ALTERNATIVE FINANCIAL SERVICE PROVIDERS AND THE SPATIAL VOID HYPOTHESIS

Tony E. Smith[*,a], Marvin M. Smith[b] and John Wackes[b]

[a]*Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA.*
[b]*Federal Reserve Bank of Philadelphia, Philadelphia, PA.*

September 10, 2007

Abstract

This paper studies the spatial relationship between traditional banking services (Banks) and alternative financial service providers (AFSPs). The main objective is to test the so-called *spatial void hypothesis* that AFSPs tend to locate in markets where traditional banking services are under-provided. The key question of interest here is whether or not AFSPs serve markets with significantly lower income levels than those of Banks. A statistical methodology is developed for addressing this question that builds on previous studies. The present approach is based largely on *K-function* analyses of both individual and paired point patterns. These Monte Carlo testing procedures are applied to Banks and AFSPs in the four-county region around Philadelphia, Pennsylvania, and are also compared with previous methods. The main finding of this research is that there is indeed strong empirical support for the spatial void hypothesis in this Philadelphia region.

Key Words: Spatial Void Hypothesis; Point Pattern Analysis; K-functions.

---

[*]   Corresponding Author. Tel. : +1-215-898-9647, Fax: +1-215-573-2068,
    E-mail Address: tesmith@seas.upenn.edu

INTRODUCTION

The majority of Americans rely on banks and other traditional financial institutions to conduct their financial transactions. However, a large segment of the population, an estimated 56 million consumers, have no affiliation with these mainstream institutions and use instead alternative financial service providers (AFSPs)—check cashing outlets, payday lenders, pawnshops, rent-to-own stores, auto title lenders, and money transmitters—for their financial needs.[1] Research has shown that many of the patrons of AFSPs are minority and low-income consumers.[2] While this fast growing non-bank segment of the financial industry seems to provide its customers with the services that they need, the price for the services are high relative to comparable services offered at many mainstream financial institutions. On the one hand, these high fees are thought to strip AFSPs' patrons of needed income to pay bills and possibly serve as the basis for asset or wealth accumulation. On the other hand, the AFSPs might fill a financial void due to the lack of mainstream financial institutions being present in areas where AFSPs' patrons are located. This latter hypothesis, which shall be designated as the *spatial void hypothesis* throughout this study, was most recently studied by the Fannie Mae Foundation (FMF).[3] The FMF study investigated the spatial void hypothesis by focusing on several sites around the country.

This study addresses essentially the same questions as the FMF study, but focuses instead on several counties in the Commonwealth of Pennsylvania. It examines whether AFSPs satisfy a financial vacuum created by the absence of traditional financial institutions and also explores whether AFSPs are disproportionately serving minority and low-income areas. In addition to applying techniques used in the FMF analysis, this study will employ certain alternative approaches using spatial statistics.

---

[1] General Accounting Office, "Electronic Transfers," Report to the Subcommittee on Oversight and Investigations, House of Representatives, September, 2002.
[2] See, Michael A. Stegman, "Banking the Unbanked: Untapped Market Opportunities for North Carolina's Financial Institutions," *Journal of the University of North Carolina School of Law*, Chapel Hill, NC, 2001.
[3] See Noah Sawyer and Kenneth Temkin, *Analysis of Alternative Financial Service Providers*, Washington, D.C.: The Fannie Mae Foundation, 2004.

BACKGROUND

Alternative financial service providers (AFSPs) have been around in various forms for sometime. According to one account, in 1986 there were about 2,200 alternative providers nationwide, but that number grew to more than 10,000 by 1994.[4] The total number of AFSPs has greatly increased since then, given that the number of pawn shops alone has been estimated to be between 12,000 and 14,000; while the number of payday lending outlets has been put at 22,000 in 2004.[5] Just in Pennsylvania, the number of licensed check cashers and pawnbrokers together grew by nearly 40 percent from August 2004 to May 2005. However, the proliferation of AFSPs in recent years has drawn a great deal of scrutiny (particularly payday lenders and check cashers) as a result of the heavy reliance on their services by minority and low-income households.[6] In many instances, this dependence has been to the exclusion of using the financial services of mainstream institutions, which offer a better opportunity to build wealth. Since the fees charged by AFSPs are typically higher than similar services available at traditional financial institutions, there continues to be some speculation regarding the popularity of these alternatives.[7] The focus has mostly been on (i) the number and kinds of AFSP establishments and their location, (ii) whether they provide needed financial services not readily available due to the absence of mainstream financial institutions, or (iii) whether they offer more convenient hours and a host of non-financial services in addition to basic financial needs. The FMF study dealt with the first two of the aforementioned areas of interest.[8] While this study will formally address the first two, it will also shed some light on the third.

---

[4] See John P. Caskey, *Fringe Banking: Check Cashing Outlets, Pawnshops, and the Poor*, New York: Russell Sage Foundation, 1994.

[5] See James H. Carr and Jenny Schuetz, *Financial Services in Distressed Communities: Framing the Issue, Finding Solutions*, Washington, D.C.: The Fannie Mae Foundation, 2001 and Mark Flannery and Katherine Samolyk, "Payday Lending: Do the Costs Justify the Price?" FDIC Center for Financial Research Working Paper, No. 2005-09, 2005.

[6] Estimates indicate that the check-cashing industry handles over 180 million checks annually valued at over 60 billion. See Gerald Goldman and James R. Wells, Jr., *Check Cashers Are Good Bank Customers*, Financial Services Centers of America, Inc., 2002.

[7] The Fannie Mae study reported that the fees can range from 15 to 17 percent for a two-week loan, while the annual percentage rate can reach 300 percent.

[8] More specifically, Fannie Mae investigated the characteristics of the neighborhoods where AFSPs are located, and whether the heavy patronage of AFSPs was due to the absence of conventional financial

Substitutes or Complements

In economics parlance, to the extent that check cashers and other AFSPs fill a void due to the absence of mainstream financial institutions, they serve as substitutes. However, to the degree that AFSPs function profitability in areas served by conventional financial institutions, they might be regarded as complements. The interplay of supply and demand forces for financial and related services helps underscore such a dichotomy. Logic would dictate that AFSPs would locate near those consumers who would most likely patronize their businesses. From an economic perspective, such behavior would be considered "rational." This is especially the case in areas with no mainstream financial institutions. Similarly, for those consumers with severely flawed credit or those with no relationship with a traditional financial institution and who frequent AFSPs (notwithstanding their high credit terms and the presence of a mainstream institution) might also be thought to be acting pragmatically.

It is understandable why the aforementioned consumers might use some AFSPs for their credit needs, but it is not so clear why consumers who have accounts at traditional financial institutions would use the services of AFSPs. While payday lenders require that borrowers have a checking account in order to receive a short-term loan, potential borrowers ostensibly have access to their mainstream financial institution for such loan needs. One study by Graves (2003) suggests that the exiting of conventional financial institutions from low-income neighborhoods might have created the impetus for some consumers with checking accounts to seek loans from payday lenders. An added explanation offered by Flannery and Samolyk (2005) is that low profit margins realized by banks and other financial institutions on small loans compelled them to forgo such loans, thus leaving an opening for payday lenders. Yet a third possible reason for the popularity of AFSPs in general is that they provide customers with more convenient hours and other non-financial services and products such as postage stamps, train and bus fares, notary services, lottery tickets, payment of utility bills, and pre-paid telephone cards.

institutions to provide the needed services. In addition, the Fannie Mae study considered the influence that laws enacted by local jurisdictions and states had on the locations of AFSPs.

While the pursuit of a comprehensive rationale why AFSPs might be complements to traditional financial institutions is beyond the scope this study, an investigation of the spatial void hypothesis as it relates to selected counties in the Commonwealth of Pennsylvania will be undertaken. [9]

METHODOLGY

This study focuses on the U.S. Census block-groups of four Pennsylvania counties – Philadelphia Allegheny, Delaware, and Montgomery – and examines the relative location patterns of Banks and AFSPs in these counties.[10,11] The analysis proceeds in several stages. First, following the FMF study, the spatial clustering of AFSPs and Banks are analyzed separately. As observed in the FMF study "the use of clusters provides a more accurate picture of the geographic distribution of the marketplace served by traditional and alternative providers."[12] In addition to these separate examinations, a second stage of analysis is carried out in which these two types of establishments are analyzed together. Here the attention focuses more directly on the relative market areas of AFSPs and Banks. First we consider markets from the demand side, and focus on the relative access of AFSPs and Banks to the spatial distribution of incomes (as characterized by median incomes at the block-group level). The key question of interest in this regard is whether or not AFSPs are serving markets with significantly lower income levels than those of Banks. Next, we consider markets from the supply side, and ask whether the residents of each given neighborhood (block-group) have significantly greater access to AFSPs than would be expected if AFSPs and Banks were indistinguishable. This same analysis is then applied to Banks as well.

---

[9] Although this study concentrates on the location of AFSPs vis-à-vis Banks or bank branches (used interchangeably), it doesn't analyze the factors that determine their location decisions. For studies that deal with the factors that underscore the location choices of payday lenders, see Mark L. Burkey and Scott P. Simkins, "Factors Affecting the Location of Payday Lending and Traditional Banking Services in North Carolina," *Review of Regional Studies*, 34(2): 191-205, 2004 and Steven M. Graves, "Landscapes of Prediction, Landscapes of Neglect: A Location Analysis of Payday Lenders and Banks," *The Professional Geographer*, 55(3):303-317, 2003.

[10] While all of the counties in Pennsylvania had bank branches, not all had AFSPs. Of those counties that did have alternative providers, only four had five or more AFSPs clustered together as well as bank branch clusters. Since the analysis relies, in part, on the clustering of bank branches and AFSPs, only the four counties that contained the requisite clustering were chosen for study. Banks and bank branches are synonymous in this paper.

[11] In this analysis, AFSPs are comprised of only check cashers and pawnbrokers.

[12] FMF study, p.7.

An important feature of the study is the comparison of the FMF's approach to determining AFSP and Bank clusters and alternative derivations employing spatial statistics.

DATA

The data for this analysis are drawn from three sources. The street addresses of the AFSPs were obtained from Pennsylvania's Department of Banking state licensing data as of May 2005. The Department of Banking's address location file provided information for the two AFSPs of interest in this analysis—check cashers and pawnbrokers. The data reflects Pennsylvania's classification of these two types of alternative providers. According to the Pennsylvania Check Casher Licensing Act of 1998, a check casher is defined as "a business entity, whether operating as a proprietorship, partnership, association, limited liability company, or corporation, that engages in the cashing of checks for a fee."[13] Similarly, the Pawnbroker's Licensing Act of 1937 stipulates that a pawnbroker "includes any person, who—(1) engages in the business of lending money on the deposit or pledge of personal property, other than choses in action, securities, or written evidences of indebtedness; or, (2) purchases personal property with an expressed or implied agreement or understanding to sell it back at a subsequent time at a stipulated price; or, (3) lends money upon goods, wares or merchandise pledged, stored or deposited as collateral security."[14]

In this study, check cashers and pawnbrokers are considered together under the rubric of AFSPs. As such, alternative providers with both check cashing and pawnbroker licenses represented only one outlet, in order to avoid double counting.

The analysis of the demographic make-up of block-groups where AFSPs and Banks are located is performed using selected variables from the 2000 Census, such as family income, percent white, percent black, and percent Hispanic. Each of these demographic characteristics is used to gauge its proportion of the census block-groups of AFSPs and Banks relative to its proportion of the counties in which the AFSPs and Banks are located.

---

[13] *Pennsylvania Bulletin*, Volume 28, Number 18, Saturday, May 2, 1998, Harrisburg, PA.
[14] Pawnbroker's Licensing Act of 1937, P.L. 200, No. 51.

Finally, the addresses of the Banks/bank branches were obtained from the Federal Deposit Insurance Corporation (FDIC) database. The FDIC database contains the locations of all the FDIC insured full-service bank branches in the Commonwealth of Pennsylvania.[15]

RESULTS

The four Pennsylvania counties in this study had a total of 333 AFSPs, with nearly 70 percent of them in Philadelphia County (see Table 1). Alternatively, all four counties had 1,339 Banks, with all but one county having three hundred twenty or more. Even though the counties have roughly four times as many Banks as AFSPs collectively, an examination of the spatial void hypothesis involves more than their relative numbers. The approach suggested here looks beyond simply whether a neighborhood (or block-group) contains a Bank in addition to an AFSP and also accounts for population density.

TABLE 1. NUMBER OF AFSPs AND BANKS

COUNTY	NUMBER OF AFSPs	NUMBER OF BANKS
Philadelphia County	230	321
Montgomery County	22	372
Delaware County	35	184
Allegheny County	46	462

Individual Cluster Analyses of AFSPs and Banks

In performing the separate cluster analyses of AFSPs and Banks, the present approach departs from that of the FMF study in several ways. First, rather than using nearest-neighbor methods to identify potential cluster members, we employ a version of

---

[15] The FDIC database does not include ATM only locations or credit unions, which are insured by the National Credit Union Administration. Credit unions are excluded from this analysis since, unlike FDIC insured depository institutions, they are not covered by the Community Reinvestment Act and thus are not required to meet the credit needs of the local communities they serve.

Ripley's[16] *K-function* approach that allows a systematic examination of clustering at alternative spatial scales. A second and even more important departure is the introduction of population density as benchmark against which to measure significant clustering. In the FMF approach, clustering is defined with respect to the simple hypothesis of "complete spatial randomness". As will become clear below, the "clusters" identified by this method are often little more than a reflection of population clustering. Hence the fundamental premise of the alternative approach suggested here is that clusters should be identified as "significant" only when they contain more establishments than would be expected on the basis of population alone.

Nearest Neighbor Hierarchical Clustering Procedure

To carry out this analysis, the locations of AFSPs and Banks were mapped using the geocoding procedure in the ArcMap software.[17]   The point locations of AFSPs and Banks are shown for Philadelphia County in Figures 1a and 1b below. This Philadelphia example will be used for illustrative purposes throughout the following discussion. [A discussion of the other counties will follow.] The red ellipses in these figures denote the clusters obtained following the same procedure as in the FMF study. These results were obtained using the *Nearest Neighbor Hierarchical Clustering* (NNHC) procedure component of Crime Stat 2.0 software [Levine (2002)], as detailed in the manual for this software.[18] For our present purposes, it is enough to say that this procedure essentially postulates that locations are realized as a completely random point pattern within the Philadelphia boundary shown. Points are then grouped into candidate clusters by first identifying point pairs that are closer together than would be expected under this randomness hypothesis. Finally, collections of linked pairs are grouped together as "first-order clusters". In this procedure there are two key parameters to be set. The first is the *p-value threshold*, p, used to identify those pairs that are "significantly" close together, and the second is the *cluster-size threshold*, *m*, used to define the minimum size of an

---

[16] See B.D Ripley, "The Second-order Analysis of Stationary Point Patterns," *Journal of Applied Probability*, 13:255-266, 1976.

[17] The geocoding was facilitated by using the StreetMap extension, which is an add-on program that assigns address locations in the mapping software.

[18] This manual is obtainable on line at *http://www.mappingcrime.org/crimestat.htm*. A full description of the Nearest Neighbor Hierarchical procedure used here starts on page 216 of Chapter 3.

"admissible" cluster. As in the FMF study, these values were here chosen to be $p = .01$ and $m = 5$. Each of these clusters is represented in terms of its associated "dispersion ellipse" shown by the red ellipses in Figures 1a and 1b below.[19]
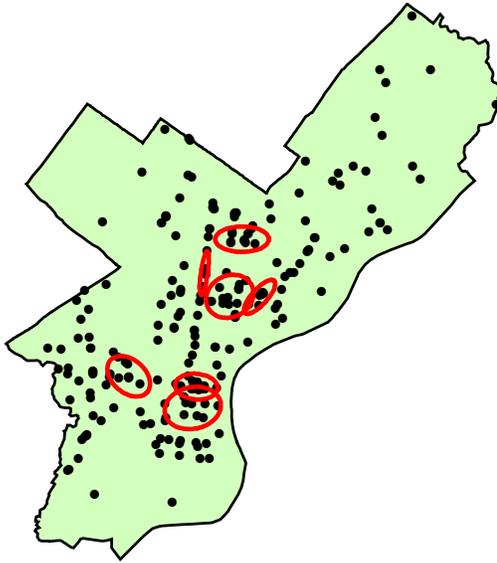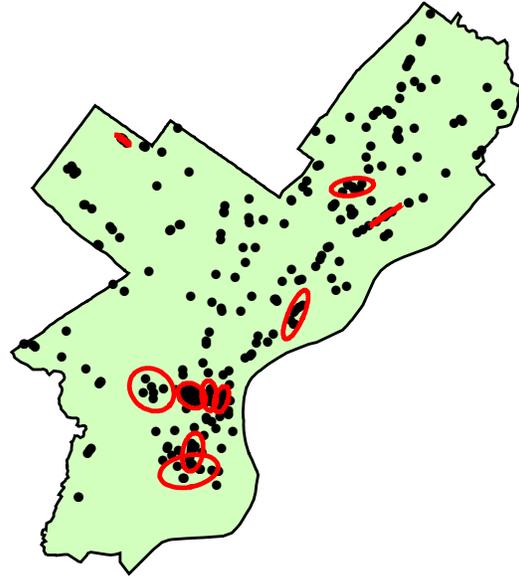

Figure 1a. Philadelphia AFSPs   **Figure 1b.** Philadelphia Bank Branches

While these regions do appear to correspond to the regions of highest concentration in each pattern,[20] it is difficult to evaluate their true significance without further information. For example, the three small clusters in the highest area of concentration in Figure 1b correspond roughly to Center City Philadelphia (see also Figure 4a below). This underscores the major limitation of this procedure, namely that concentrations of commercial services such as Banks are to be expected in areas of high population density. This can be seen more clearly in Figure 3a and 4a below, which involve enlargements of Figures 1a and 1b, respectively. Here the green background now represents population

---

[19] These dispersion ellipses are calculated by first rotating axes around the centroid of the point cluster, and then setting the major axis of the ellipse in the direction of maximum dispersion (standard deviation) of the y-axis. The size of this ellipse (in terms of standard deviations) is also a choice parameter in this graphical output. To ensure that each cluster corresponds closely to the points inside the ellipse, this parameter was set to its minimum value of one standard deviation.

[20] Here it should be noted that many establishments are so close together that they do not register as separate dots in these figures; so simple visual inspections of the patterns are necessarily limited by this fact.

densities in each of the block-groups shown on the map.[21] For example, the clusters of both AFSPs and Banks in West Philadelphia (just to the left of Center City) are actually in an area of unusually high population density (which includes both the University of Pennsylvania and Drexel University). Hence the concentration of financial institutions here is again not very surprising. So the key question is whether these concentrations are actually higher than would be expected given the local population.

K-Function Procedure

To control for these effects, a second procedure is now developed which takes the underlying populations in each block-group as a reference measure. In contrast to the above hypothesis of complete spatial randomness in which the likelihoods of point locations are taken to be proportional to area of each block-group, the present approach postulates that these likelihoods are proportional to population.[22] To motivate this *K-function* approach, consider a given point pattern, $X^0 = \{x_i^0 : i = 1,..,n\}$, consisting of n point locations (such as the AFSP locations in Figure 1a). For any point, x_i^0, let $K_i^0(d)$ denote the number of other points in X^0 within distance d of x_i^0. For example, the point x_i^0 in Figure 2 below corresponds to one of the AFSPs near the western edge of Philadelphia County. Here d equals one half mile, and the figure shows that $K_i^0(d) = 2$.[23] To test whether this observed count is "unusually large" given the local population size (represented by the color intensities of each block-group in the figure), one can employ Monte Carlo methods to estimate the sampling distributions of such point counts. To do so, one need only simulate a large number of replicate point patterns,

---

[21] The block-groups left blank were excluded from the analysis for reasons including zero population levels.

[22] Here it should be noted that a similar "risk-adjusted" version of NNHC procedure is available in Crime Stat 2.0 (starting on p.235 of Chapter 3). However, this procedure is considerably more complex than the present approach and is more difficult to interpret. For example, rather than sample directly from such a reference measure to test the "risk adjusted" hypothesis of randomness, this measure is used to construct local asymptotic normal approximations to point densities in each point neighborhood. While such approximations may be reasonable for large numbers of points, they are more questionable given the small sample sizes in such local neighborhoods.

[23] Technically, the function of d defined by $K_i^0(d)$ is the sample estimate of a *local K-function*. As with most of the spatial statistics used here, this is an instance of the general class of LISA statistics proposed by L. Anselin, "Local Indicators of Spatial Association: LISA," *Geographical Analysis*, 27: 93-116, 1995.

$X^{(s)} = \{x_i^{(s)} : i = 1,.., n-1\}, s = 1,.., N$, of size, $n-1$, from a probability distribution proportional to population.[24] Each pattern $X^{(s)}$ then constitutes a possible set of locations for all points other than x_i^0, under the null hypothesis that location probabilities are
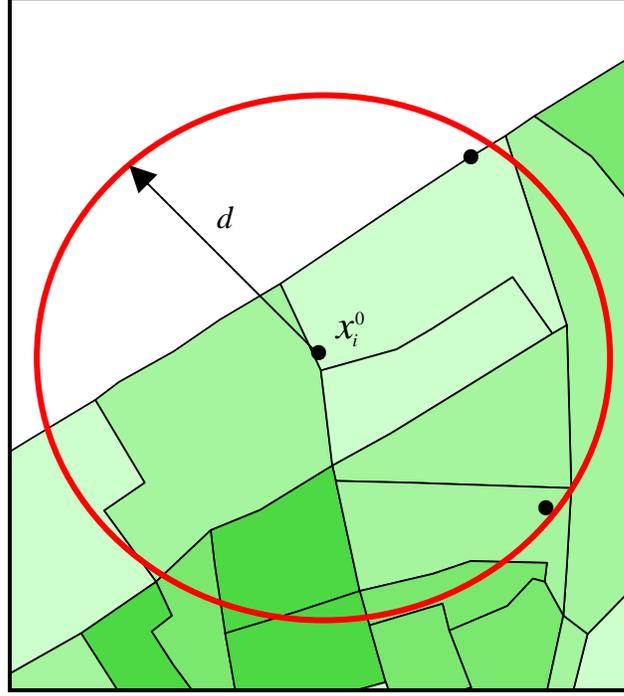

Figure 2. Point Counts for Patterns

proportional to population. If $K_i^{(s)}(d)$ denotes the number of points in each pattern, $X^{(s)}$, within distance d of x_i^0, and if $K_i^0(d)$ were simply another sample from this distribution, then each possible ranking of this particular value in the list of values, $\{K_i^{(0)}(d), K_i^{(1)}(d), .., K_i^{(N)}(d)\}$, should be equally likely. Hence if $M_i(d)$ denotes the number of these $N+1$ values that are at least as large as $K_i^0(d)$, then the ratio

---

[24] Technically this amounts to a two stage sampling procedure. If p_i denotes the fraction of population in block group i, then a random point is first assigned to block group i with probability p_i. This point is then assigned to a random location inside this block group (by standard rejection sampling methods). Hence the actual distribution used here is a step-function approximation to population density. A more elaborate procedure could be constructed by employing kernel-smoothing methods to estimate this density over a much finer grid.

(1) $\qquad P_i(d) = \dfrac{M_i(d)}{N+1}$

yields a (maximum likelihood) estimate of the probability of observing a count as large as $K_i^0(d)$ in a sample of size $N+1$ from this hypothesized distribution. By construction, $P_i(d)$ is thus the *p-value* for a (one-sided) test of this hypothesis. Here the value $N = 1000$ was used for all simulations in this paper.[25] So for example if $M_i(d) = 10$, then $P_i(d) = 10/1001 < .01$ would imply that the estimated chance of observing a value as small as $K_i^0(d)$ is less than one in a hundred. Hence in this case $K_i^0(d)$ might indeed be regarded as "unusually large" after accounting for population.[26]
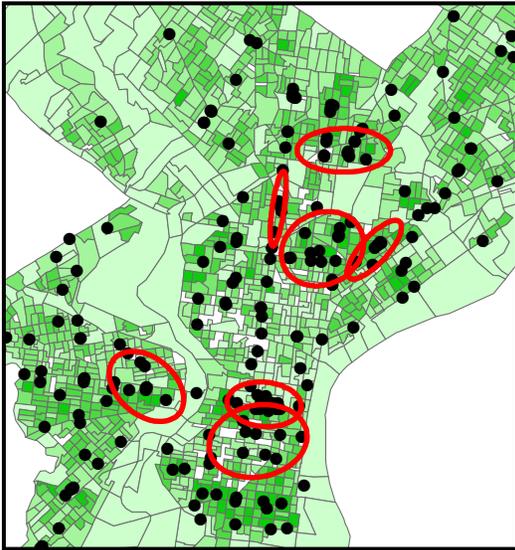
Figure 3a. Philadelphia AFSPs (NNHC)   **Figure 3b.** Philadelphia AFSPs (K-Function)
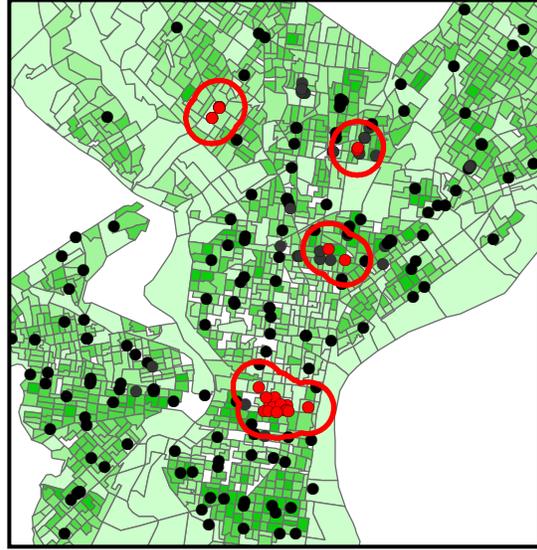
These p-values can be mapped, and provide a clear visual picture of where counts are unusually high. To compare the present results with those of the NNHC procedure above,

---

[25] This simulation procedure and all others discussed in this paper were programmed in MATLAB by one of the authors.

[26] Here it should be noted that one additional limitation of both the NNHC procedure and its risk-adjusted version in footnote 18 above is overcome by the present approach. This limitation is well illustrated by the example in Figure 2 above, where x_i^0 is so close to the boundary that the full circle of radius d is not contained in the Philadelphia region. Such "edge effects" are ignored in both NNHC procedures, and can in principle lead to a serious bias in the identification of nearest-neighbors close to the boundary. However, since all simulated point pattern counts in the present Monte Carlo approach are subject to the *same* edge effects, the p-values obtained are not biased by this effect.

a single *p-value threshold* of $p = .01$ has again been selected. For the Philadelphia example with d equal to one half mile, those locations with p-values not exceeding .01 are depicted by red dots in Figures 3b and 4b. In addition, we have again used a *cluster-size threshold* of $m = 5$, so that meaningful "clusters" are taken to include at least five points. Hence each red point shown also has at least four other points within distance d of its location. The corresponding red boundaries shown outline the unions of d-radius circles around each of these red dots.[27] With this definition, the points inside these boundaries are exactly those included in the K-counts of at least one of the red points. These groupings are then taken to constitute the natural clusters defined by these two threshold criteria.
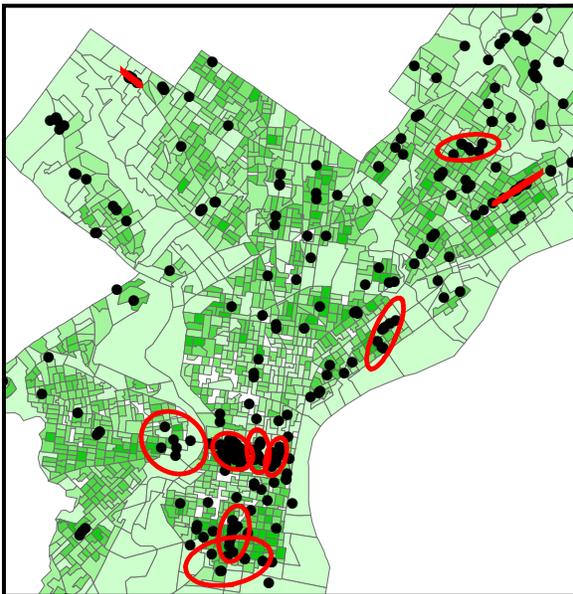
Figure 4a. Philadelphia Banks (NNHC) **Figure 4b.** Philadelphia Banks (K-Function)
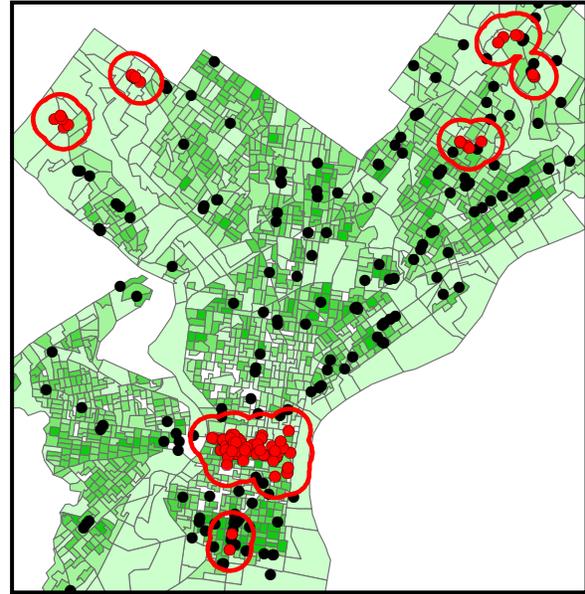
Before comparing these two procedures, it is important to emphasize again that the results shown are specific to certain choices of parameter values. In particular, they depend on the *threshold p-value*, *p*, the *threshold cluster-size*, *m*, as well as the *radial-distance threshold*, *d*, in the K-function approach. Hence it should be clear that a range of other values can be considered. This has in fact been done, and it can be reported that the values of $p = .01$ and $m = 5$ used in the FMF study turn out to be quite appropriate for the four counties in the present study as well. For example, choices of cluster sizes

---

[27] These unions were obtained by using the "buffer" option available in ARCMAP.

smaller than 5 tend to yield many small isolated clusters. Moreover, values larger than 5 tend to eliminate certain groupings that just meet this threshold. As one example, the left-most Bank cluster in Figure 4b happens to consist of exactly 5 Banks. This cluster is of particular interest since it is *not* significant under the NNHC procedure (Figure 4a). Thus the significance of this cluster in the present approach is largely due to its location in a relatively sparsely populated area. (We shall return to this point later).

Turning next to the p-value threshold, $p = .01$, observations can be made similar to those for the cluster-size threshold. But there is one additional consideration that should be mentioned. This local definition of p-values at each point location has been criticized as involving "multiple testing" that can in principle lead to the identification of "too many" clusters. But, as pointed out by Rushton and Lolonis (1996), [28] who were among the first to use this method, each such p-value is only valid as a *local* test of the above null hypothesis. The fact that such tests are necessarily correlated (in terms of overlapping circles) means only that this significance level, *p*, should *not* be applied to the cluster as a whole. As with the NNHC method above, these correlated groupings serve mainly to identify areas of potentially significant clustering.[29]

Turning finally to the radial-distance threshold, *d*, it should be clear from the very name "K-function" that this approach is used primarily to examine changes in significance levels over a range of *d*-values. However, for purposes of comparison with NNHC, it is most convenient to choose a single value for presentation. Here the value of *d* chosen depends heavily on the density of population in a given area.[30] In the case of Philadelphia, the value of one half mile for *d* appears to yield the best compromise between many small isolated clusters (for small *d*) and a few large diffuse clusters (for large *d*). However, for the lower population densities in Allegheny, Delaware, and

---

[28] See G. Ruston and P. Lolonis, "Exploratory Spatial Analysis of Birth Defects in an Urban Population," *Statistics in Medicine*, 15: 717-726, 1996.

[29] It should also be mentioned that a number of methods have been proposed for treating this problem, including the sequential testing procedure of J. Besag and J. Newell, "The Detection of Clusters in Rare Diseases," *Journal of the Royal Statistical Society, A,* 154: 327-333, 1991 and the SaTScan procedure of M. Kulldorff, "A Spatial Scan Statistic," *Communications in Statistics: Theory and Methods*, 26: 1487-1496, 1997. Here it suffices to say that the results produced by such procedures are in fact very similar to those obtained here – and again depend most critically on the choice of the threshold p-value that is chosen. An excellent discussion of this issue can be found in M.C. Castro and B.H. Singer, "Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Tests in Local Statistics of Spatial Association," *Geographical Analysis*, 38: 180-208, 2006.

[30] Ideally such distances should reflect some meaningful "market size" for each financial institution. But in the absence of such data, the present choices are again based largely on more subjective considerations.

Montgomery counties (discussed below), a larger radial-distance threshold of one mile for *d* appeared to yield more meaningful clustering results.

Comparisons of Results for the Two Clustering Methods

Given these general observations, we next compare the major differences in results between these two methods. Overall, the K-function approach generally yields less Bank clusters than the NNHC method as well as less AFSP clusters in Philadelphia County (see Table 2). But this result masks other worthwhile findings. Turning first to a comparison of AFSP clusters in Figures 3a and 3b, it is of interest to note that the cluster around the area containing the University of Pennsylvania and Drexel University in West Philadelphia observed in the NNHC method (Figure 3a) is absent in Figure 3b. Once the high population density in the West Philadelphia area has been accounted for, the concentration of AFSPs in this area is no longer significant. This is also true of a number of other clusters, such as the South Philadelphia cluster in Figure 3a. Again this urban area is densely populated. A similar pattern can be observed for Banks in Figures 4a and 4b, where the West Philadelphia cluster is again absent in Figure 4b. However, in South Philadelphia the concentration of Banks seen in Figure 4a continues to be significant in Figure 4b, indicating that there are even more Banks in this area than would be expected on the basis of population alone.[31]

TABLE 2. COMPARISON OF CLUSTERS BETWEEN NNHC AND K-FUNCTION

	NUMBER OF ASFPs CLUSTERS		NUMBER OF BANK CLUSTERS	
	NNHC	K-FUNCTION	NNHC	K-FUNCTION
Philadelphia County	7	4	10	7
Delaware County	1	1	20	11
Allegheny County	1	1	11	4
Montgomery County	1	1	30	12

---

[31] For a full set of the comparison of the two clustering methods for the four counties, see Appendix.

Next observe that a number of new clusters have appeared in the K-function approach, such as the small cluster of Banks in the upper left of Figure 4b (mentioned above). The significance of this cluster is due largely to the relatively low population density in this area of Philadelphia. In other words, one would not expect to find this concentration of Banks based on population alone. Rather, the cluster identified appears to be more readily explainable in terms of the high level of median incomes in this area – which is about $66,000 versus an approximate median income of $30,000 for all of Philadelphia (as discussed in more detail below). This is also true of the two new clusters in the Northeast Philadelphia area of Figure 4b. With respect to AFSPs, there is also one new cluster in Northwest Philadelphia in Figure 3b. Here again, the significance of this cluster is due to the low population density. But the median income ($26,000) in this area is actually lower than that for all of Philadelphia ($30,000). Hence, as observed in the FMF study, this suggests that such clustering of AFSPs may be directed more toward low-income customers.

Finally the case of Center City is of interest for different reasons. Here we see that there is a very significant clustering of both Banks and AFSPs – even after population is accounted for. But further reflection suggests that the relevant population densities in such centers of business activity may not be well captured by residential population levels. For while there are indeed many high-rise apartment buildings in Center City, the actual day-time population levels are more accurately reflected by employment data than residential data. Hence even by controlling for residential density, the actual population density in such areas during business hours may be severely underestimated. This highlights one clear limitation of the present K-function approach. What is needed here is a more accurate reference measure of relevant *consumer* population densities. Hence the question of whether the strong clustering observed in Center City is actually significant remains open for further study.

Comparative Market Analyses of AFSPs versus Banks

We now turn to a consideration of the *relative locations* of AFSPs versus Banks. Here again there are many approaches that could be taken. For example, one potentially relevant extension of the K-function approach would be to construct *cross K-functions* in which, for example, the number of AFSPs within a given distance of each Banks is

analyzed.[32] Using this approach, one could in principle determine whether there are significantly fewer AFSPs in the neighborhoods of Banks than would be expected if AFSPs and Banks were indistinguishable.[33] Such a "repulsion" effect would in principle be consistent with the spatial void hypothesis.

However, since this approach considers market-area effects only indirectly, we choose to adopt several alternative approaches that focus more directly on questions related to the market areas of AFSPs relative to Banks. As mentioned above, we consider here two approaches: one focusing on the *demand side* and the other on the *supply side*.

Comparisons of Typical Incomes for Potential Customers of Financial Institutions

Turning first to the demand side, we seek to develop a spatial model reflecting possible differences in the incomes of potential customers for financial institutions at different locations. To do so, we begin by approximating potential demand in each block-group j in terms of the *median income level*, y_j, in that block-group. By employing these median incomes, we next construct a measure of "typical incomes" for the potential customers of each financial institution as follows. First we assume that the likelihood that any individual in block-group j is a potential customer of institution i depends on the accessibility of j to i, which is taken to be a decreasing function of the distance, d_{ij}, from i to the centroid of block-group j.[34] This *individual accessibility function*, $a(d_{ij})$, is assumed here to have the following quadratic (kernel) form,

$$(2) \qquad a(d) = a(d \mid b) = \begin{cases} \left[1 - (d/b)^2 \right]^2 & , d \leq b \\ 0 & , d > b \end{cases}$$

which starts at $a(0) = 1$, and falls to zero at some distance, b, designated as the *extent* (or "positive support") of the function. This extent parameter, b, determines the maximum

---

[32] For a discussion of cross K-functions, see for example T.E. Smith, "A Scale-Sensitive Test of Attraction and Repulsion between Spatial Point Patterns," *Geographical Analysis*, 36: 315-331, 2004. Also see the many references cited in that paper.

[33] For further discussion of this "indistinguishability" concept see footnote 39 below.

[34] Again it should be emphasized that without finer information about the locations of individual consumers, these centroids serve as a convenient representative or "typical" locations for the consumers in each block-group.

distance at which institution i can expect to draw customers, as illustrated for $b = 0.5$ (one half mile) in Figure 5.[35] The role of this parameter will be discussed further below.
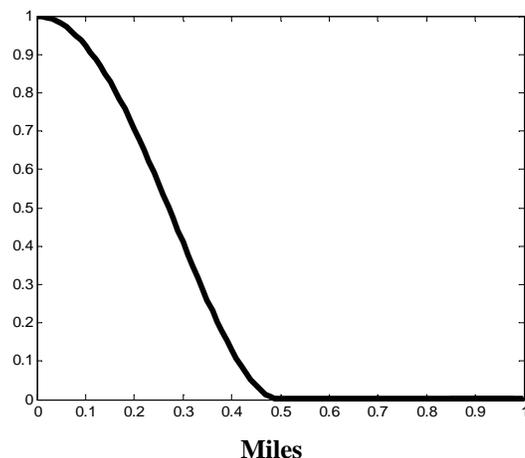

Figure 5. Individual Accessibility Function

Given this individual accessibility function, the expected number of potential i-customers from j is the product of $a(d_{ij})$ and the population, n_j, of block-group j. Hence the probability that any potential i-customer is from j is given by

$$(3) \qquad p_{ij} = \frac{n_j a(d_{ij})}{\sum_{k=1}^{n} n_k a(d_{ik})}$$

In terms of this distribution of potential customers from each block-group, we now take the *typical income*, Y_i, of potential i-customers to be given in terms of *median incomes*[36] by

$$(4) \qquad Y_i = \sum_{j=1}^{n} p_{ij} y_j$$

Without further information about actual market sizes, we assume throughout that the individual access functions in (3) are the *same* for both Banks and AFSPs. Hence the

---

[35] This of course serves only as a rough approximation to actual customer potential, and (as in footnote 30 above) should ideally be estimated on the basis of actual market data.

[36] Note that an estimate of the *expected income* of a potential i-customer would be based on average incomes in block-groups rather than median incomes. Hence our use of median incomes, together with the term "typical", reflects the fact that medians are less sensitive to outliers than are means, and are taken here to be more appropriate *representative incomes*.

present analysis focuses mainly on differences between median-income values in the neighborhoods of these institutions.

Given the spatial model of typical income levels defined by (3) and (4),[37] the key question of interest here is whether these incomes are higher for potential customers of Banks than of AFSPs. If so, then this finding would lend support to the spatial void hypothesis. To answer this question, we again proceed by postulating the null hypothesis that typical income levels of potential customers are the same for Banks and AFSPs, and then test this hypothesis by Monte Carlo methods.

Here we begin by estimating the *expected values* of Y_i for randomly sampled institutions i of each type. If the relevant sets of Bank and AFSP locations (say for Philadelphia County) are denoted respectively by $L_B = \{i : i = 1,..,n_B\}$ and $L_{AFSP} = \{i : i = 1,..,n_{ASFP}\}$,[38] then (maximum likelihood) estimates of these expected values are given respectively by

(5) $$Y_B^0 = \frac{1}{n_B} \sum_{i \in L_B} Y_i^0$$

(6) $$Y_{AFSP}^0 = \frac{1}{n_{AFSP}} \sum_{i \in L_{AFSP}} Y_i^0$$

where each value, Y_i^0, is based on the distances $\{d_{ij} : j = 1,..,n\}$ from the *observed* location of institution i to each block-group (say in Philadelphia County).

To test the hypothesis that expected typical-income values are higher for Banks than for AFSPs, one could proceed in a manner analogous to the tests above by generating random patterns of Banks and AFSPs that are proportional to income distributions. But this would ignore many of the key factors constraining the actual locations of these institutions (such the given street network and local zoning restrictions). Hence to preserve these factors (which may actually influence the extent of potential markets), we choose to formulate an alternative null hypothesis based on random-permutation tests. The key idea here is to take the full set of *institution locations*,

---

[37] It should also be noted that this spatial measure is closely related to the types of "*G*-statistics" developed by Getis and Ord. See A. Getis and J.K. Ord, "The Analysis of Spatial Association by Distance Statistics," *Geographical Analysis*, 24:189-207, 1992.

[38] For ease of exposition, we implicitly assume here that the locations of all financial institutions are distinct, so that locations can be identified with unique institutions. Obvious adjustments can be made when two or more institutions share a common location.

(7) $\quad L = L_B \cup L_{AFSP} = \{i : i = 1,..,n_L\}, \ n_L = n_B + n_{AFSP},$

as given, and to ask what expected income differences would look like if the location behavior of Banks and AFSPs were completely indistinguishable. Here "indistinguishable" is taken to mean that the specific locations called "Banks" are simply one of many equally likely choices of n_B sites from L.[39] Since each choice of these sites amounts to a random re-labeling of sites, the distribution of income differences for Banks and AFSPs under this *indistinguishability hypothesis* can readily be estimated by sampling a large set of N relabeling and re-computing the values in (5) and (6) for each of these samples. This distribution can then be used to test the desired hypothesis as before.

To make these ideas precise, let $\pi = [\pi(1),..,\pi(n_B),\pi(n_B+1),..,\pi(n_L)]$ denote a random permutation (relabeling) of the numbers $(1,..,n_B,n_B+1,..,n_L)$ and for each $i = 1,..,n_L$ let $Y_i^\pi = Y_{\pi(i)}$. Then for this permutation, π, the estimates corresponding to (5) and (6) above are now given respectively by:

(8) $\quad Y_B^\pi \ = \ \dfrac{1}{n_B}\sum_{i \in L_B} Y_i^\pi$

(9) $\quad Y_{AFSP}^\pi \ = \ \dfrac{1}{n_{AFSP}}\sum_{i \in L_{AFSP}} Y_i^\pi$

Of particular interest for our purposes is the difference, $\Delta(\pi) = Y_B^\pi - Y_{AFSP}^\pi$, between these expected typical-income values. If the corresponding observed difference between (5) and (6) is denoted by $\Delta(0) = Y_B^0 - Y_{AFSP}^0$, and if we simulate a large number of random relabeling, $\pi_k, \ k = 1,..,N$,[40] then our interest again focuses on the relative ranking of $\Delta(0)$ in the list of values $[\Delta(0),\Delta(\pi_1),..,\Delta(\pi_N)]$. If the number of values at least as large as $\Delta(0)$ is now denoted by M_Δ, then as a parallel to (1) above, the chance of observing a

---

[39] Statistically, the labels of locations are then referred to as "exchangeable events." From a behavioral viewpoint, this is similar to Debreu's well-known example of choosing "red buses" versus "blue buses." See G. Debreu, "Review of R.D. Luce, Individual Choice Behavior: A Theoretical Analysis," *American Economic Review*, 50: 186-188, 1960.

[40] Here it should be noted that since the individual Y_i values remain the same in each sample, it might appear that simulation is not required at all. However in the Philadelphia case for example, where $n_B = 321$ and $n_{AFSP} = 230$, there are more than 10^{160} distinct subsets of n_B locations. So while exact computations

difference as large as $\Delta(0)$ under the null hypothesis of indistinguishable institutions is estimated to be

$$(10) \quad P_\Delta = \frac{M_\Delta}{N+1}$$

Hence P_Δ now serves as the *p-value* for a one sided test of this indistinguishability hypothesis. Here we again use $N = 1000$, so that if $M_\Delta = 10$ for example, then $P_\Delta = 10/1001 < .01$ would imply in the present context that the chance of observing a difference in expected typical-income values as large as $\Delta(0)$ is less than one in a hundred under this null hypothesis.

Before discussing the results of these tests, it should again be emphasized that such tests implicitly rely on the specific parameterization of the individual accessibility function in (2) above. Moreover (as mentioned in footnote 30 above) this behavioral function is generally not directly observable. However, it has been observed by many researchers that the key parameter determining the *statistical* behavior of such kernel functions is their *extent* (also called their "bandwidth").[41] Hence one advantage of the present testing procedure is that this parameter can easily be varied, so that the test can be conducted for a range of extent values. Thus, in cases where the established test results are robust to variations in extent, it can be argued that this robustness lends considerable support to the test conclusions. To determine robustness in the present case, a range of extent values from one quarter mile up to one and a half miles was used.[42]

In *all* counties tested, the results obtained were very robust to variations in these scale effects, and confirmed that Banks do indeed have significantly higher typical-income levels of potential customers than do AFSPs. Such results are well illustrated by the case of Philadelphia, with an extent, b, of one half mile (paralleling the clustering results presented above). Figure 7 below shows the histogram of values obtained for the observed difference, $\Delta(0)$, together with $N = 1000$ simulated differences,

---

can be done in some simple cases, simulation is still the most practical alternative. For further discussion of this issue see for example E.S. Edgington, *Randomization Tests*, Marcel Dekker: New York, 1995.

[41] This observation was first made by Silverman (1978), but has since been confirmed by many others. See B.W. Silverman, "Choosing a Window Width when Estimating a Density," *Biometrika*, 65: 1-11, 1978 and N.A.C. Cressie. *Statistics for Spatial Data*, Wiley: New York, section 8.5, 1993.

[42] In units of miles, the actual extent values used were $b = [.25,.50,.75,1.00,1.50]$.

$[\Delta(\pi_1),..,\Delta(\pi_N)]$, at the same extent value. Here the observed difference in expected typical-income levels for potential customers of Banks and AFSPs is little more than
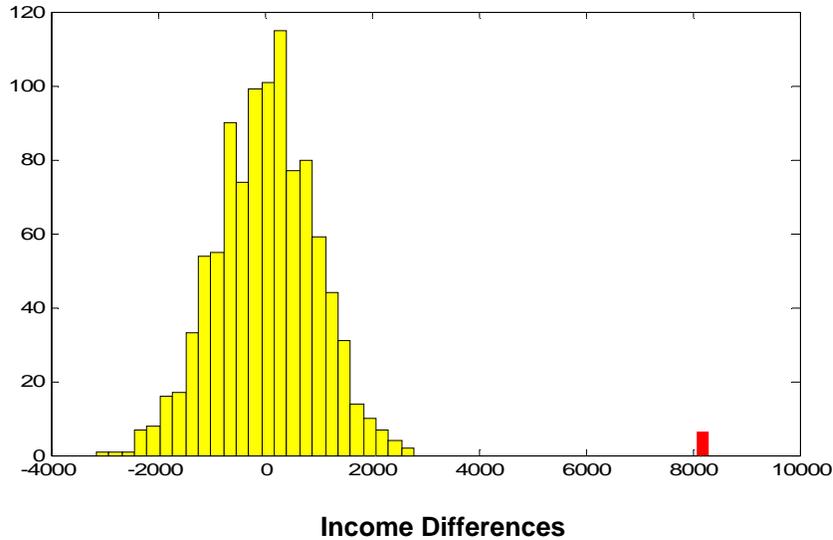

Figure 7. Expected Income Differences for Philadelphia

$8000, as shown by the red bar in the figure. The results of 1000 simulations (random relabellings) are shown in yellow. Thus, in this case (as with all other extent values simulated) the estimated p-value is $P_\Delta \leq 1/1001 \approx .001$. Moreover since the observed value, $\Delta(0)$, is an extreme outlier, it is clear that larger simulations would produce even stronger results. Hence it can be concluded on the basis of these tests that typical customer incomes (as we have defined them) are significantly higher for Banks than for ASFPs.

Comparisons of the Relative Access of Potential Customers to Banks and AFSPs

While the above testing procedure is able to determine whether there is a significant overall difference in typical customer incomes between Banks and AFSPs, it provides no information as to *where* these differences are occurring in space. Hence the objective of the present approach is to add a *spatial dimension* to the analysis above. To do so, we turn to the supply side of the market, and focus on the question of which block-groups of potential customers have significantly greater access to one type of institution than the

22

other. Once these block groups are identified, we shall revisit the question of income differences between these block-groups.

In view of the analysis above, one natural approach to this question would be to employ the individual accessibility functions in (5) to determine access. In particular, one could use appropriate modifications of (5) to calculate summary measures of the *relative access* of individuals in block-group j to Banks and AFSPs at their respective locations in L_B and L_{AFSP}. The corresponding distribution of *access differences* could then be simulated under the indistinguishability hypothesis above, and tested for each block-group. But since this would require the recalculation of access differences for every block-group in every simulation,[43] we choose to develop a simpler approach which yields exact probability calculations and avoids the need for Monte Carlo methods altogether.

In particular, we now revisit the *K-function* approach developed for the analysis of individual clusters above, and apply an appropriate modification of this approach in the present setting. Recall from Figure 2 that to identify clustering of financial institutions, the key idea was to count the number of additional institutions of the same type within distance d of any given institution, i. Here the attention shifts to customers, so that for any given block-group, j, we now count the numbers of the institutions (both Banks and AFSPs) within distance d of the centroid, c_j, of j. An illustrative example is shown in Figure 8 below, where thirteen financial institutions are seen to be within one half mile ($d = 0.5$) of c_j.[44] In this case there are eleven AFSPs (red) and only two Banks (blue). So, given the fact that there are more Banks ($n_B = 321$) than AFSPs ($n_{AFSP} = 230$), this would appear to be a very significant concentration of AFSPs.

---

[43] For the case of Philadelphia with 1806 block-groups and 1000 simulations, this would require close to 2 million accessibility-difference calculations.

[44] The actual location of this block-group is shown in Figure 9 below.
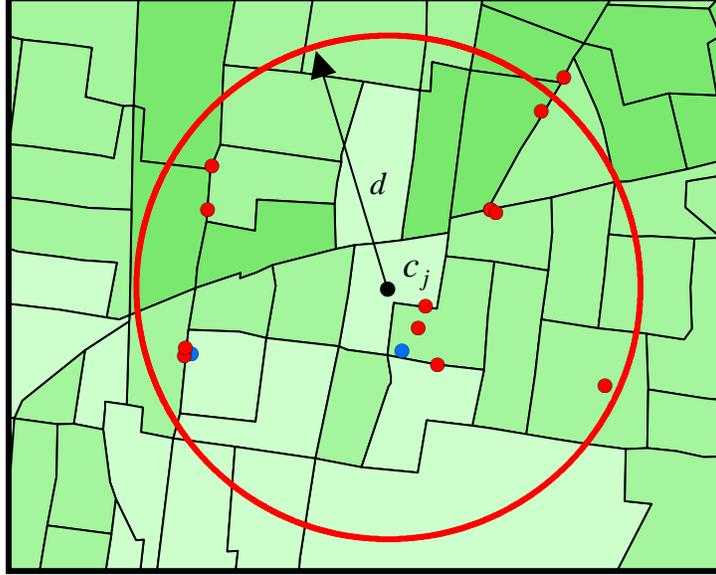
Figure 8. Institution Counts for Block-Groups

To test this conjecture, we again appeal to the *indistinguishability hypothesis* above, and suppose that this assignment of "Banks" and "AFSPs" to locations in L is only one among many possible equally-likely labeling of these locations. Under this hypothesis, if there are a total of m institutions within distance d of c_j, and if there are $K \, (= n_{AFSP})$ AFSPs in the total population of $M \, (= n_L)$ institutions (say in Philadelphia County), then the chance that k of these m institutions will be AFSPs is well known to be given by the *hypergeometric* probability:[45]

$$(11) \qquad p(k \,|\, m, K, M) = \frac{\binom{K}{k}\binom{M-K}{m-k}}{\binom{M}{m}} = \frac{\left(\dfrac{K!}{k!(K-k)!}\right)\left(\dfrac{(M-K)!}{(m-k)!(M-K-m+k)!}\right)}{\left(\dfrac{M!}{m!(M-m)!}\right)}$$

Hence if the random count variable, C, denotes the number of AFSPs within distance d of c_j, then the chance of observing at least k AFSPs is given by

---

[45] Of course, these probabilities are well approximated by the Binomial distribution when *M* and *K* are both sufficiently large relative to *m*. But this may fail in the present case when there are relatively few AFSPs compared to Banks. A recent discussion of this approximation issue can be found in F. Lopez-Blazquez and

(12) $P_j(d) = \text{Prob}(C \geq k \mid m, K, M) = \sum_{c=k}^{m} p(c \mid m, K, M)$

In the present case, it is this cumulative probability, $P_j(d)$, that serves as the appropriate *p-value* for a one-sided test of the *indistinguishability hypothesis* for block-group j at radial distance d. For example, in the case illustrated in Figure 9 (where $d = 0.5$),

(13) $P_j(d) = \text{Prob}(C \geq 11 \mid 13, 230, 551) = .0018$

Hence if Banks and AFSPs were in fact indistinguishable, then the chance of such an extreme concentration of ASFPs would be extremely small. On this basis, it may thus be concluded that this block-group has an unusually large supply of AFSPs relative to Banks. Here it is also worth noting that *no* AFSPs or Banks are actually inside this block-group. This underscores the need to define "accessibility" more broadly. Notice also from the figure that there is one additional AFSP just beyond this half-mile cutoff. So it should also be clear that one must consider a range of extents (d values) in order to gage the proper scale of this relative concentration.

An additional attractive feature of this method is that such p-values can be *mapped* in a manner similar to Figures 3b and 4b above, to yield a picture of "relative concentration hot spots." Here we choose to show a broader range of p-values than the simple ".01-threshold" value in Figures 3b and 4b. Hence in Figure 10 below, all p-values, $P_j(d) \leq .10$ are shown, where the darkest p-values are those where the most significant relative concentration of ASFPs occurs.[46] Again, these values are shown only for the half-mile scale ($d = 0.5$), which appears to provide the most meaningful concentrations for the Philadelphia case (with larger radii tending to produce more diffuse concentrations, and smaller radii more granular concentrations).[47] For example, the case shown in Figure 8 above corresponds to the block-group at the center of the red half-mile radius shown on the map. This p-value is in the darkest category, and is seen to correspond roughly to the center of a large area of relatively concentrated AFSPs in North Philadelphia. This relative concentration is quite consistent with Figures 3b and 4b above, where there is seen to be a significant clustering of AFSPs in Figure 3b but no

B. Salamanca Mino, "Binomial Approximation to Hypergeometric Probabilities," *Journal of Statistical Planning and Inference*, 87: 21-29, 2000.

[46] Again, the p-values for adjacent block-groups are necessarily positively correlated. So while this approach (in our view) is very informative, no precise conclusions can be drawn about the overall statistical significance of each "hot spot".

[47] An illustration of two alternative scales is shown for Banks in Figure 10 below.

significant clustering of Banks in Figure 4b. At the other extreme, it is interesting to notice that while there was no significant clustering of either AFSPs or Banks in West Philadelphia, there is a significant relative concentration of AFSPs precisely in the area where significant clustering of AFSPs was indicated by the NNHC approach in Figure 3a.
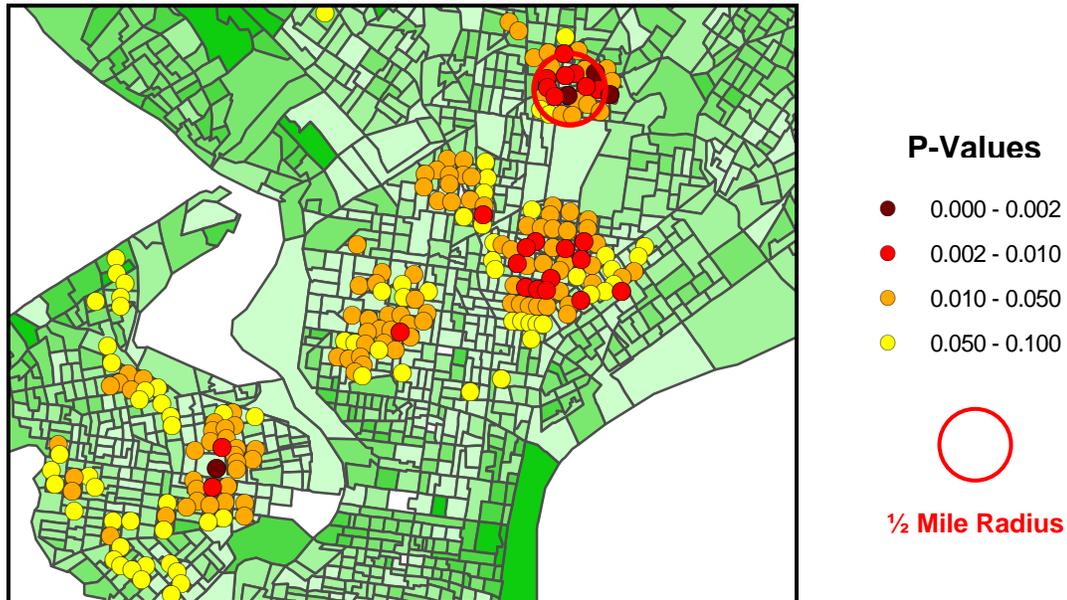

Figure 9. Relative Concentrations of AFSPs versus Banks

So even though the residential population density is too high to attribute any significance to the number of AFSPs in this area, their *relative* number compared to Banks is quite significant.[48] Finally, it is of interest to note while Figure 3b and 4b showed significant clustering of both AFSPs and Banks in Center City, it is evident from Figure 9 that there is no significant *relative concentration* of AFSPs compared to Banks.

Before discussing issues related to the income levels of potential customers in these concentrations, it is appropriate to conduct the same analysis for Banks versus AFSPs. The results are shown in Figure 10 below (using the same p-value scale as in Figure 9). Results for a half-mile radius ($d = 0.5$) are shown on the left; and for comparison, the corresponding results for a one-mile radius ($d = 1.0$) are shown on the right. Turning first to the half-mile case, which is directly comparable to Figure 9 above,

---

[48] It is of interest to notice from Figure 4a above that a significant concentration of Banks was also identified by NNHC in roughly the same area. But closer inspection shows that this cluster is precisely in

the most striking feature here is the significantly high relative concentration of Banks versus AFSPs in the Center City area of Philadelphia. This is of course not surprising, given the nature of central business districts. But notice that this result does add information to Figure 9 – since it is in principle possible that neither Banks nor AFSPs would exhibit significant relative concentration compared to the other. Observe next that

$(d = 0.5)$ $(d = 1.0)$
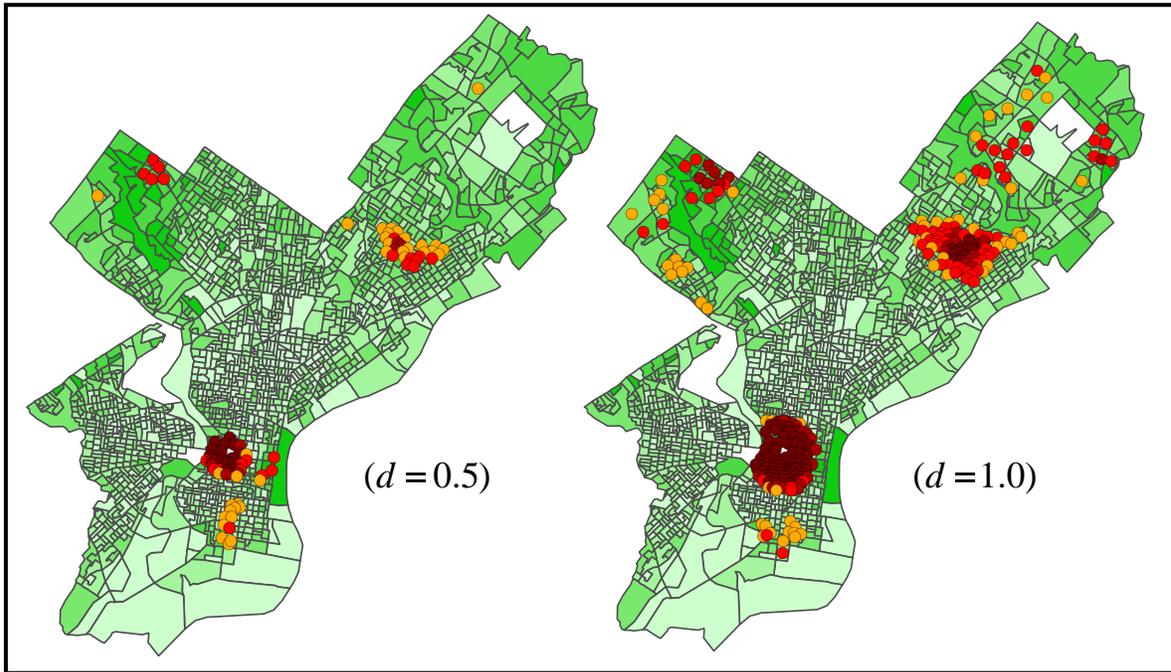
Figure 10. Relative Concentration of Banks

Banks tend to exhibit some degree of significant concentration relative to AFSPs in almost all areas where they are significantly clustered in Figure 4b. This significance is weaker in Northwest and Northeast Philadelphia primarily because there are fewer financial institutions in these areas. For example, the significant cluster of Banks farthest to the right in Figure 4b happens to have exactly five banks. But for the two block-groups within one half mile of this entire cluster there is also one AFSP within the same half-mile radius. Thus, given the fact that there are considerably more banks in Philadelphia than AFSPs, the chance of five Banks occurring in this sample of size six is actually more than one in five.[49] Note finally that while the results for a one-mile radius are

---

the University area mentioned above, and that the AFSP cluster is just to the west of this University area, where there is a marked shift in local demographics.

[49] In terms of the notation in (12) the exact probability here is $\text{Prob}(C \geq 5 \mid 6,321,551) = .206$.

qualitatively similar, the areas of significantly high relative concentration are larger and somewhat more difficult to interpret. For example the "Center City" concentration for the half-mile case has now spread into North and South Philadelphia, and is much less clearly identified with Center City itself.

Finally, we employ these relative concentration results to re-examine the question of income differences between potential customers of Banks and AFSPs. Here one could of course conduct a number of standard difference-between-means tests. But to obtain a result that is more in the spirit of Figure 7 above, it is of interest to consider a Monte Carlo test based on an appropriate modification of the *indistinguishability hypothesis*. Here we again start with AFSPs and identify the set of block-groups with significant relative concentration of AFSPs. More precisely, if the set of all block-groups in the given area (e.g., Philadelphia County) is denoted by, $BG = \{j : j = 1,..,n\}$, then in terms of (12) with $d = 0.5$, we first identify the subset

(14) $BG_0 = \{j \in BG : P_j(d) \le .05\}$

of block-groups with significant relative concentration of AFSPs at the .05 level. If n_0 denotes the number of block-groups in BG_0, and if the average median income for this set of block-groups is denoted by

(15) $\overline{Y}_0 = \dfrac{1}{n_0} \sum_{j \in BG_0} y_j$,

then the appropriate null hypothesis for our present purposes is that median incomes for BG_0 should be typical of those for random samples of n_0 block-groups from BG. In particular, $\overline{Y}_0$ should constitute a typical sample from the distribution of $\overline{Y}$ for samples of size n_0 from BG. This distribution is again easily simulated by Monte Carlo methods. In the present case, if we draw N independent random samples of size n_0 from BG, and denote their realized average median incomes by $\overline{Y}_k, k = 1,..,N$, then we can construct p-values as before by letting M_0 denote the number of samples in the list $[\overline{Y}_k : k = 0,1,..,N]$ that are no greater than $\overline{Y}_0$. The appropriate p-value, P_0, for a one sided test of this null hypothesis is then given by

(16) $P_0 = \dfrac{M_0}{N+1}$

As in the case of Figure 7 above, these results turn out to be so decisive that it is more informative to simply plot the simulated values for this test. The resulting histogram is shown in Figure 11 below, where the red bar denotes the observed value, $\bar{Y}_0 = \$20,107$, and the yellow bars represent the 1000 simulated $\bar{Y}$ values (with the scale in thousands of dollars). The average median income for all Philadelphia block-groups is $30,686, which (not surprisingly) is just about at the center of the yellow bars. So it is clear from these results that the median incomes in those block-groups with a significantly high concentration of AFSPs versus Banks are definitely below those that would be expected for a comparable random sample of block-groups. Hence this result again provides strong support for the spatial void hypothesis in Philadelphia County.
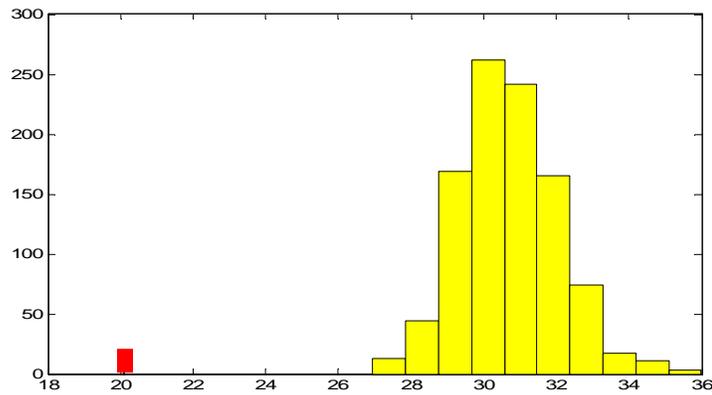
Figure 11. Income Test Results for AFSPs

Exactly the same analysis can be performed for Banks, except that the appropriate question is now reversed. In particular, for block-groups with significant relative concentrations of Banks [as in (14)] we now ask whether their average median income is significantly *higher* than would be expected for random block-group samples of the same size. Here there are $n_0 = 51$ block-groups with significant relative concentration of Banks (at the .05 level), and the average median income for these block-groups is $\bar{Y}_0 = \$40,796$. The results for this case are again extreme, and are best shown by a histogram of values,
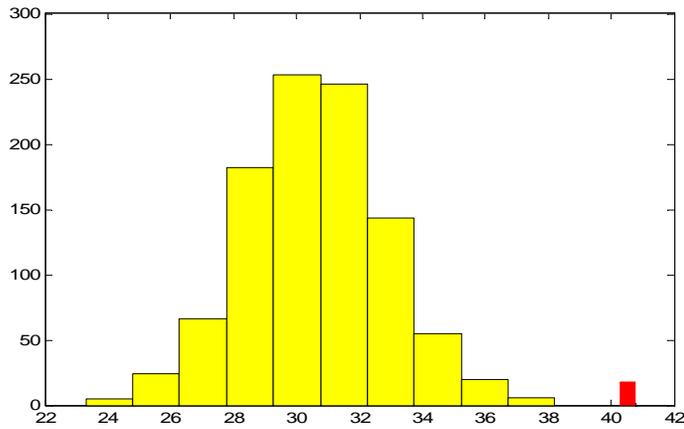
Figure 12. Income Test Results for Banks

as in Figure 12. Thus, the average medium income is here seen to be significantly *higher* than would be expected for random block-group samples of the same size. Though this result is not as dramatic as that for AFSPs, it is nonetheless surely significant beyond the .001 level [as estimated by (16)]. Hence, as a "flip side" of the spatial void hypothesis above, it also appears that areas where Banks are significantly more concentrated than AFSPs also have significantly high median incomes.

Finally we note that closer parallels to Figure 7 above could easily be constructed by focusing directly on *differences* between $\overline{Y}_0$ values associated with the significant block-group sets, BG_0, for Banks and AFSPs respectively. These differences could then be used to construct similar Monte Carlo tests of "no difference" hypotheses – which would of course yield dramatic results, as we have seen. But the present approach provides useful additional information: namely, that $\overline{Y}_0$ for APSPs is in fact significantly *lower* than typical median incomes in Philadelphia and that $\overline{Y}_0$ for Banks is significantly *higher*.

Table 3 shows a comparison of median incomes of block-groups with significant relative access to AFSPs clusters to median incomes for the county as a whole. Again using Philadelphia County as an example, the first column indicates that there are 28 block-groups where relative accessibility to AFSPs versus Banks is significant at the .01 level. The second column shows that the average of medium incomes of these block-groups is $19,900 which, in turn, is compared to an average median income of $30,686 (column three) over all block-groups in Philadelphia County. This implies that the potential customers of AFSPs have relatively lower incomes. Finally, the last column reveals that compared to 10,000 random samples of 28 block-groups in Philadelphia County, the chance of getting an average medium income as low as $19,900 is less than 1 in 10,000.

TABLE 3. MEDIAN INCOME OF BLOCK-GROUPS

	ASFPs CLUSTERS	ASFPs MEDIAN INCOMES	COUNTY MEDIAN INCOME	P-VALUE
Philadelphia County	28	$19,900	$30,686	<.0001
Delaware County	44	$35,426	$63,347	<.0001
Allegheny County	14	$34,592	$49,988	0.0008
Montgomery County	35	$40,078	$75,830	<.0001

As can be seen in the Table 3, the number of block-groups among the four counties ranges from 14 in Allegheny to 44 in Delaware County. As in the case of Philadelphia County, the median incomes of the pertinent block-groups in the other three counties are markedly less than the median incomes of all block-groups in their respective counties. This indicates support for the spatial void hypothesis in all four counties. Moreover, the results are quite significant, even in Allegheny County where there are only 14 out of 1103 block-groups with significantly more access to AFSPs than Banks.

Finally, it should be noted that we have not included a parallel set of results for Banks (as suggested by Figure 12 above). The reason is that there are so few ASFPs in

counties other than Philadelphia, that *no* block-groups exhibit significantly high relative access to Banks (at any scale of analysis). Indeed, even if *all* service providers within a given radius of a block-group are Banks, this will not be a "rare event" if there are sufficiently few ASFPs. So the present notion of "relative accessibility" is clearly of limited use in such cases.

Demographic Makeup of AFSP Locations

In addition to the issue of whether AFSPs fill a void due to the absence of or relative accessibility to Banks, the question often raised is whether they are located predominantly in minority areas. This question is examined by considering the location patterns of AFSPs in our four study counties.

Location of AFSPs and Minorities

Figure 13 shows pie charts depicting a comparison of the ethnic or racial makeup of block-groups with significant relative access to AFSP clusters to the ethnic or racial makeup in the entire county. By comparing the ethnic composition of these block-groups with that in their county, it can be determined whether minorities are disproportionately represented in the designated block-groups. For example, minorities comprise 60.3 percent of the population in Philadelphia County while they are 88.1 percent of those block-groups with significant relative access to AFSP clusters. Thus, minorities are overrepresented in those designated block-groups. A similar pattern of minorities being overrepresented occurs in the other three counties.

A closer look at the representation of specific minority groups yields an interesting picture. African Americans are over-represented in block-groups with significant relative access to AFSP clusters in Montgomery, Delaware, and Allegheny counties, but under-represented in Philadelphia County (where they are 46.1 percent in

32

Philadelphia County (City of Philadelphia)

3.9% 1.9%
8.4%
39.7%
46.1%

5.6% 2.8% 11.9%
45.5%
34.2%
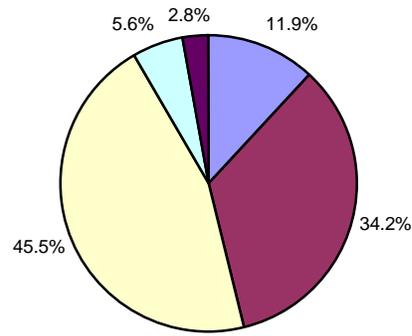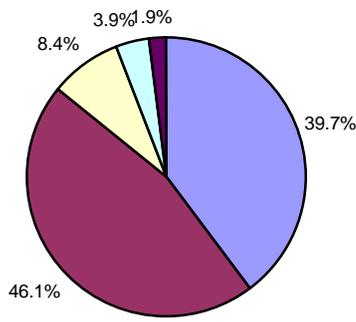
□White ■African American □Hispanic □Asian ■Other

Montgomery County

2.0% 3.3% 0.9%
7.8%
86.0%

2.0%
2.9%
10.6%
54.2%
30.3%
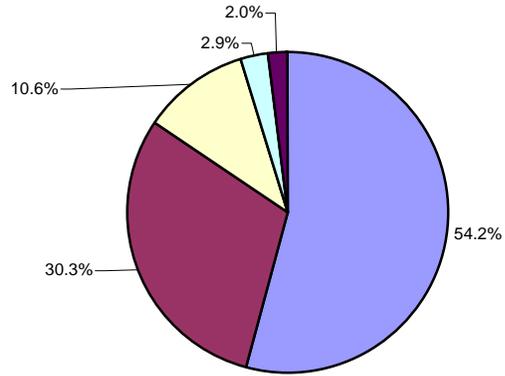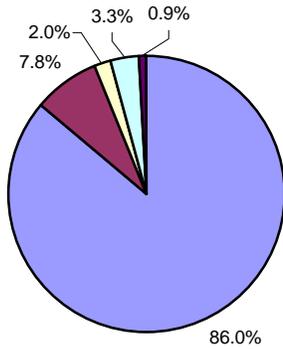
□White ■African American
□Hispanic □Asian
■Other

Delaware County

3.0% 0.9%
1.4%
15.3%
79.4%

2.2%
10.6%
3.2%
34.0%
50.0%
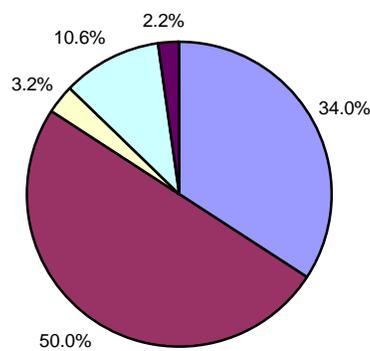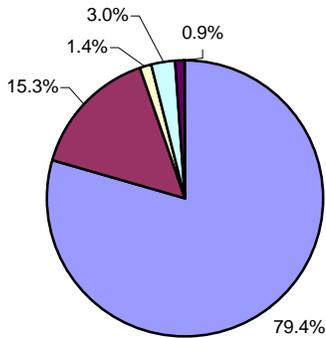
□White ■African American
□Hispanic □Asian
■Other

Allegheny County

1.3% 0.9%
0.8%
13.2%
83.8%

1.3%
1.3%
1.1%
38.9%
57.4%
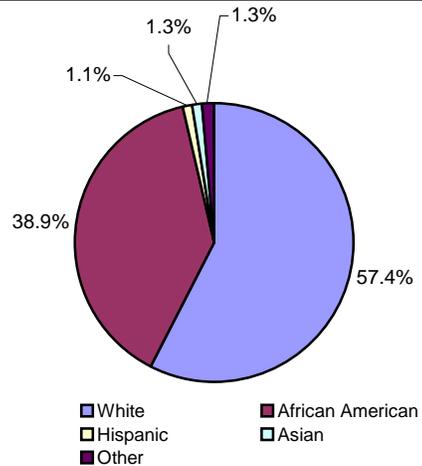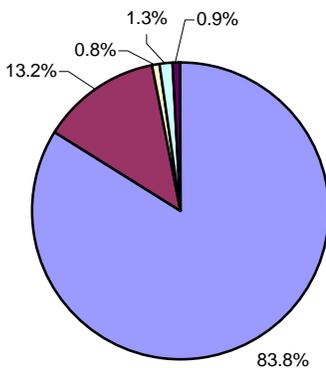
□White ■African American
□Hispanic □Asian
■Other

*Other includes American Indian Not-Hispanic, Other Race Not-Hispanic, Hawaiian Pacific Islander Not-Hispanic, and Two or More Races Not-Hispanic.

33

FIGURE 13. RACIAL OR ETHNIC COMPOSITION OF BLOCK-GROUPS

the county and 34.2 percent in the designated block-groups).[50]   However, there is uniformity in the over-representation of Hispanics in the pertinent block-groups in all of the counties.

The representation of Asian in block-groups with significant relative access to AFSP clusters is more varied. Asians are overrepresented in the designated block-groups in Philadelphia and Delaware counties and underrepresented in Montgomery and (slightly in) Allegheny counties.

It is interesting to note that in Philadelphia County, Hispanics are 8.4 percent of the county but 45.5 percent in the pertinent block-groups. This over-representation of Hispanics more than offsets the under-representation of African Americans, thus resulting in an overall overrepresentation of minorities as mentioned earlier.[51]

SUMMARY AND CONCLUSIONS

The increased prominence of alternative financial providers (AFSPs) in our communities across the U.S. has not gone without notice. In general, AFSPs provide some of the financial services offered by mainstream financial institutions, but typically at a higher price. Moreover, AFSPs do not supply the types of products and services that foster asset creation.   In light of the prevailing wisdom that many minority and low-income households rely heavily on their services, these consumers might forego valuable wealth-building opportunities. The present study has focused on (i) whether the high patronage of this non-bank segment of the financial industry is due to the absence of mainstream financial institutions in the area—known as the *spatial void hypothesis*—and (ii) whether these AFSPs are disproportionately located near minority and low-income households. This investigation was carried out by focusing on four counties in the Commonwealth of Pennsylvania (Philadelphia, Delaware, Allegheny, and Montgomery).

A previous attempt to investigate the "spatial void hypothesis" relied on the relative location of AFSP and Bank clusters identified by the Nearest Neighbor Hierarchical Clustering (NNHC) procedure. However, in the present study it was shown that by failing to account for population concentrations (and also by ignoring "edge effects"), this NNHC approach can sometimes produce misleading results. Hence an

---

[50] These results were also confirmed by employing the same test as summarized in Table 3 above, with "%African American" replacing "Median Income".

[51] These results were again confirmed by the method in footnote 50 above.

alternative approach (based on Ripley's *K-function*) was proposed which not only overcomes these limitations but also permits a systematic evaluation of clustering at different spatial scales.

The difference between these approaches was illustrated for the case of Philadelphia County, where some clusters identified as significant by the NNHC method are no longer significant once (high) population density is accounted for. Similarly, some new clusters are identified by the K-function method, which become significant once (low) population density is accounted for.

In order to test the spatial void hypothesis with respect to these four study counties, we have developed two complementary approaches that focus more directly on issues related to the market areas of AFSPs relative to Banks. On the demand side, we employed a simple measure of "typical incomes" together with a modification of the K-function method above to examine possible differences in the incomes of potential customers of AFSPs and Banks at different locations. On the supply side, we developed similar tests of the relative accessibility of Banks and AFSPs to customers at different spatial locations. In each of these approaches, we found support for the spatial void hypothesis in all four counties studied. In addition, we found that the neighborhoods served by AFSPs are indeed characterized by an over-representation of minority groups.

But, perhaps the most compelling result of this study is to demonstrate the usefulness of certain alternative statistical methods for investigating the spatial void hypothesis. These flexible Monte Carlo techniques help overcome many of the shortcomings of prior efforts, and in our view, offer a fruitful approach for further investigations of the *spatial void hypothesis*.

References

Anselin, L. "Local Indicators of Spatial Association: LISA," *Geographical Analysis*, 27:93-116, 1995.

Besag, J. and J. Newell. "The Detection of Clusters in Rare Diseases," *Journal of the Royal Statistical Society*, *A*, 154: 327-333, 1991.

Burkey, Mark and Scott Simkins. "Factors Affecting the Location of payday Lending and Traditional Banking Services in North Carolina," *Review of Regional Studies*, 34(2): 191-205, 2004.

Carr, James and Jenny Schurtz. *Financial Services in Distressed Communities: Framing the Issue, Finding Solutions*, Washington, D.C.: The Fannie Mae Foundation, 2001.

Caskey, John. *Fringe Banking: Check Cashing Outlets, Pawnshops, and the Poor*, New York: Russell Sage Foundation, 1994.

Castro, M.C. and B.H. Singer. "Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Tests in Local Statistics of Spatial Association," *Geographical Analysis*, 38: 180-208, 2006.

Cressie, N.A.C. *Statistics for Spatial Data*, Wiley: New York, section 8.5, 1993.

Debreu, G. "Review of R. D. Luce, Individual Choice Behavior: A Theoretical Analysis," *American Economic Review*, 50: 186-188, 1960.

Edgington, E.S. *Radomization Tests*, Marcel Dekker: New York, 1995.

Flannery, Mark and Katherine Samolyk. "Payday Lending: Do the Cost Justify the Price?" FDIC Center for Financial Research, Working Paper, No. 2005-09, 2005.

General Accounting Office. "Electronic Transfers." Report to the Subcommittee on Oversight and Investigations, House of Representatives, September, 2002.

Getis, A. and J.K. Ord. "The Analysis of Spatial Association by Distance Statistics," *Geographical Analysis*, 24: 189-207, 1992.

Goldman, Gerald and James R. Wells, Jr., *Check Cashers Are Good Bank Customers*, Hackensack, N.J.: Financial Services Centers of America, Inc., 2002.

Graves, Steven, M. "Landscapes of Prediction, Landscapes of Neglect: A Location Analysis of Payday Lenders and Banks," *The Professional Geographer*, 55(3): 303-317, 2003.

Kulldorff, M. "A Spatial Scan Statistic," *Communications in Statistics: Theory and Methods*, 26: 1487-1496, 1997.

Levine, N., *CrimeStat II*, Washington, D.C.: The National Institute of Justice, 2002.

Lopez-Blazquez, F. and B. Salamanca Mino. "Binomial Approximation to Hypergeometric Probabilities," *Journal of Statistical Planning and Inference*, 87: 21-29, 2000.

Pawnbroker's Licensing Act of 1937, P.L. 200, No. 51.

Pennsylvania Bulletin, Volume 28, Number 18, Saturday, May 2, 1998, Harrisburg, PA.

Ripley, B. D. "The Second-order Analysis of Stationary Point Patterns," *Journal of Applied Probability*, 13: 255-266, 1976.

Ruston, G. and P. Lolonis. "Exploratory Spatial Analysis of Birth Defects in an Urban Population," *Statistics in Medicine*, 15: 717-726, 1996.

Sawyer, Noah and Kenneth Temkin. *Analysis of Alternative Financial Service Providers*, Washington, D.C.: The Fannie Mae Foundation, 2004.

Silverman, B.W., "Choosing a Window Width when Estimating a Density," *Biometrika*, 65: 1-11, 1978.

Smith, T.E. "A Scale-sensitive Test of Attraction and Repulsion between Spatial Point Patterns," *Geographical Analysis*, 36: 315-331, 2004.

Stegman, Michael. "Banking the Unbanked: Untapped Market Opportunities for North Carolina's Financial Institutions," *Journal of the University of North Carolina School of Law*, Chapel Hill, NC, 2001.

APPENDIX

Comparison of the Two Clustering Methods for the Study Counties
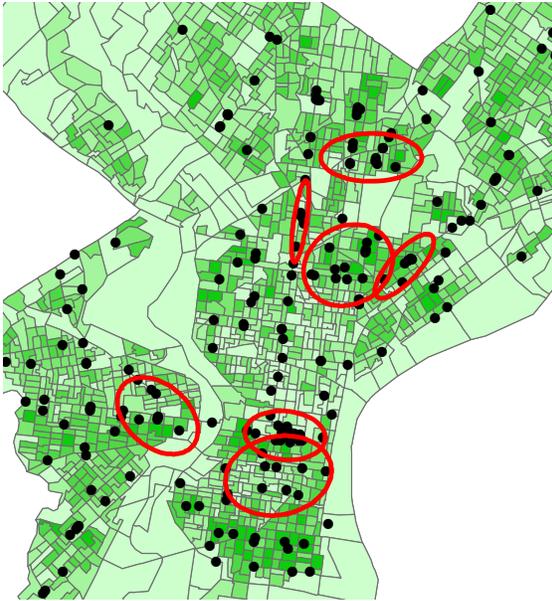
Crime Stat cc-Clusters
(P-value = .01, Count >= 5)
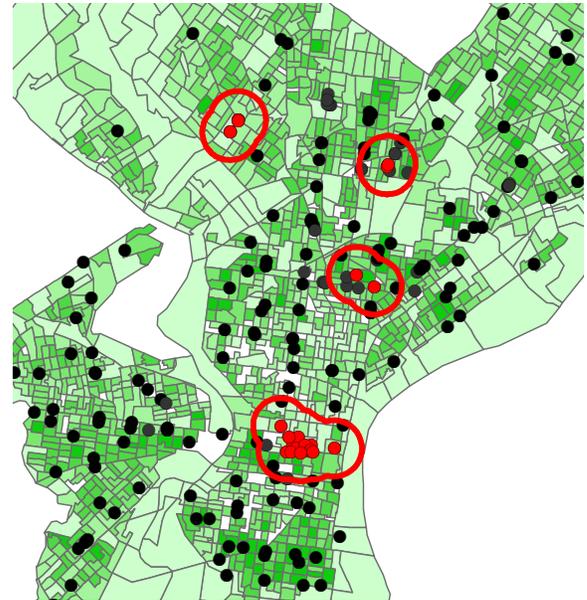
Philly cc-Clusters
(One-Half Mile radius, Count >= 5)
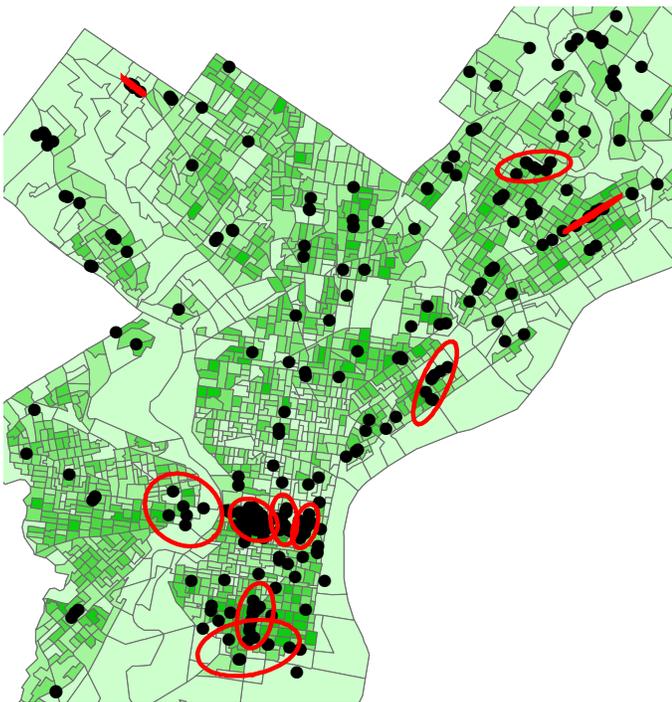
⬤ **P-Value <= .01**

Crime Stat b-Clusters
(P-value = .01, Count >= 5)
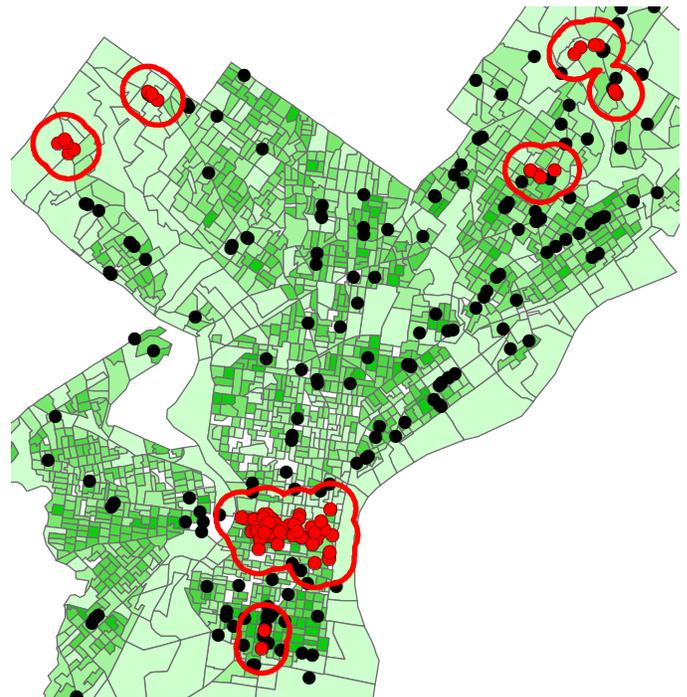
Philly b-Clusters
(One-Half Mile radius, Count >= 5)
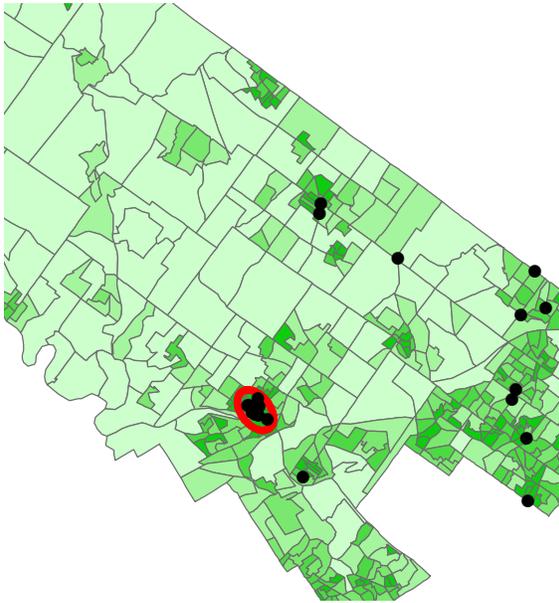
⬤ **P-Value <= .01**

Crime Stat cc-Clusters
(P-value = .01, Count >= 5)
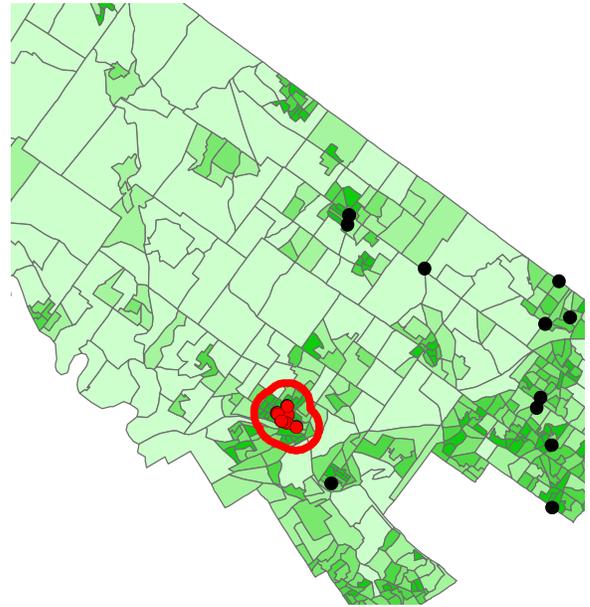
Montgomery cc-Clusters
(One Mile radius, Count >= 5)
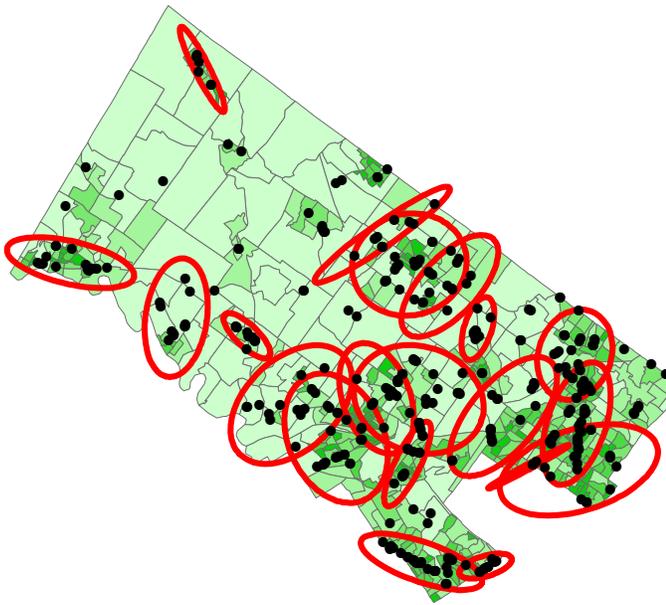
🔴 **P-Value <= .01**

Crime Stat b-Clusters
(P-value = .01, Count >= 5)
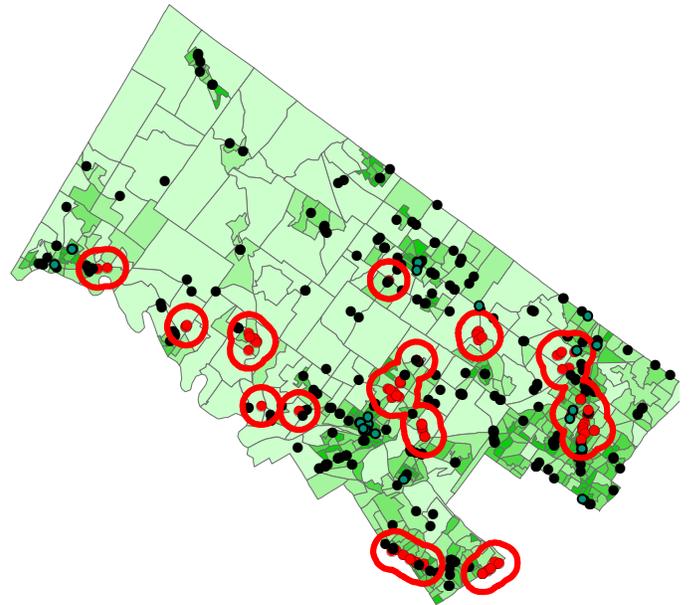
Montgomery b-Clusters
(One Mile radius, Count >= 5)
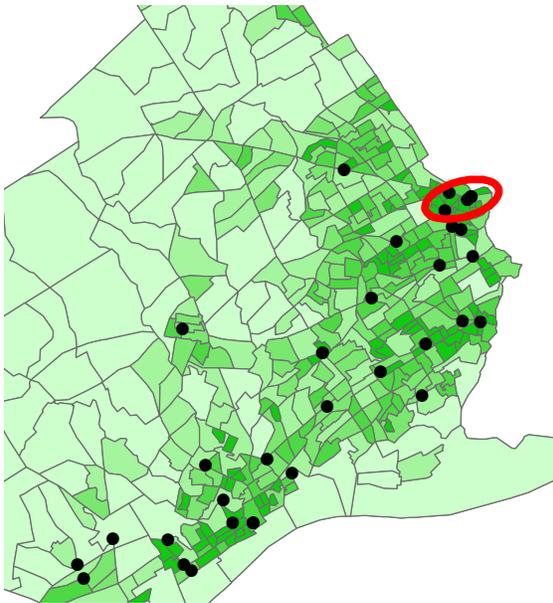
🔴 **P-Value <= .01**

Crime Stat cc-Clusters
(P-value = .01, Count >= 5)
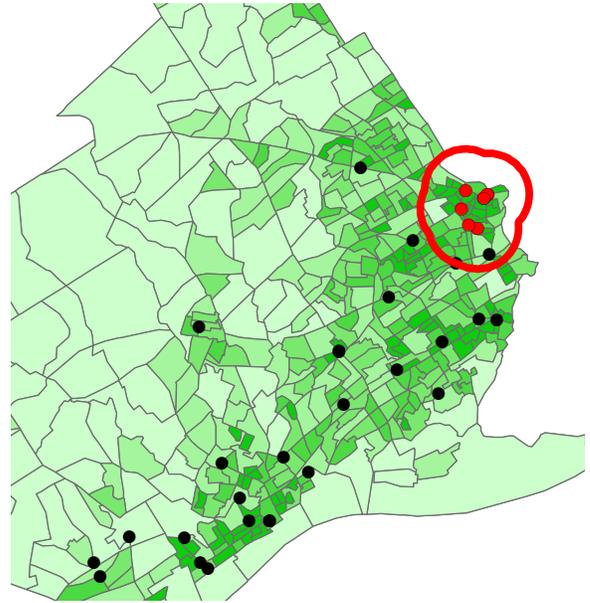
Delaware cc-Clusters
(One Mile radius, Count >= 5)
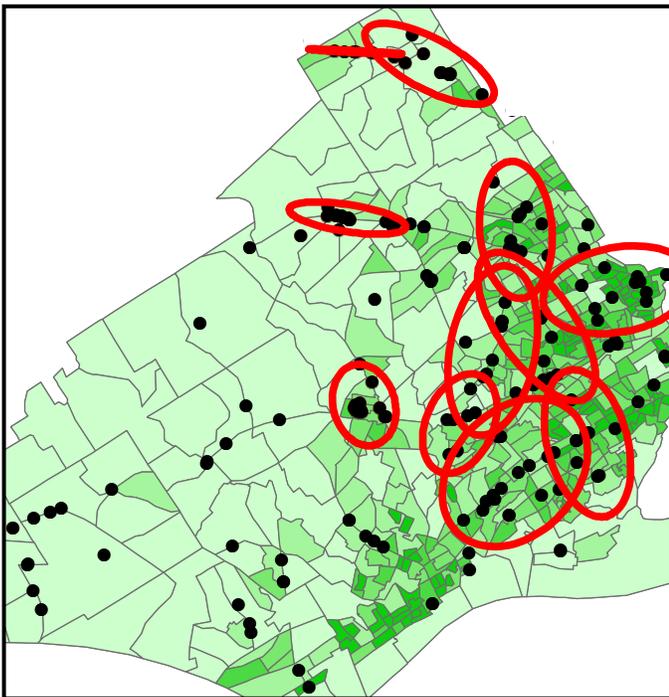
🔴 **P-Value <= .01**

Crime Stat b-Clusters
(P-value = .01, Count >= 5)
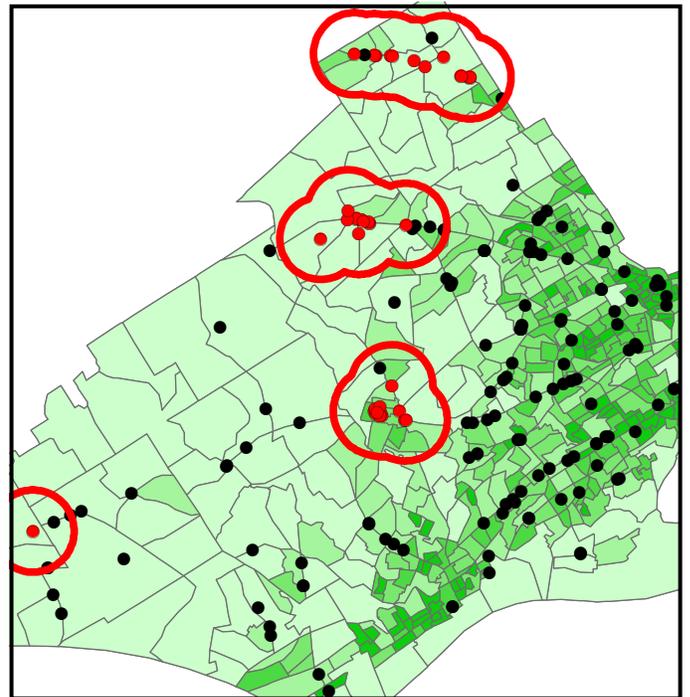
Delaware b-Clusters
(One Mile radius, Count >= 5)
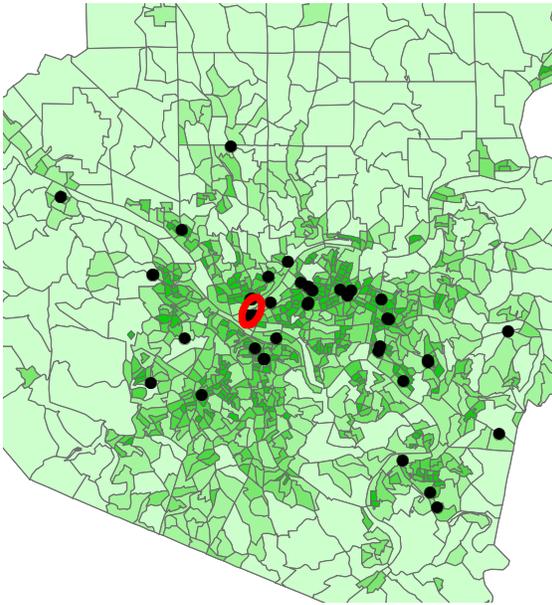
🔴 **P-Value <= .01**

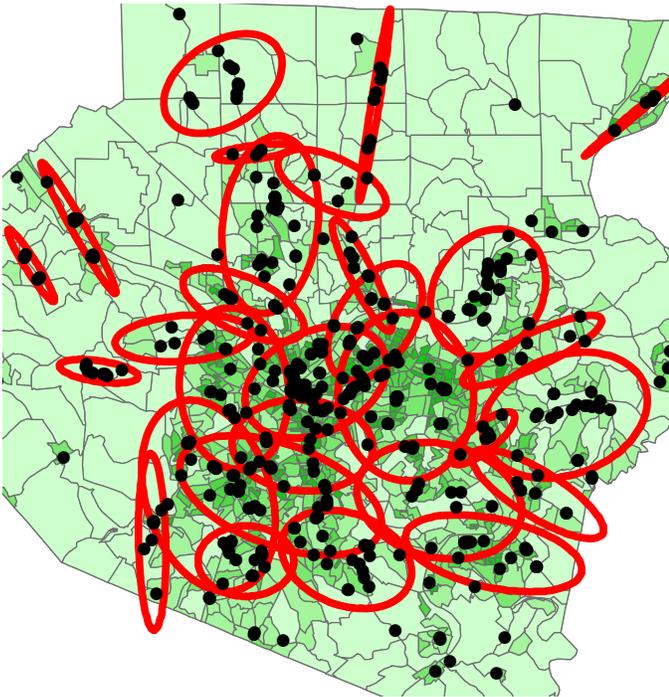**Crime Stat cc-Clusters**
(P-value = .01, Count >= 5)
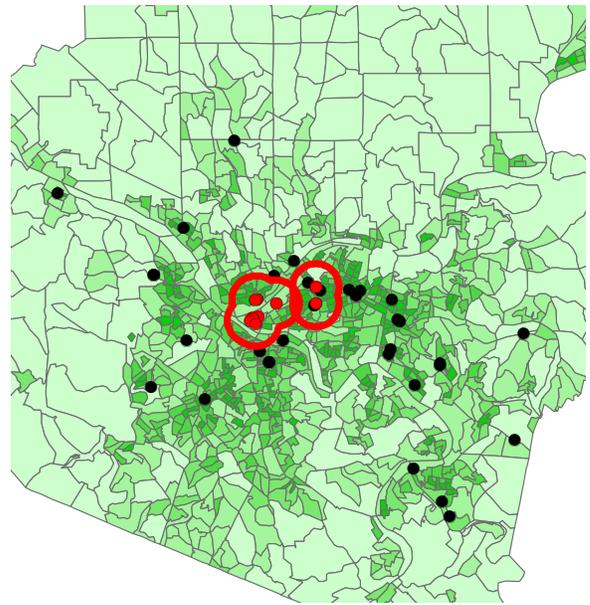
Allegheny cc-Clusters
(One Mile radius, Count >= 5)

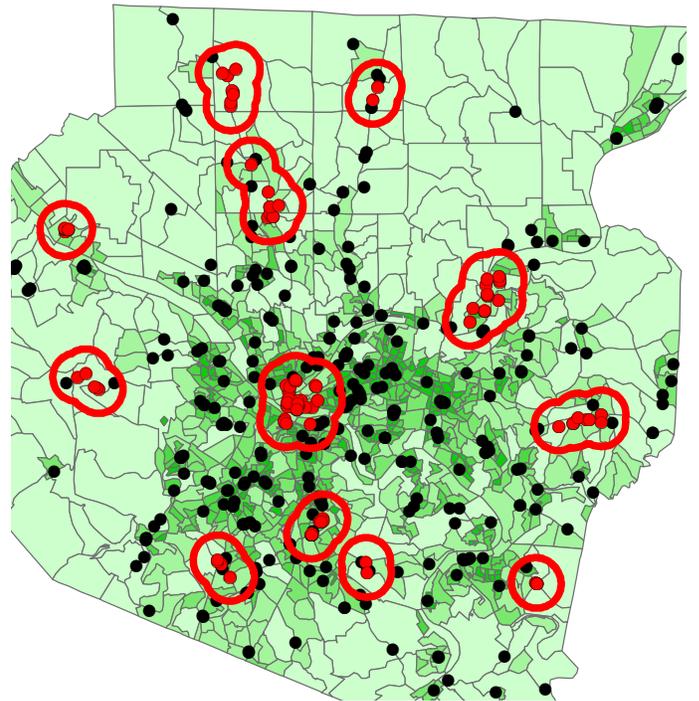🔴 **P-Value <= .01**

Crime Stat b-Clusters
(P-value = .01, Count >= 5)

Allegheny b-Clusters
(One Mile radius, Count >= 5)

🔴 **P-Value <= .01**