these locations can be used as points in a GIS (Geographic Information System). Each point in a GIS is a record in a database table that has a unique id associated with the particular survey respondent. The unique ids facilitate the joining of the rest of the survey data with the points.

Subsequently, the survey data can be examined and described geographically through maps. For example, each trip origin point could be colored according to its respondent's opinion of whether the ramp meter system should be modified or not. In addition, geocoding allows for spatial querying and aggregation of the data. For example, all of the trip origins could be aggregated to the county level to see if the average opinion of the ramp meters differs from county to county. Spatial aggregation in a GIS is a quick and efficient way to discern whether geographic patterns exist in the data.

GIS was also used in this study to create variables for the ANOVA analysis. First, the study area was divided into groups of zip codes. Then each respondent was allocated into one of these groups based on its home zip code point, trip origin point, and destination point. Categorical variables were created to store this general information about the respondent's location. These variables allowed the study of whether the nature of the respondent's trip affects their view of the ramp meters, or whether the respondent's home location affects or colors their experiences with the ramp meters along with the other variables in the ANOVA.

Data Preparation

The data originated in text file format. The geographic data were separated and information from each survey was stored in a series of lines. The first line states the case id, the unique id representing the respondent. The second line has the respondent's home zip code; and the next three lines contain information about the intersection where the respondent's trip started in the form of "street one," "street two." The final three lines contain information about the intersection where the respondent's trip ended in the form of "street one" and "street two."

```
Pre-Wave Case ID:  5001

Zip Code:  55427

Q3d. Intersection where started trip

Street1:  Plymouth Ave
Street2:  Napor Ave

Q3h. Intersection where ended trip

Street1:  94th Ave
Street2:  James Ave
```

A program was written in ArcView's Avenue language to parse the intersection and zip code data into a database format. The resulting database represents each survey's information in a record where the first column or key field contains the unique case id. The second column contains street 1 of intersection 1, the third column contains street 2 of intersection 1, the fourth column contains street 1 of intersection 2, and the fifth column contains street 2 of intersection 2. The database tables are stored in an ArcView GIS.

There were many obvious misspellings and even miswordings evident in the data, so each record in the database was coarsely edited manually to correct such obvious errors. After the initial data cleaning was completed, two new fields were created in the database, which concatenated the two streets in each intersection into one field and separated them with an "&." For example, if "Street1" were Plymouth Ave and "Street2" were Napor Ave, the new field would contain "Plymouth Ave & Napor Ave." This was done to format the data for automatic geocoding as described below.

In order to automatically geocode or pinpoint locations in a GIS environment, a detailed basemap must be used that contains address precision. The dataset used by Cambridge Systematics for this task is The Lawrence Group (TLG) Street Centerline Data. The dataset or layer was developed by The Lawrence Group for the express purpose of facilitating automated routing and address matching applications. This street data layer was brought into ArcView and spatially indexed by ArcView in preparation for geocoding the intersection data.

Geocoding results in layers of points that have pinpoint precision. However, the points will only be as accurate as the original data. The best and most precise geocoding is done with address data. However, since the survey collected intersection data, it was assumed that the intersections represent a precise point in space and that point is indeed one of their trip ends. The data were obtained through telephone surveyors asking the respondent to identify the trip end intersections. Since not all respondents are expected to know the exact intersection where they began and ended their trips, the respondents gave information that describes the general area they began or ended their trip. In the process of geocoding, the general intersection descriptions of varying accuracy were assigned a precise point in space.

Geocoding Methodology

ArcView's automated geocoding function is called, "Geocode Addresses." It requires the user to specify the table where the location data are stored and the field where the address to locate is. In this case, the database table created that stores the intersection data and the field we created with the streets of each intersection separated by the "&" sign were selected. ArcView's function goes through each record in the table and looks for a point on the basemap that matches the intersection. The user is allowed to loosen spelling sensitivity so that, if the street name is spelled wrong in the data, a match can still be made.

ArcView's geocoding function was only able to automatically match between 40 and 50 percent of each survey's origins and destinations (see Table 6C.1). ArcView allows interactive geocoding of the remaining unmatched surveys. For each, it allows the user to choose from a list of possible intersections. However, the user is not allowed to see the intersections on a map. Because of this, only another 22 to 30 percent more could be geocoded semi-automatically. The remaining unmatched intersections were geocoded by hand using a combination of paper and electronic maps. Each origin and destination were located on a map, and then a point was created in ArcView at that location. After this effort, there were still some intersections that could not be geocoded, because the respondent didn't know or wasn't sure where their trip started and/or ended. In the end, about 92 percent of all trip origins and destinations were geocoded.

Table F.1 Summary of Geocoding Effort and Resulting Database

Geocoding Results	Number of Surveys	Automatically Matched	Semi-Automatically Matched	Manually Matched	Unmatched	Total Match Percentage
"With meters" Survey Origins	760	343	205	169	43	94.3%
"With meters" Survey Destinations	760	380	171	136	73	90.4%
"Without meters" Survey Origins	760	366	174	150	70	90.8%
"Without meters" Survey Destinations	760	309	238	148	65	91.5%
Total						**91.7%**