Paper 3386-2015

Defining a Reasonable Distance for Consumer Access to Retail Locations

Sarah Campbell, Federal Deposit Insurance Corporation

ABSTRACT

We use a series of SAS® data steps and SQL joins to calculate reasonable distance thresholds useful for exploring consumer access to bank branches. We demonstrate the calculations using geocoded addresses from the FDIC Summary of Deposits data and Census geospatial data, including point shapefiles containing population-weighted centroids of census tracts.

We start with a PROC SQL step to employ a Cartesian join on bank branch location data with the dataset containing population-weighted centroid coordinates. Using the GEODIST function in SAS, we are able to calculate the distance to the nearest bank branch from the population-weighted centroid of each census tract. The tract dataset is then grouped **–** by Metropolitan Statistical Area (MSA), Metropolitan Division (MD), or non-metropolitan (non-MSA) area for each state **–** and sorted in ascending order within each grouping by distance to the nearest bank branch using the RETAIN function. We calculate the cumulative population and cumulative population percent for each MSA/MD; the reasonable threshold distance is established where the cumulative population is closest to 90%.

INTRODUCTION

It is useful to understand bank branch patterns as they relate to consumers in nearby communities. Determining the extent of market accessibility for a particular area requires defining a threshold value for what may be considered a "reasonable distance" for consumers to travel to nearby retail locations. Previous research methods on consumer proximity to markets often rely on a relatively uniform distance threshold value to define market accessibility. However, a more dynamic threshold value that reflects the variance in population distribution and commuting patterns across the country may yield more salient results in identifying communities within the potential service area of a retail location.

Figures 1a-c below exhibit results from a preliminary, geospatial analysis of the data on branch coordinates relative to coordinates of population-weighted centroids of census tracts. Roughly 90% of the US population is within 4 miles of the nearest full-service bank branch. A similar proportion (89.9%) of the population living in metropolitan areas is within 3 miles of the nearest branch.[1] However, the vast majority (90.1%) of the population in non-metropolitan areas is within 8 miles of the nearest branch location.

Figure 1. Proximity analysis results for (A) all areas; (B) MSAs/MDs only; and (C) non-MSAs only

A

Miles to Bank Branch	No. of Tracts	Cumulative No. of Tracts	Cumulative Tract Population (1000s)	Cumulative Tract Population (%)
.	125	.		.
1	39,569	39,569	163,684	52.4
2	16,968	56,537	240,603	77.0
3	5,642	62,179	266,397	85.3
4	3,029	65,208	279,666	**89.5**
5	2,156	67,364	288,643	92.4
6	1,636	69,000	295,167	94.5
7	1,298	70,298	300,219	96.1
8	967	71,265	303,838	97.2
9	666	71,931	306,245	98.0
10	519	72,450	308,075	98.6
>10	1,552	74,002	312,471	100.0
TOTAL	74,127*	.		.

B

Miles to Bank Branch	No. of Tracts	Cumulative No. of Tracts	Cumulative Tract Population (1000s)	Cumulative Tract Population (%)
.	0	0		0.0
1	35,485	35,485	147,279	56.2
2	14,928	50,413	215,594	82.3
3	4,219	54,632	235,453	**89.9**
4	1,952	56,584	244,488	93.4
5	1,293	57,877	250,272	95.6
6	886	58,763	254,109	97.0
7	609	59,372	256,644	98.0
8	401	59,773	258,255	98.6
9	296	60,069	259,488	99.1
10	196	60,265	260,261	99.4
>10	475	60,740	261,852	100.0
TOTAL	60,740	.		.

C

Miles to Bank Branch	No. of Tracts	Cumulative No. of Tracts	Cumulative Tract Population (1000s)	Cumulative Tract Population (%)
.	125	.		.
1	3,868	3,868	15,597	30.8
2	2,176	6,044	24,816	49.0
3	1,431	7,475	30,717	60.7
4	1,073	8,548	35,005	69.2
5	884	9,432	38,388	75.8
6	789	10,221	41,291	81.6
7	653	10,874	43,705	86.3
8	532	11,406	45,605	**90.1**
9	359	11,765	46,759	92.4
10	324	12,089	47,834	94.5
>10	964	13,053	50,619	100.0
TOTAL	13,178*	.		.

[1] Consistent with findings from the Federal Reserve's Survey of Consumer Finances in which the median distance between depository institutions and consumers was 3 miles, remaining constant from 1992 to 2004 (Brevoort and Wolken 2008).

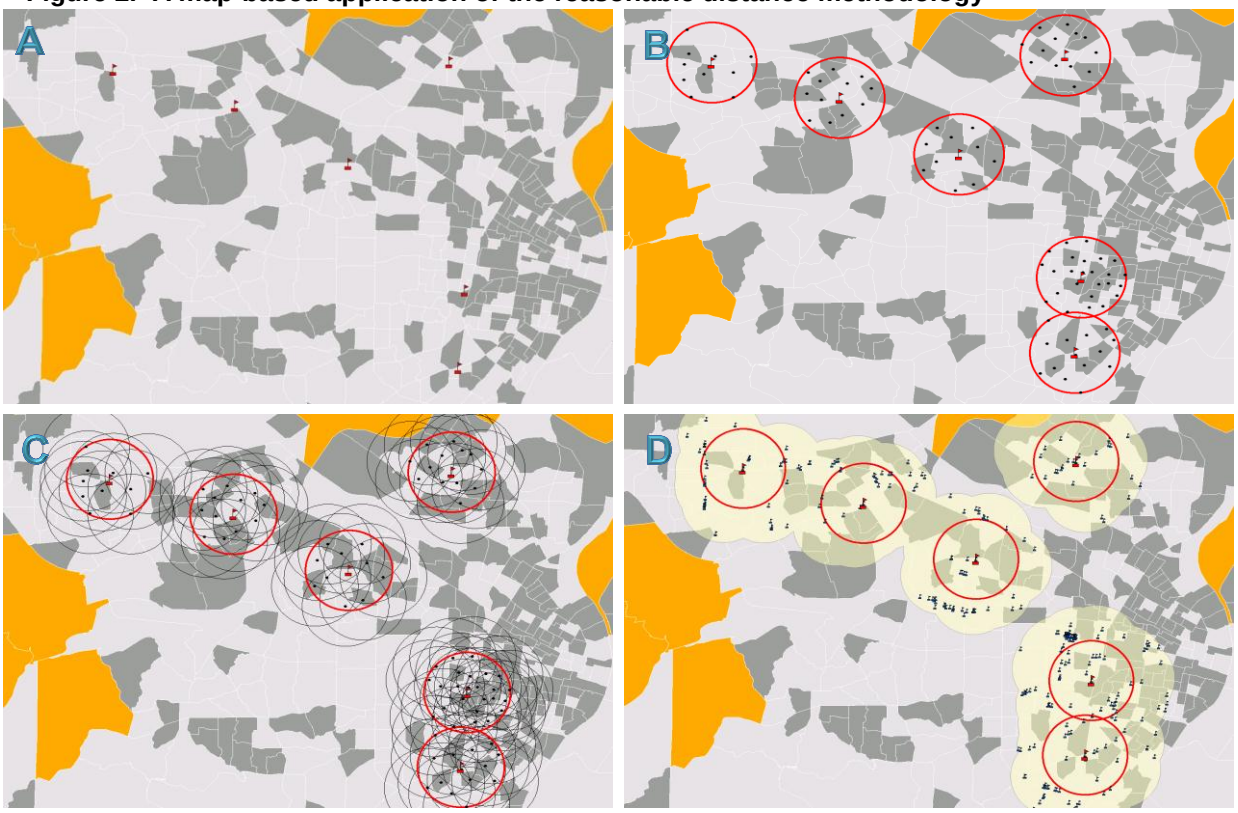
Further analysis of distance thresholds for unique metropolitan areas revealed that even these three distinct classes of geographic groupings may lack exactness. This preliminary analysis is motivation for calculating reasonable distance thresholds for unique MSAs, MDs, and non-MSAs by state.

FIND THE NEAREST BRANCH LOCATION

Based on findings from the preliminary analysis, it was determined that a "reasonable distance" for consumer market access would be quantified as the unique distance for a particular geography that, if traveled by 90 percent of the people in that area, would result in access to at least one branch location.

The four images below in Figure 2 illustrate a map-based application of the "reasonable distance" methodology. The images are enlarged maps of a community in the St. Louis, MO-IL metropolitan area where the reasonable distance threshold is 2.89 miles. The red buildings with flags represent the six bank branches of an institution headquartered in St. Louis, MO (2A). A buffer with a radius of 2.89 miles is drawn around each of the six banks. The small black dots are the population-weighted centroids of each of the nearby tracts (2B). For each of these tracts, we draw 2.89 mile buffers around the population-weighted centroids (2C). Lastly, the number of alternate nearby branches is the sum of all full service branches within 2.89 miles of the centroid of each affected tract (2D).

Figure 2. A map-based application of the reasonable distance methodology



The first step in calculating the "reasonable distance" threshold value for unique geographies is to locate the nearest bank branch from each tract's population-weighted centroid. A Cartesian join in a PROC SQL procedure is used to match tract and branch coordinates:

```
proc sql;
   create table work.cartjoin as
      select *,
      geodist(cntrd_lat, cntrd_long, br_lat, br_long, 'M') as d_cntrd_to_bb
   from work.tractfile, work.branchfile
      where abs(cntrd_lat-br_lat)<0.1 OR abs(cntrd_long-br_long)<0.1;
quit;
```

Once matched, the procedure then calculates the geodetic distance between paired coordinates using the GEODIST function. To reduce disk space and processing time, a WHERE statement is added to the Cartesian join so that the resulting rows are restricted to paired coordinates that are in proximity to one another. One-tenth of a decimal degree is roughly 70 miles, so the WHERE statement used in the above code limits the search radius for paired matches to 70 miles in each cardinal direction (NCGIA 1997).

A second two-step PROC SQL procedure is then used to isolate the nearest branch from the population center of each tract and join the distance to the nearest branch back to the tract dataset:

```
proc sql;
   create table work.neartable as
      select distinct geo_id as near_fips,
      min(d_cntrd_to_bb) as dist_nearest_bb
   from work.cartjoin group by near_fips;

proc sql;
   create table work.join_dist as
      select *
   from work.neartable, work.tractfile
      where near_fips = geo_id
         group by near_fips;
quit;
```

Lastly, a series of PROC SQL procedures are used to summarize population at the metropolitan area level and tract population as a percent of the MSA/MD population:

```
proc sql;
   create table work.msacounts as
      select distinct msamdcode as msamd_fips,
      count(*) as tractcount,
      sum(tract_pop) as sum_msa_pop
   from join_dist
      group by msamd_fips;

proc sql;
   create table work.jointotals as
      select *
   from work.msacounts, join_dist
      where codex = msamdcode;
quit;

proc sql;
   create table work.msapct as
      select distinct geo_id as tract_fips,
      tract_pop*100/sum_msa_pop as msa_pct
   from work.jointotals;
quit;

proc sql;
   create table work.totals as
      select *
   from work.msapct, work.jointotals
      where tract_fips=geo_id;
quit;
```

The population summary variables are then used in a DATA STEP to isolate the tract and respective distance to the nearest branch for that tract where the cumulative population percent in that MSA/MD is

closest (in either direction +/-) to 90. The distance to the nearest branch for that particular tract is the reasonable distance threshold value for that MSA/MD.

CALCULATE THE REASONABLE DISTANCE FOR UNIQUE GEOGRAPHIES

Once the distance to the nearest branch from each tract's population center is determined, and the MSA/MD summary statistics are calculated, the tract dataset is sorted by MSA/MD FIPS code and then by distance to the nearest branch in miles.

```
proc sort data = totals out=tractset;
   by msamd_fips dist_nearest_bb;
run;
```

The following DATA STEP uses the RETAIN statement to create two new variables that calculate the cumulative MSA/MD population and cumulative population percent for every tract in the dataset. Note that the WHERE statement limits the data to tracts where the civilian population is greater than zero. The final variable created, `dist90` is used in the subsequent data step to determine where the cumulative population for each tract within a particular MSA/MD grouping is closest to 90%.

```
data work.cbsa90;
   set work.tractset (keep = geo_id tract_pop sum_msa_pop msa_pct
                             msamd_fips msaname dist_nearest_bb);
      by msamd_fips dist_nearest_bb;
         where fl0_pop not in ('X');

      retain msa_cum_pop_pct msa_cum_pop;

         if first.msamd_fips then do;
            msa_cum_pop_pct = msa_pct;
            msa_cum_pop = tract_pop;
         end;
           else do;
           msa_cum_pop_pct = sum(msa_cum_pop_pct, msa_pct);
           msa_cum_pop = sum(msa_cum_pop, tract_pop);
        end;

        dist90 = abs(90-msa_cum_pop_pct);
run;
```

Figure 3. Isolating the marginal tract for the St. Louis, MO-IL metropolitan statistical area

msamd	msaname	dist_nearest_bb	geo_id	tract_pop	msa_pct	sum_msa_pop	msa_cum_pop	msa_cum_pop_pct	dist90
41180	ST. LOUIS, MO-IL	2.84648	17163502200	2101	0.07469	2812896	2508439	89.176	0.8236
41180	ST. LOUIS, MO-IL	2.85789	29099700401	4746	0.16872	2812896	2513185	89.345	0.6549
41180	ST. LOUIS, MO-IL	2.8583	29189221627	6513	0.23154	2812896	2519698	89.577	0.4234
41180	ST. LOUIS, MO-IL	2.87151	17119403802	4596	0.16339	2812896	2524294	89.74	0.26
41180	ST. LOUIS, MO-IL	2.88546	17117956500	2652	0.09428	2812896	2526946	89.834	0.1657
41180	ST. LOUIS, MO-IL	2.89477	17119402802	8158	0.29002	2812896	2535104	90.124	0.1243
41180	ST. LOUIS, MO-IL	2.92991	29099700303	4806	0.17086	2812896	2539910	90.295	0.2952
41180	ST. LOUIS, MO-IL	2.99376	29113810100	4269	0.15177	2812896	2544179	90.447	0.447
41180	ST. LOUIS, MO-IL	3.01782	29099700208	5818	0.20683	2812896	2549997	90.654	0.6538

As illustrated in Figure 3 above, a final DATA STEP returns the marginal tract that when included, yields a cumulative population percent that is closest to 90 percent of the total population for each metropolitan area and statewide non-metropolitan area. The reasonable distance threshold value, `thresh90`, for each MSA/MD is then equaled to the distance to the nearest bank branch for the marginal tract:

```
proc sql;
   create table work.mindist90 as
      select distinct msamd_fips, msaname, dist_nearest_bb as thresh90
   from work.cbsa90
      group by msamd_fips
         having dist90 = min(dist90);
quit;
```

CONCLUSION

This paper provides a simple solution executable in SAS to determine the extent of market access for consumers within a specified geography or groups of geographic entities. By locating the nearest retail location from each census tract, a reasonable distance threshold value can be calculated to better understand the feasibility of commuting within any given market area.

The solution is a starting point with room for improvement. For example, the GEODIST function in SAS uses the Vincenty formula to return geodetic distance between latitude and longitude coordinates. It does not account for topological features such as rivers or lakes; and it does not measure along roads and highways to convey actual drive time distance. See 'Recommended Reading' below for additional resources available for incorporating these features into the routine.

REFERENCES

Brevoort, K.P. and J.D. Wolken. 2008. "Does distance matter in banking?" *Working Paper 2008/34*, 9–10. Washington, DC: Board of Governors of Federal Reserve System.

National Center for Geographic Information and Analysis (NCGIA). 1997. "The NCGIA Core Curriculum in GIS Science: Unit 014 - Latitude and Longitude, Section 3.1.1. - Latitude and Distance" Accessed January 25, 2014. http://www.ncgia.ucsb.edu/education/curricula/giscc/units/u014/u014_f.html

ACKNOWLEDGMENTS

The author acknowledges the contribution of her colleagues at the Federal Deposit Insurance Corporation: Jane Coburn, Kris Rengert, and Lariece Brown for contributions to the methodology discussed in this paper; and Dhruv Sharma for his technical advice on query optimization.

RECOMMENDED READING

- ***Driving Distances and Times Using SAS® and Google Maps***
  http://support.sas.com/resources/papers/proceedings10/050-2010.pdf
- ***Batch Production of Driving Distances and Times Using SAS® and Web Map APIs***
  http://support.sas.com/resources/papers/proceedings12/091-2012.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Sarah Campbell
Federal Deposit Insurance Corporation
Email: sarcampbell@fdic.gov